Spectral density analysis of time correlation functions in lattice QCD using the maximum entropy method

H. Rudolf Fiebig*

Physics Department, Florida International University, 11200 SW 8th Street, Miami, Florida 33199
(Received 11 February 2002; published 14 May 2002)

We study various aspects of extracting spectral information from time correlation functions of lattice QCD by means of Bayesian inference with an entropic prior, the maximum entropy method (MEM). Correlator functions of a heavy-light meson-meson system serve as a repository for lattice data with diverse statistical quality. Attention is given to spectral mass density functions, inferred from the data, and their dependence on the parameters of the MEM. We propose to employ simulated annealing, or cooling, to solve the Bayesian inference problem, and discuss the practical issues of the approach.

I. INTRODUCTION

Numerical simulations of quantum chromodynamics (QCD) on a Euclidean space-time lattice provides access to mass spectra of hadronic systems through the analysis of time correlation functions. In theory the latter are linear combinations of exponential functions

$$C(t,t_0)=Z_1 e^{-E_1(t-t_0)}+Z_2 e^{-E_2(t-t_0)}+\cdots, \qquad (1)$$

where the E_n are the excitation energies of the system and the strength coefficients

$$Z_n=|\langle n|\hat{\Phi}(t_0)|0\rangle|^2 \qquad (2)$$

are matrix elements of some vacuum-subtracted operator $\hat{\Phi}(t_0)=\Phi(t_0)-\langle 0|\Phi(t_0)|0\rangle$ between the vacuum $|0\rangle$ and a ground or excited state $|n\rangle, n>0$. In practice the exponential model (1) is fitted to noisy numerical simulation "data." The statistical quality of simulation data rarely is good enough for the two-exponential fit (1) to succeed. It is common practice to look at the large-t behavior of the correlation function $C(t,t_0)$ in a t interval where it is dominated by only one exponential, with the lowest energy, and then make a one-parameter fit to a plateau of the effective-mass function $\mu_{\text{eff}}(t,t_0)=-\partial \ln C(t,t_0)/\partial t$. Possible discretizations are

$$\mu_{\text{eff},0}(t,t_0)=-\ln\left(\frac{C(t+1,t_0)}{C(t,t_0)}\right)\simeq m_{\text{eff},0} \qquad (3a)$$

$$\mu_{\text{eff},1}(t,t_0)=\frac{C(t+1,t_0)}{C(t,t_0)}\simeq e^{-m_{\text{eff},1}} \qquad (3b)$$

$$\mu_{\text{eff},2}(t,t_0)=\frac{C(t+1,t_0)-C(t-1,t_0)}{2C(t,t_0)}$$

$$\simeq -\sinh(m_{\text{eff},2}) \qquad (3c)$$

$$\mu_{\text{eff},3}^2(t,t_0)=\frac{C(t+1,t_0)+C(t-1,t_0)-2C(t,t_0)}{C(t,t_0)}$$

$$\simeq 2(\cosh(m_{\text{eff},3})-1). \qquad (3d)$$

The expressions after the $\simeq$ are the values of μ_{eff} for a pure plateau of mass m_{eff}. The procedure implies the selection of consecutive time slices $t=t_1 \ldots t_2$ for which $\mu_{\text{eff}}=$ const, within errors, and an appropriate fit. The selection of this, so-called, plateau is a matter of judgment. A condition for reliable results is that the correlation function (1) is dominated by just one exponential term, usually the ground state. The latter can be enhanced by the use of smeared operators [1] and fuzzy link variables [2]. This analysis procedure discourages consideration of excited states. In fact it will only reliably produce results if those are suppressed. Workarounds involve diagonalization of a correlation matrix of several operators or variational techniques [3]. Those, however, still rely on plateau selection without utilizing the information contained in the entire available time-slice range of a correlation function.

As lattice simulations of QCD now aim at excited hadron states, N^*'s for example [4,5], this situation is unsatisfactory. Alternative methods employing Bayesian inference [6] are a viable option. The maximum entropy method (MEM), which involves a particular choice of the Bayesian prior probability, falls in this class. Bayesian statistics [7] is a classic subject with a vast range of applications. However, application within the context of lattice QCD is relatively new [8–12].

In this work we report on our experience using the MEM for extracting spectral mass density functions $\rho(\omega)$ from lattice-generated time correlators

$$C(t,t_0)=\int d\omega\, \rho(\omega)e^{-\omega(t-t_0)}, \qquad (4)$$

where a discrete set of time slices t is understood. Discretization of the ω integral with reasonably fine resolution leads to an ambiguous problem where the number of parameter values $\rho(\omega)$ is (typically much) larger than the number of lattice data $C(t,t_0)$. In the MEM an entropy term involving the spectral density is used as a Bayesian prior to infer $\rho(\omega)$ from the data.

*Electronic address: fiebig@fiu.edu

We here apply MEM analysis to sets of lattice correlation functions of a meson-meson system. Those particular simulations are aimed at learning about mechanisms of hadronic interaction. This will be discussed separately [13]. The lattice data generated within that project involve local and nonlocal operators. They exhibit a wide range of statistical quality from "very good" to "marginally acceptable."

Our focus here is to utilize those data as a testing ground for Bayesian MEM analysis. In contrast to other works we employ simulated annealing to the solution of the Bayesian inference problem. The main aim of this work is to explore the feasibility of this approach for extracting masses from a lattice simulation using realistic lattice data, including excitations. For the most part this translates into studying the sensitivity of the method to its native parameters.

II. BAYESIAN INFERENCE FOR CURVE FITTING

From a Bayesian point of view the spectral density function ρ in Eq. (4) is a random variable subject to a certain probability distribution functional $\mathcal{P}[\rho]$. Solution of the curve fitting problem consists in finding the function ρ which maximizes the conditional probability $\mathcal{P}[\rho \leftarrow C]$, the *posterior probability*, given a "measured" data set C. Computation of ρ is then based on Bayes' theorem [6]

$$\mathcal{P}[\rho \leftarrow C]\mathcal{P}[C] = \mathcal{P}[C \leftarrow \rho]\mathcal{P}[\rho], \qquad (5)$$

also known as "detailed balance" in a different context. The functional $\mathcal{P}[C]$, the *evidence*, gives the probability of measuring a data set C. The conditional probability $\mathcal{P}[C \leftarrow \rho]$, the *likelihood function*, determines the probability of measuring C given a spectral function ρ. Finally $\mathcal{P}[\rho]$, the Bayesian *prior*, defines a constraint on the spectral density function ρ. Its choice is a matter of judgment. Ideally, the prior should reflect the physics known about the system, for example an upper limit on the hadronic mass scale. The posterior probability is the product of the likelihood function and the prior $\mathcal{P}[\rho \leftarrow C] = \mathcal{P}[C \leftarrow \rho]\mathcal{P}[\rho]/\mathcal{P}[C]$, where the *evidence* merely plays the role of a normalization constant [6]. Indeed, the normalization condition $\int[d\rho]\mathcal{P}[\rho \leftarrow C] = 1$ applied to Eq. (5) gives $\mathcal{P}[C] = \int[d\rho]\mathcal{P}[C \leftarrow \rho]\mathcal{P}[\rho]$. Thus, for a fixed C, we have

$$\mathcal{P}[\rho \leftarrow C] \propto \mathcal{P}[C \leftarrow \rho]\mathcal{P}[\rho]. \qquad (6)$$

The curve fitting problem requires the product of the *likelihood function* and the *prior* function.

A. Spectral density

Our lattice data come from correlation functions built from heavy-light meson-meson operators

$$\Phi_v = v_1\Phi_1 + v_2\Phi_2, \qquad (7)$$

where Φ_1 and Φ_2 involve local and non-local meson-meson fields, respectively, at relative distance r, and v are some coefficients [14,15]. On a finite lattice the corresponding cor-

relator $C_v(t,t_0) = \langle \hat{\Phi}_v^\dagger(t)\hat{\Phi}_v(t_0)\rangle$, where $\hat{\Phi} = \Phi - \langle\Phi\rangle$, has a purely discrete spectrum

$$C_v(t,t_0) = \sum_{n \neq 0} |\langle n|\Phi_v(t_0)|0\rangle|^2 e^{-\omega_n(t-t_0)}. \qquad (8)$$

Here $|n\rangle$ denotes a complete set of states with energies ω_n, some of which may be negative due to periodic lattice boundary conditions and operator structure. Our normalization conventions for forward and backward going propagators are determined by defining

$$\exp_T(\omega,t) = \Theta(\omega)e^{-\omega t} + \Theta(-\omega)e^{+\omega(T-t)}, \qquad (9)$$

where $0 \leq t < T$, and Θ denotes the step function. We then expect the lattice data to fit the following model

$$F(\rho_T|t,t_0) = \int_{-\infty}^{+\infty} d\omega \rho_T(\omega)\exp_T(\omega,t-t_0), \qquad (10)$$

where $\rho_T(\omega)$ is a spectral density function, defined for positive (forward) and negative (backward) frequencies. The requirement that the model be exact, $F(\rho_T|t,t_0) = C_v(t,t_0)$, leads to

$$\rho_T(\omega) = \sum_{n \neq 0} \delta(\omega - \omega_n)|\langle n|\Phi_v(t_0)|0\rangle|^2$$

$$\times [\Theta(\omega_n) + \Theta(-\omega_n)e^{-\omega_n T}]. \qquad (11)$$

Thus a discrete sum over δ peaks is the theoretical form of the spectral function. Our objective is to compute $\rho_T(\omega)$ from lattice data using Bayesian inference.

B. Likelihood function

Toward this end we proceed to construct the likelihood function. The lattice data come in the form of an average over N_U gauge configurations

$$C_v(t,t_0) = \frac{1}{N_U}\sum_{n=1}^{N_U} C_v(U_n|t,t_0), \qquad (12)$$

where $C_v(U_n|t,t_0)$ is the value of an operator, in this case $\hat{\Phi}_v^\dagger(t)\hat{\Phi}_v(t_0)$, in one gauge field configuration U_n. Correlation function data on different time slices are stochastically dependent. Their errors are described by the covariance matrix

$$\Gamma_v(t_1,t_2) = \frac{1}{N_U}\sum_{n=1}^{N_U} (C_v(t_1,t_0) - C_v(U_n|t_1,t_0))$$

$$\times (C_v(t_2,t_0) - C_v(U_n|t_2,t_0)). \qquad (13)$$

The χ^2 distance of the spectral model (10) from the lattice data then is

$$\chi^2 = \sum_{t_1,t_2} (C_v(t_1,t_0) - F(\rho_T|t_1,t_0))\Gamma_v^{-1}(t_1,t_2)$$

$$\times (C_v(t_2,t_0) - F(\rho_T|t_2,t_0)). \qquad (14)$$

For numerical work a discretization scheme of the ω integral in Eq. (10) is required. Our choice is

$$F(\rho_T|t,t_0) \simeq \sum_{k=K_-}^{K_+} \rho_k \exp_T(\omega_k, t-t_0) \qquad (15)$$

where $\omega_k = \Delta\omega k$, $\Delta\omega$ is an appropriate (small) interval, $\rho_k = \Delta\omega\rho_T(\omega_k)$, and $K_- < 0 < K_+$.

The likelihood function $\mathcal{P}[C \leftarrow \rho]$ describes the probability distribution of the data C given a certain parameter set ρ. If we imagine that the data are obtained by a large number of measurements, at fixed ρ, then the probability distribution for C is Gaussian by virtue of the central limit theorem,

$$\mathcal{P}[C \leftarrow \rho] \propto e^{-\chi^2/2}. \qquad (16)$$

This is the standard argument for employing the above form of the likelihood function in the context of Bayesian inference [6,16].

C. Entropic prior

In case some information is available about the physics of the system it can be used to constrain the parameter space of the model. This is the role of the Bayesian prior. In the standard approach plateau methods are a severe form of imposing restrictions. A two-exponential fit (1), if feasible, is less constraining. In a Bayesian context it is possible to gradually increase the number of exponentials until convergence is reached. This is a strategy advocated in [12]; see also [17]. There, the model for the correlation function is $\Sigma_n A_n e^{-E_n t}$, initially with small number of terms, which is then constrained by the Bayesian prior $\exp(-\Sigma_n[(A_n - \bar{A}_n)^2/2\bar{\sigma}_{A_n}^2 + (E_n - \bar{E}_n)^2/2\bar{\sigma}_{E_n}^2])$. The quantities $\bar{A}_n, \bar{\sigma}_{A_n}, \bar{E}_n, \bar{\sigma}_{E_n}$ are input. Their choice is inspired by prior knowledge about the physics of the system.

On the other hand, there is usually no *a priori* information about the location and the strengths of the peaks in the mass spectrum. The view that only *minimal information* is available about the spectral density function can also be implemented in the Bayesian prior. The information content, in the sense of [18–20], is measured by the entropy $S = -\Sigma_k \rho_k \ln(\rho_k/m)$, on some scale m. Rather, a commonly used variant is the Shannon-Jaynes entropy [6]

$$S[\rho] = \sum_{k=K_-}^{K_+} \left(\rho_k - m_k - \rho_k \ln\frac{\rho_k}{m_k} \right). \qquad (17)$$

Note that $\rho_k \geqslant 0$, according to Eq. (11). The configuration $m = \{m_k : K_- \leqslant k \leqslant K_+\}$ is called the default model. We have $S \leqslant 0$, $\forall\rho$, while $S = 0 \Leftrightarrow \rho = m$. The default model is a unique absolute maximum of S. Choosing the prior probability as

$$\mathcal{P}[\rho] \propto e^{\alpha S} \qquad (18)$$

entails that $\mathcal{P}[\rho]$ is maximal in the absence of information about ρ. An argument for Eq. (18) can be found in [6]. The entropy strength α and the default model m are parameters.

D. Computing the spectral density

With Eqs. (16) and (18) the posterior probability (6) becomes

$$\mathcal{P}[\rho \leftarrow C] \propto e^{-(\chi^2/2 - \alpha S)}. \qquad (19)$$

We wish to maximize $\mathcal{P}[\rho \leftarrow C]$ with respect to ρ, at fixed C. It can be shown that both $\chi^2[\rho]$ and $-S[\rho]$ are convex functions of $\rho = \{\rho_k : K_- \leqslant k \leqslant K_+\}$. Thus

$$W[\rho] = \chi^2/2 - \alpha S \qquad (20)$$

has a unique absolute minimum. The functional $W[\rho]$ is nonlinear and maximally nonlocal since all degrees of freedom ρ_k are coupled via the covariance matrix (13) in Eq. (14). To find the minimum of $W[\rho]$ one option is to use singular value decomposition (SVD); see [11].

In keeping with the Bayesian probabilistic interpretation of ρ an attractive alternative is to employ stochastic methods to solve the optimization problem $W[\rho] = \min$. In this work we employ simulated annealing [21], equivalently known as cooling. The algorithm is based on the partition function

$$Z_W = \int [d\rho] e^{-\beta_W W[\rho]}. \qquad (21)$$

It involves the generation of equilibrium configurations ρ while gradually increasing β_W from an initially small value, following some annealing schedule. The latter is subject to experimentation. We have used the power law

$$\beta_W(n) = (\beta_1 - \beta_0)(n/N)^\gamma + \beta_0 \qquad (22)$$

with annealing steps $n = 0 \ldots N$ between an initial β_0 and a final β_1.

A standard Metropolis algorithm was used to generate configurations ρ with the distribution in Eq. (21). In consecutive sweeps local updates were done by multiplying the spectral parameters with positive random numbers, $\rho_k \rightarrow x\rho_k$. Some experimenting showed that Γ-distributed random deviates of order two, $p_a(x) = x^{a-1}e^{-x}/\Gamma(a), a=2$, work quite efficiently at an acceptance rate centered at about 50%.

III. RESULTS

All simulations were done on an $L^3 \times T = 10^3 \times 30$ lattice. The gauge field and fermion actions are both anisotropic, with a bare aspect ratio of $a_s/a_t = 3$, and tadpole improved. The gauge field action is that of [3] with $\beta = 2.4$, leading to a spatial lattice constant of $a_s \simeq 0.25$ fm, $a_s^{-1} \simeq 800$ MeV. For the light fermions we use a clover improved Wilson action. The hopping parameter $\kappa = 0.0679$ results in a mass ratio $m_\pi/m_\rho \simeq 0.75$. Following [3] only spatial directions are improved with spatial tadpole renormalization factors $u_s = \langle \Box \rangle^{1/4}$, while $u_t = 1$ in the time direction. Clover terms involving time directions are omitted.

Some guidance for a reasonable ω discretization (15), of Eq. (10), may be derived from the physical value of the lattice constant a_t, and the time extent Ta_t of the lattice.

Admissible lattice energies thus lie approximately between $\pi/a_t \approx 7.5$ GeV and $\pi/Ta_t \approx 250$ MeV, or ≈ 3 and ≈ 0.1 in units of a_t^{-1}. In practice these are somewhat extreme bounds. Typical hadronic excitation energies are much less than $\pi/a_t \approx 7.5$ GeV. The lower bound, on the other hand, may well be ignored as a criterion for choosing the discretization interval $\Delta\omega$, because the theoretical form of ρ is a superposition of δ peaks. Thus the resolution $\Delta\omega$ should be small, in fact much smaller than ≈ 0.1. A reasonable lower bound is the likely statistical error on spectral masses. For most of the results presented here $\Delta\omega = 0.04$, and $K_- = -40, K_+ = +80$, leading to $-1.6 \leq \omega \leq +3.2$, were used with Eq. (15).

With the annealing schedule (22), we have used $N = 2048$ cooling steps, at 128 sweeps per temperature, starting at $\beta_0 = 1.0 \times 10^{-5}\beta$ and ending at $\beta_1 = 1.0 \times 10^{+5}\beta$, with a geometric average of $\beta = 1.0 \times 10^{+3}$. These choices are an outcome of experimentation. With $\gamma \simeq 16.61$ in Eq. (22) about half of the cooling steps operate in the regions $\beta_W(n) < \beta$ and $\beta_W(n) > \beta$, respectively. The average value β is such that $\beta_W W[\rho]$ fluctuates about one at around $N/2$ cooling steps. With the final annealing temperature kept constant, $\beta_W = \beta_1$, an additional 1024 steps were done keeping 16 configurations ρ in order to measure cooling fluctuations.

Results are robust within reasonable changes of the annealing schedule parameters; they were used throughout this work.

A. Entropy weight dependence

The extent to which the spectral density ρ depends on the value of the entropy weight parameter α, in Eq. (20), is a primary concern. We are interested in testing the α dependence for a case where both ground and excited states are prominently present in a time correlation function. For this reason we have constructed a mock correlator $C_X(t,t_0)$. Its building blocks were the eigenvalues of the 2×2 correlation matrix $C_{ij}(t,t_0) = \langle \hat{\Phi}_i^\dagger(t)\hat{\Phi}_j(t_0) \rangle$ using the above mentioned local and non-local meson fields. Pieces of those were arbitrarily matched and enhanced in order to exhibit a multiexponential correlation function. While $C_X(t,t_0)$ bears no physical significance, its rich structure provides a useful laboratory for testing the α dependence of the spectral density function.

In Fig. 1 we show a sequence of six pairs of Bayesian fits to the mock correlator $C_X(t,t_0)$ and the corresponding spectral densities ρ for a wide range of entropy weights α. The stability of the global structure of ρ while α changes from 1.4×10^{-2} to $1.4 \times 10^{+7}$ is most notable [22]. As α becomes larger entire peaks vanish starting with the smallest one. The reason is that the annealing action (20) gradually loses memory of the data, contained in χ^2, in favor of the entropy. The fit at $\alpha = 1.4 \times 10^{+7}$ exhibits the onset of a smoothing of the microstructure, starting with the largest peak. This is the signature of emerging entropy dominance over the data. In practice this situation should be avoided. In our case entropy strengths in the region $\alpha < 10^{+6}$ over eight orders of magnitude give stable consistent results. It has been proposed that spectral functions be integrated over α to avoid the param-
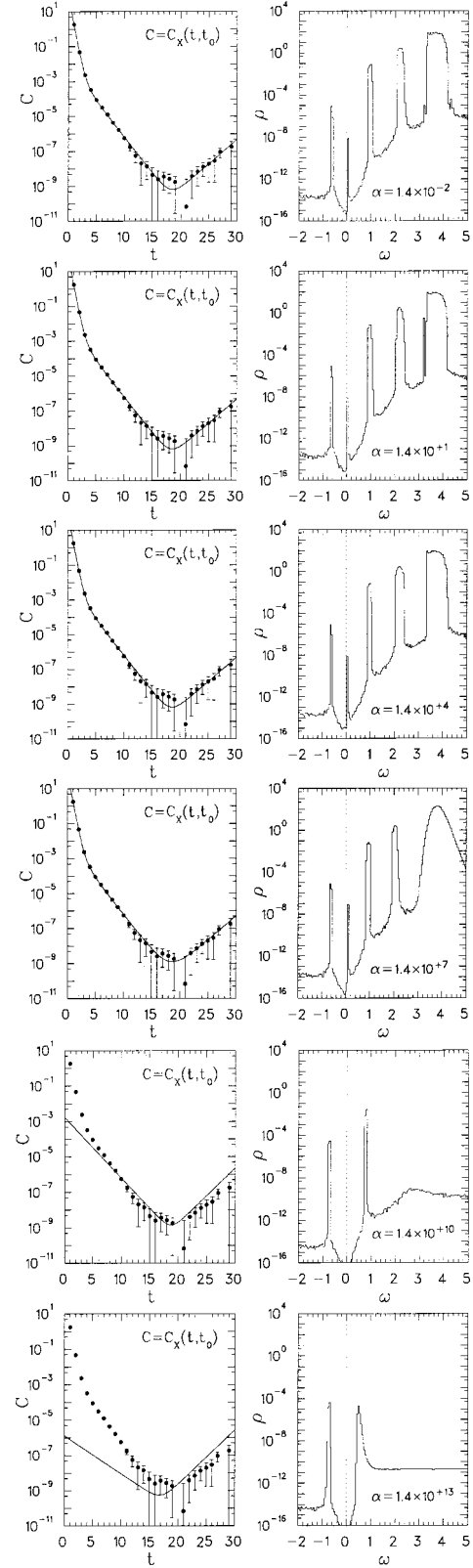

FIG. 1. Mock time correlation functions C_X shown with their Bayesian fits (solid lines), and the corresponding spectral densities ρ. The sequence of six pairs of figures shows how the spectral fit evolves through a change of the entropy weight α through 15 orders of magnitude.
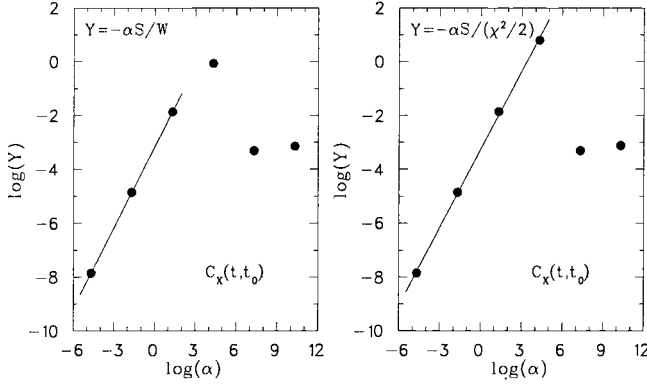
FIG. 2. Empirical dependence of the entropy loads $Y_{S/W}$ and Y_{S/χ^2} on the entropy weight parameter α, see Eqs. (23),(24). These results are for the mock correlator $C_X(t,t_0)$. The lines indicate the extent of linear relationships.

eter dependence [6]. Inspection of our results clearly indicates that averaging over α would be without consequence to the gross structure of ρ; only the microstructure would be affected. Even the region $\alpha > 10^{+6}$ could be included, since the magnitude of ρ quickly becomes insignificant.

In order to decide on a tuning criterion for α it is useful to monitor quantities like

$$Y_{S/W} = \frac{\langle -\alpha S \rangle_{\beta_W \to \infty}}{\langle W \rangle_{\beta_W \to \infty}} \quad (23)$$

$$Y_{S/\chi^2} = \frac{\langle -\alpha S \rangle_{\beta_W \to \infty}}{\langle \chi^2/2 \rangle_{\beta_W \to \infty}}, \quad (24)$$

where $\langle \ldots \rangle_{\beta_W \to \infty}$ refers to the annealing average measured at the final cooling temperature, β_1. We will refer to the above quantities as entropy loads. Those are shown in Fig. 2. It turns out that $\log(Y)$ depends linearly on $\log(\alpha)$ in the regions $\log(\alpha) < +1$ and $\log(\alpha) < +4$, for $Y_{S/W}$ and Y_{S/χ^2}, re-
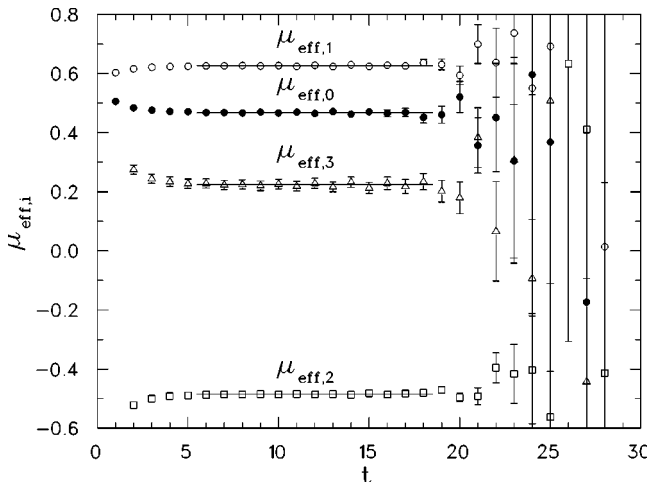
TABLE I. Plateau masses derived from Eqs. (3a)–(3d) on the time slice range $6 \leqslant t \leqslant 18$. The entry E_1 is the Bayesian result with Δ_1 being the peak width (standard deviation) computed from the spectral density function ρ. Statistical errors are derived from a gauge configuration jackknife analysis.

$m_{\text{eff},0}$	$m_{\text{eff},1}$	$m_{\text{eff},2}$	$m_{\text{eff},3}$	E_1	Δ_1
0.468(8)	0.468(7)	0.468(3)	0.47(2)	0.471(15)	0.017(6)

spectively. (In fact $Y \approx 6.2 \times 10^{-4} \alpha$.) Beyond the linear region too much entropy is loaded into the annealing action W, leading to a smoothing of peaks, as seen in Fig. 1. Empirically, the criterion emerging from this observation is to tune the entropy weight such that $\log(Y) \approx -2 \pm 1$ within the linear region. The precise value of $\log(Y)$ is not important; also $Y = Y_{S/W}$ and $Y = Y_{S/\chi^2}$ work equally well. As is evident from Fig. 1 results are extremely robust against varying α.

B. Single-meson spectrum

The correlation function $c(t,t_0) = \langle \hat{\Phi}^\dagger(t) \hat{\Phi}(t_0) \rangle$ of a single pseudoscalar heavy-light meson operator $\phi(t) = \Sigma_{\vec{x}} \bar{Q}_A(\vec{x}t) \gamma_5 q_A(\vec{x}t)$ delivers high quality data in this simulation. We use these to compare with plateau methods and make some observations relevant to the present stochastic approach to the MEM.

In Fig. 3 plots of the mass function discretizations (3a)–(3d), built from $c(t,t_0)$, and the corresponding plateau fits are displayed. Plateau fits were made directly to $\mu_{\text{eff},i=0\ldots3}$. The resulting masses, other than $m_{\text{eff},0}$, are from solving Eqs. (3b)–(3d). Table I shows that those are consistent within statistical (jackknife) errors.

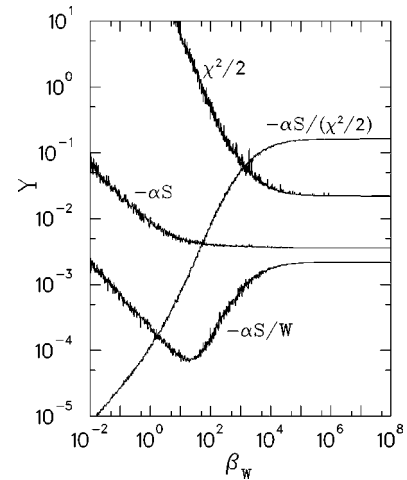Figure 4 gives a sense of the annealing dynamics. Beside Eqs. (23) and (24) also shown are

FIG. 3. Effective mass functions (3a)–(3d) for a single heavy-light meson. The horizontal lines are plateau fits in the time slice range $6 \leqslant t \leqslant 18$.

FIG. 4. Annealing dynamics in terms of the tuning functions $Y_{S/W}$, Y_{S/χ^2}, and Y_S, Y_{χ^2}, versus the cooling parameter β_W. The graphs are labeled with reference to the entropy loads (23),(24), and (25),(26). This example is for the single-meson correlator, with entropy strength $\alpha = 5.0 \times 10^{-5}$ and a constant default model $m = 1.0 \times 10^{-12}$.
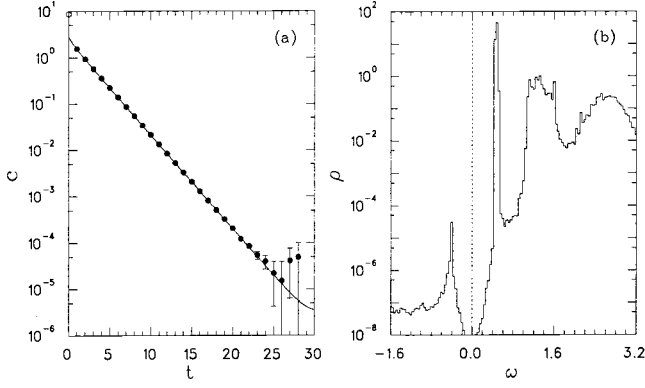
FIG. 5. Time correlation function for a single heavy-light meson together with a Bayesian fit (a), and the corresponding spectral density function (b). This result stems from a single random start, with entropy weight $\alpha = 5.0 \times 10^{-5}$, and a constant default model $m = 1.0 \times 10^{-12}$.

$$Y_S = \langle -\alpha S \rangle_{\beta_W \to \infty} \qquad (25)$$

$$Y_{\chi^2} = \langle \chi^2/2 \rangle_{\beta_W \to \infty}. \qquad (26)$$

In [15] and [14] the use of $Y_{S/W}$ was advocated as a tuning criterion. In view of Fig. 4 Y_{S/χ^2} appears to be a better choice given its monotonic nature. A target entropy load of $Y_{S/\chi^2} \approx 10^{-1 \pm 1}$ is a safe tuning criterion, provided the cooling algorithm runs in the (upper) linear region; see Fig. 2.

The Bayesian analysis of the time correlation function $c(t, t_0)$ is shown in Fig. 5. The solid line in Fig. 5(a) derives from the computed spectral density ρ, via Eq. (15). With the exception of $t_0 = 0$ all available time slices were used. Parameters are $\alpha = 5.0 \times 10^{-5}$, for the entropy strength, a constant default model $m = 1.0 \times 10^{-12}$, and a random annealing start about m. The graph of ρ in Fig. 5(b) exhibits a global structure consisting of distinct peaks, some broad, and a microstructure of fluctuations on the scale of $\Delta \omega$. The microstructure depends on details of the annealing process, particularly the start configuration. Clearly, it makes no sense to infer the microstructure from the data. The reason is that only $T - 1 = 29$ data points do not contain enough information to determine $K_+ - K_- + 1 = K = 121$ spectral parameters (with any sizable probability).

On the other hand, the global structure is a stable feature. In the region $\omega > 0$ three peaks can be distinguished in Fig. 5(b). By way of inspection we loosely define

$$\delta_n = \{\omega : \omega \in \text{peak } n\}, \quad n = 1, 2 \dots. \qquad (27)$$

Then, for each peak n, we may calculate the volume Z_n, the mass E_n, and the width Δ_n, according to

$$Z_n = \int_{\delta_n} d\omega \rho_T(\omega) \qquad (28)$$

$$E_n = Z_n^{-1} \int_{\delta_n} d\omega \rho_T(\omega) \omega \qquad (29)$$
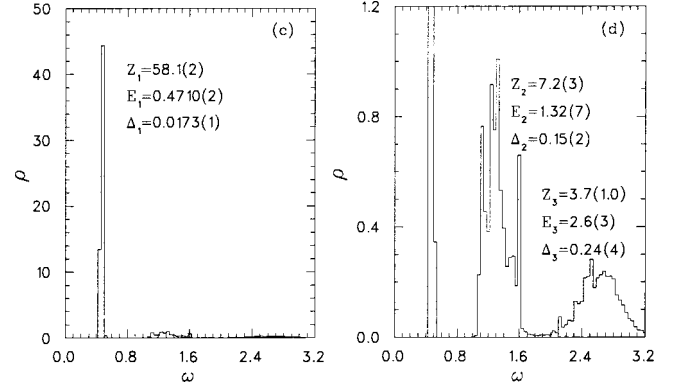

FIG. 6. Spectral density ρ for a single heavy-light meson, same as in Fig. 5(b), but on linear scales, emphasizing the ground and the excites states (c) and (d), respectively. The uncertainties of Z_n, E_n, and Δ_n are standard deviations from eight annealing runs.

$$\Delta_n^2 = Z_n^{-1} \int_{\delta_n} d\omega \rho_T(\omega)(\omega - E_n)^2. \qquad (30)$$

These integrated, low moment, quantities are evidently insensitive to the microstructure. They constitute the information that reasonably can be expected to flow from the Bayesian analysis.

The spectral density of Fig. 5(b) is replotted in Fig. 6 on linear scales. The tall narrow peak in Fig. 6(c) corresponds to the plateau masses of Fig. 3, as listed in Table I. There, the entries E_1 and Δ_1 are the Bayesian results. Their statistical errors are derived from a jackknife analysis selecting four subsets of gauge configurations. (Note that the uncertainties in Fig. 6 are standard deviations from eight annealing starts.) Cold starts from the default model m were used to suppress the dependence on the annealing start configuration. The peak width Δ_1 is comparable to the gauge configuration statistical error. This is the exception. With correlation function data of lesser quality (as with the two-meson operators below) the size of the peak width is typically larger than the statistical error. It appears that the peak width Δ_n is related to the size Θ_n of the corresponding effective mass function plateau, like in Fig. 3, or the size of the log-linear stretch in a plot as in Fig. 5(a). As a very coarse description $\Delta_n \Theta_n \approx \text{const}$ comes to mind. Using $\Theta_1 = 12$ and $\Delta_1 = 0.017$ we have const≈ 0.2. The peaks $n = 2$ and $n = 3$ seen in Fig. 6(d) would thus appear to originate from $\Theta_n \approx 0.2/\Delta_n$, or 1.3 and 0.8 time slices, respectively. (By inspection of Fig. 3 as many as 5 time slices appear involved, however.) The physical relevance of, at least, peak $n = 3$ is therefore questionable. On the other hand it is remarkable that the maximum entropy method is sensitive to the slightest details in the correlation function data.

C. Default model dependence

The Shannon-Jaynes entropy (17) implies the possible dependence of the computed spectral density ρ on the default model $m = \{m_k : K_- \leq k \leq K_+\}$. We explore the m dependence
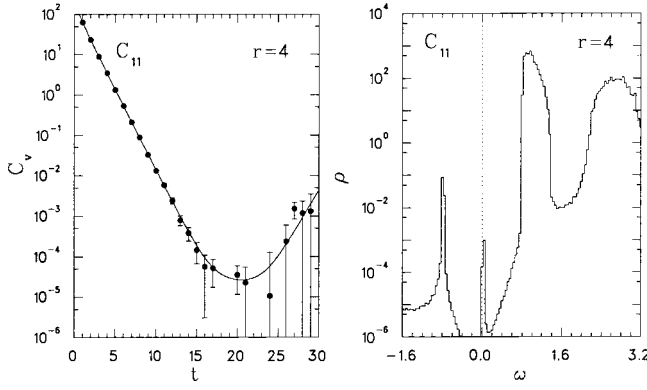
FIG. 7. Correlation function $C_{11} = \langle \hat{\Phi}_1^\dagger(t) \hat{\Phi}_1(t_0) \rangle$ of a heavy-light meson-meson operator at relative distance $r=4$. The Bayesian fit (solid line) is from the spectral density ρ shown on the right. At $\alpha = 2 \times 10^{-6}$ and constant default model $m = 1.0 \times 10^{-12}$ the spectral density ρ is obtained from an average over eight random annealing start configurations. The average entropy load is $Y_{S/\chi^2} = 0.477$ for these runs.

using as an example the time correlation function C_v with $v_1 = 1$, $v_2 = 0$, in the notation of Eq. (7), at relative distance $r = 4$.

Figure 7 shows the time correlation function data together with the Bayesian fit, and the corresponding spectral density ρ. The latter is the average over eight random annealing start configurations. This has the effect of smoothing out the microstructure of ρ. We have used a constant default model $m_k = 1.0 \times 10^{-12}$, all k.

The stability of this result is tested by varying the default model through 15 orders of magnitude, $m = 10^{-12} \ldots 10^{+3}$, as shown in Fig. 8. To keep effects of the annealing start configuration small cold starts from $\rho = m$, using the same random seed, were employed for all values of m. In each case the entropy strength parameter α was tuned such that the entropy load Y_{S/χ^2} remained constant. Aside from the familiar microstructure fluctuations, the global (physical) features are stable within the range of, a remarkable, fifteen orders of magnitude. Numerical experiments with non-constant m do not change this assessment. In Table II are listed the three integral quantities (28)–(30) averaged over the six default models together with the corresponding standard deviations. Their smallness (0.3–3 %) attests to the default model independence of the Bayesian fits. Given the huge variation of the default model the stability of ρ is remarkable.

D. Annealing start dependence

The annealing algorithm starts with some initial spectral configuration ρ_{ini}. Depending on the purpose we have used cold starts from the default model, $\rho_{\text{ini}} = m$, or random starts from the default model, $\rho_{\text{ini},k} = x_k m_k$, where the x_k are drawn from a gamma distribution of order two, $p_a(x) = x^{a-1} e^{-x} / \Gamma(a), a = 2$. The global features of the final spectral density are of course independent of the start configuration, but the microstructure of ρ is not. The reason is that in practice the annealing process is neither infinitely slow nor is
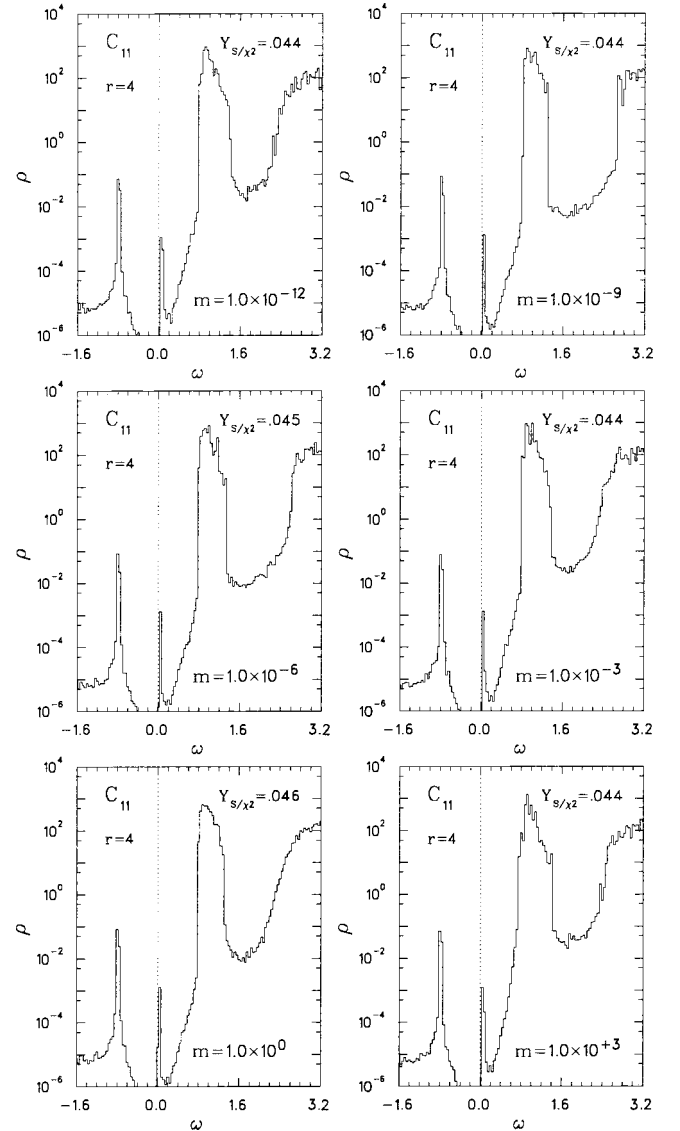

FIG. 8. A sequence of spectral densities ρ obtained from a wide range of constant default models m; see inserts. The entropy strength parameter α was tuned to keep the entropy load constant, $Y_{S/\chi^2} \approx 0.045$. The operator is the same as in Fig. 7.

the final cooling temperature β_1^{-1} exactly zero. Therefore the annealing result for ρ settles close to the global minimum, say ρ_{min}, of $W[\rho]$. Considering annealing (thermal) fluctuations only, we expect the deviation $|\rho - \rho_{\text{min}}|$ to be large in directions (of ρ space) where the minimum is shallow. Ther-

TABLE II. Averages of volume, energy, and width of the dominant peak seen in Fig. 8 over the six default model choices $m = 10^{-12} \ldots 10^{+3}$ at fixed entropy load $Y_{S/\chi^2} \approx 0.045$. The uncertainties are the corresponding standard deviations. The entry $m_{\text{eff},0}$ is the plateau mass (3a) from Fig. 11 with the statistical (jackknife) error, see Sec. III E.

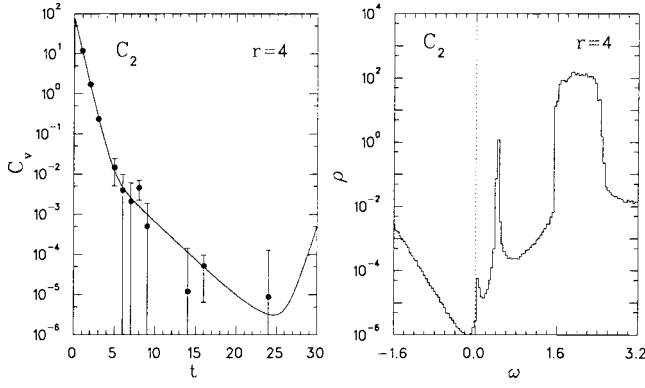| Z_1	E_1	Δ_1	$m_{\text{eff},0}$
3923.(18.)	0.972(3)	0.100(3)	0.94(1)

FIG. 9. Excited state correlation function C_2 of a heavy-light meson-meson operator at relative distance $r=4$. The Bayesian fit (solid line) is from the spectral density ρ shown on the right. At $\alpha = 5 \times 10^{-7}$ and constant default model $m = 1.0 \times 10^{-12}$ the spectral density ρ is obtained from an average over eight random annealing start configurations.

mal fluctuations are easily controlled, however. Those were kept negligible in the present study. More importantly, there may be local minima close to ρ_{min} which are only slightly larger than $W[\rho_{min}]$. This situation invites computing a set of spectral densities from different, say random, initial configurations. The averages and standard deviations of the ρ_k then give some insight into the structure of the peak and the nature of the minimum of W and its neighborhood.

To present an example we have selected an excited state time correlation function $C_2(t,t_0)$ of the meson-meson system at relative distance $r=4$. $C_2(t,t_0)$ is the smaller of the eigenvalues of the 2×2 correlation matrix $C_{ij}(t,t_0) = \langle \hat{\Phi}_i^\dagger(t) \hat{\Phi}_j(t_0) \rangle$, on each time slice. The reason for selecting this operator is to see how the MEM responds to a data set that is marginally acceptable, at best. Figure 9 shows the correlator and the corresponding spectral density obtained from an average over eight Bayesian fits based on different random annealing start configurations. The same spectral density is displayed in the first frame of Fig. 10 on a linear scale. The dotted lines represent the limits within one standard deviation. The remaining three frames of Fig. 10 show spectral functions from selected single start configurations. They illustrate the microstructure fluctuations.

We argue that the microstructure, on a fine discretization scale $\Delta\omega$, is extraneous information. On the basis that the number of measured data points, as supplied by the time correlation function $C_v(t,t_0)$, is much smaller than the number of inferred parameters ρ_k, exact knowledge of ρ would actually constitute an information gain not supported by the data. Rather, only averages of suitable observables based on the inferred spectral density, like Eqs. (28)–(30) for example, are relevant information that can be extracted from the Bayesian analysis. Whether or not the ρ average of a certain observable is relevant information supported by the data may possibly be decided by the criterion that the standard deviation with respect to different annealing starts be small. From Table III we see that the standard deviations for the small-moment averages (28),(29),(30) are comparable to typical gauge configuration statistical errors, for example those in
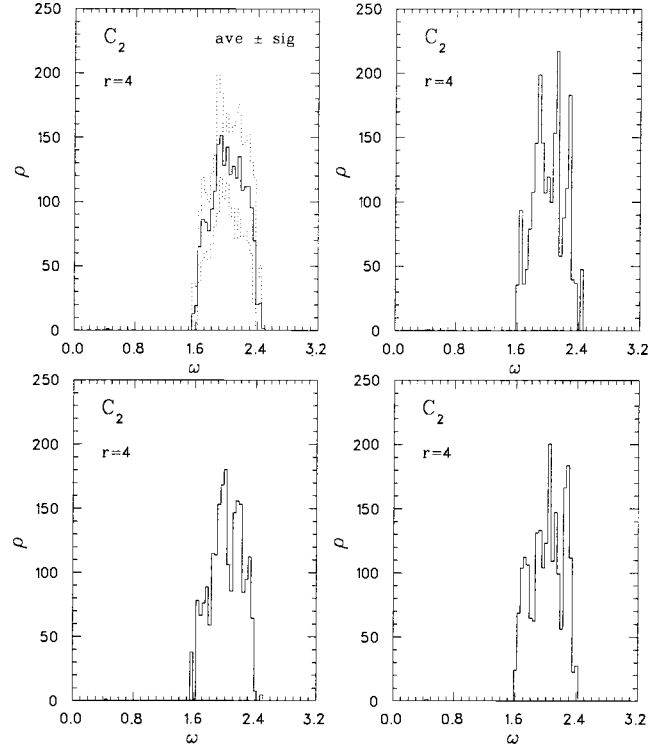

FIG. 10. Spectral densities ρ of the excited state correlation function of Fig. 9. The sequence of four frames shows the average (ave) over a sample of eight random annealing start configurations including the bounds (dotted lines) of one standard deviation ($\pm$ sig), and three selected examples of spectral functions making up that sample.

Table I. This should be an acceptable test; certainly high resolution operators would fail it.

E. Relation to plateau methods

Aside from the obvious differences in algorithm and philosophy it is important to understand that the traditional plateau method and the celebrated Bayesian approach also are distinctly different in the way they utilize the lattice correlator data. First, the former uses data on only a (subjectively) truncated contiguous set of time slices while completely ignoring the rest, whereas the latter utilizes the data on all available time slices without bias. Second, in the plateau method the stochastic dependence of the data between the plateau time slices is often ignored (see *Note added in proof*), whereas in the Bayesian approach the dependence is fully accounted for through the covariance matrix (13).

TABLE III. Averages of volume, energy, and width of the dominant peak seen in Fig. 10 over eight random annealing start configurations, at fixed $\alpha = 5.0 \times 10^{-7}$ and constant default model $m = 10^{-12}$. The entry $m_{\text{eff},0}$ is the plateau mass (3a) from Fig. 11 with the statistical (jackknife) error; see Sec. III E.

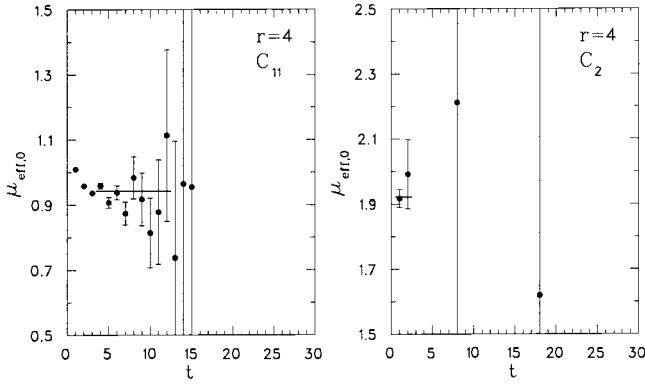| Z_1	E_1	Δ_1	$m_{\text{eff},0}$
2156.(11.)	2.012(7)	0.214(11)	1.92(3)

FIG. 11. Effective mass functions $\mu_{\text{eff},0}$, see Eq. (3a), of the correlator examples C_{11} and C_2 shown in Figs. 7 and 9. The plateaus are shown as horizontal lines extending over 9 and 2 time slices, respectively.

Hence, the traditional plateau method and the Bayesian inference approach cannot be compared on an equal footing. In particular, their systematic errors are in principle different.

A comparison of those methods is thus reduced to observing their responses to the same data sets. If the numerical quality of data is very good both methods (in fact any two methods) will of course give the same answers. An example is the single-meson case discussed above; see Table I. In case of imperfect numerical data, however, the two methods should be expected to give different results. We illustrate this point by showing in Fig. 11 the effective mass functions (3a) of the correlators C_{11} and C_2 displayed in Figs. 7 and 9, respectively. While the C_{11} data are somewhat level within 9 time slices, the C_2 data are extreme in the sense that only 2 data points are available to the plateau method. Bayesian inference, as illustrated by Fig. 9 and also Fig. 10, has no problem responding with a distinct peak. The reason, of course, is that the entire set of correlator data including their correlations is available to the Bayesian approach.

In Tables II and III we compare the plateau masses $m_{\text{eff},0}$ obtained from Eq. (3a) to the Bayesian results E_1. The numbers differ by about $3-5\%$. Note that the statistical (jackknife) errors on the plateau masses are much smaller. Because of the data truncation the method has no way of "knowing" about the poor quality of the correlator data, particularly in the C_2 case of the excited state correlator. The Bayesian method, on the other hand, is fully "aware" of this fact and conveys this information by responding with a sizable peak width Δ_1, which easily encompasses the plateau masses.

This raises the question whether Bayesian peak widths or plateau mass statistical errors are a better measure for the uncertainty of masses extracted from lattice simulations. The answer is beyond the scope of this work.

IV. SUMMARY AND CONCLUSION

We have reported on our experience using Bayesian inference with an entropic prior, the maximum entropy method, to extract spectral information from lattice generated time correlation functions. The latter were taken from a simulation aimed at studying hadronic interaction, but used here only as a repository of simulation data of diverse quality.

In contrast with other works the method of choice for extracting spectral densities was simulated annealing.

Between the maximum entropy method and simulated annealing there were three major concerns about the parameter and algorithm dependence of the results: Dependence on (i) the entropy weight, (ii) the default model, and (iii) the annealing start configuration. Besides suggesting strategies for parameter tuning, independence of the Bayesian inferred spectral density ρ on (i) the entropy weight, and (ii) the default model could be demonstrated within a range of eight and fifteen orders of magnitude of the parameters, respectively. Concerning the annealing start configuration dependence (iii) we argued that only spectral density averages of certain operators are acceptable. From an information theory point of view [18], those should be operators insensitive to the microstructure of the inferred spectral density. In particular, keeping in mind that the theoretical structure of the lattice spectral function is a superposition of distinct peaks, those operators include the spectral peak volume Z_n, or normalization, the peak energy E_n, or mass, and the peak width Δ_n, or standard deviation.

Bayesian inference has too long been ignored by the lattice community as an analysis tool. It has an advantage over conventional plateau methods for extracting hadron masses from lattice simulations because the entire information contained in the correlator function, or matrix, is utilized. This aspect is particularly important where excited state masses are desired, since the noise contamination of their signal can be significant. The maximum entropy method is very robust with respect to changing its parameters. Simulated annealing is practical for obtaining spectral density functions. The method should be given serious consideration as an alternative to conventional ways.

Note added in proof. Uncorrelated fits to a mass function may be justified if the number of gauge configurations N is large compared to the number of plateau times slices D; see Ref. [23]. There the condition $N > \max(D^2, 10(D+1))$ applied to the situation of Fig. 3 gives $708 > 169$.

[1] C. Alexandrou, S. Güsken, F. Jegerlehner, K. Schilling, and R. Sommer, Nucl. Phys. **B414**, 815 (1994).

[2] C. Albanese *et al.*, Phys. Lett. B **192**, 163 (1987).

[3] C.J. Morningstar and M.J. Peardon, Phys. Rev. D **60**, 034509 (1999).

[4] F.X. Lee and D.B. Leinweber, Nucl. Phys. B (Proc. Suppl.) **73**, 258 (1999); S. Sasaki, T. Blum, and S. Ohta, Phys. Rev. D **65**, 074503 (2002).

[5] QCDSF Collaboration, M. Göckeler *et al.*, hep-lat/0106022.

[6] M. Jarrell and J.E. Gubernatis, Phys. Rep. **269**, 133 (1996).

[7] G. Box and G. Tiao, *Bayesian Inference in Statistical Analysis* (Addison-Wesley, Reading, MA, 1973).

[8] Y. Nakahara, M. Asakawa, and T. Hatsuda, Phys. Rev. D **60**, 091503 (1999).

[9] Y. Nakahara, M. Asakawa, and T. Hatsuda, Nucl. Phys. B (Proc. Suppl.) **83**, 191 (2000).

[10] M. Asakawa, Y. Nakahara, and T. Hatsuda, Nucl. Phys. B (Proc. Suppl.) **86**, 191 (2000).

[11] M. Asakawa, T. Hatsuda, and Y. Nakahara, Prog. Part. Nucl. Phys. **46**, 459 (2001).

[12] G.P. Lepage *et al.*, Nucl. Phys. B (Proc. Suppl.) **106**, 12 (2002).

[13] H.R. Fiebig, "A lattice study of interaction mechanisms in a heavy-light meson-meson system."

[14] Lattice Hadron Physics Collaboration, H.R. Fiebig, Nucl. Phys. B (Proc. Suppl.) **106**, 344 (2002).

[15] LHP Collaboration, H.R. Fiebig, Nucl. Phys. B (Proc. Suppl.) **109**, 207 (2002).

[16] S. Brandt, *Statistical and Computational Methods in Data Analysis* (North-Holland, New York, 1976).

[17] C. Morningstar, Nucl. Phys. B (Proc. Suppl.) **109**, 185 (2002).

[18] C. Shannon and W. Weaver, *The Mathematical Theory of Communication* (University of Illinois Press, Urbana, 1949).

[19] E. Jaynes, Phys. Rev. **108**, 171 (1957).

[20] E. Jaynes, Phys. Rev. **106**, 620 (1957).

[21] S. Kirkpatrick, J. Stat. Phys. **34**, 975 (1984).

[22] In [15] and [14] a different normalization of the covariance matrix (13) was used. This can be accounted for by a rescaling of the entropy weight $\alpha = (N_U - 1)\bar{\alpha}$, where $\bar{\alpha}$ refers to the above references and $N_U = 708$.

[23] C. Michael, Phys. Rev. D **49**, 2616 (1994); C. Michael and A. McKerrell, *ibid.* **51**, 3745 (1995).