DIMENSIONALLY EXACT FORM-FREE ENERGY CONFINEMENT SCALING IN W7-AS

R. PREUSS, V. DOSE

Max-Planck-Institut für Plasmaphysik, EURATOM Association

D-85748 Garching b. München, Germany

W. VON DER LINDEN

Institut für Theoretische Physik, Technische Universität Graz

A-8010 Graz, Austria

(August 12, 1999)

Abstract

A semi empirical theory for global energy confinement in the stellarator W7-AS is presented. It employs elementary rules of probability theory together with the transformation properties of the basic equations describing plasma behavior. The semi empirical theory allows model comparison between four different kinetic plasma models. The analysis is applied to the 153 $\iota \approx 1/3$ W7-AS data documented in the international stellarator data base. We find that the collisional low-β kinetic model describes best the physics of W7-AS energy confinement. Finally the predictive power of the semi empirical theory is successfully tested in the prediction of the energy content in (single variable) density and power scans performed in W7-AS.

I. INTRODUCTION

The characterization of plasma behavior in terms of global energy confinement scaling functions has been attempted for more than twenty years. It serves mainly the purposes of inter machine comparison and extrapolation from current experiments to future devices. Two prominent current examples are those of the tokamak line from which the ITER performance is inferred and, similarly, confinement property projections of W7-X from stellarator experience. The general assumption that the scaling function may be expressed in terms of a power law has survived over the years. On the one hand this is due to its great simplicity and convenience, on the other hand we must concede that the simple power law scaling function has shown some success in predicting the operation of larger machines as, for example, TFTR and JET on the basis of data collected from small and mid-sized devices. Moreover, power law functions have a good-natured monotonic extrapolation behavior. As time went by more and more indications occurred that the power law form may have shortcomings. Already De Boo et al. [1] noted that their L-mode results could be fitted equally well by either a P^{α_P} or a (A+B/P) power dependence. Similar observations led to suggestions of offset linear scaling expressions as, for example, the Rebut-Lallia function [2]. Even the very recent paper by Stroth et al. [3] on stellarator energy confinement concludes that the density dependence might be more complicated than n^{α_n} and seems to be stronger at low densities.

In addition to engineering scaling expressions which aim at a best fit of the data regardless of dimensional correctness, dimensional constraints derived from invariance requirements of the basic equations describing plasma behavior under similarity transformations [4] have been used as auxiliary conditions in the fit of power law functions to experimental data. This would make sense if we knew that the scaling function must be of the power law type. However, this is by no means to be taken for granted as pointed out by various authors [5,6]. As a consequence we addressed in a previous paper [7] the question of how close power law scaling functions fall to dimensional constraints and obtained the result that none of the Connor Taylor (CT) models [4] was satisfied within experimental error.

This is again another indication that the power law is not the appropriate scaling function. In this paper we shall follow another route and present the calculation in detail; a short discussion of the results has already been given in [8]. We shall consider the dimensional constraints as the most general and basic information on the physics of energy confinement in a given experiment. In exploring this assumption we shall abandon the power law or any other analytical ansatz altogether and construct instead a form-free inherently dimensionally exact scaling function. Elementary probability theory is used to determine its optimal form. These form-free functions will further be employed to decide which of the four kinetic CT models applies most likely to the experimental data from W7-AS, and therefore describes the physics underlying the data set we are using. Finally we will look for the result of the energy content for certain variable settings given by these functions in the most probable model. Here an important topic is single variable scans (e.g. the variation of the energy content as function of the density with all other variables fixed) which constitute a very stringent test on any energy confinement function. Such scans are not directly accessible from published databases. On the other hand, single variable scans are experimentally cumbersome, and expensive experiments have to be performed for each and every input variable of interest. It is therefore highly desirable to extract single variable scans from existing databases by employing improved data analysis techniques.

In the second chapter we present the idea of Connor and Taylor of dimensionally exact confinement scaling explicitly. Due to the extensive use of probability theory the third chapter is exclusively devoted to this field. In chapter IV we then apply these principles to our problem. While we do not recommend this, the experienced reader may directly proceed to chapter V where the main results are presented.

II. DIMENSIONALLY EXACT FORM-FREE SCALING

Our aim in this paper is the use of dimensionally exact energy confinement scaling functions. Connor and Taylor [4] were the first to interpret experimental scaling functions in terms of constraints derived from the requirement of physical invariance under similarity transformations of the basic equations describing plasma behavior. Previous scaling functions were built on a purely empirical basis (e.g. Hugill and Sheffield [9]) or formulated in terms of (arbitrary) dimensionless parameters which were obtained by some physical considerations [10]. Since Connor and Taylor found that the experimental scaling was incompatible with any of their plasma models, they suggested that the theoretically derived dimensional constraints should be incorporated directly in the power law ansatz. This proposal has been followed subsequently on many occasions [11]. It consists of expressing the energy content W of the toroidal magnetic confinement device by

$$W^{theo} \propto na^4RB^2 \left(\frac{P}{na^4RB^3}\right)^{x_1} \left(\frac{a^3B^4}{n}\right)^{x_2} \left(\frac{1}{na^2}\right)^{x_3} \tag{1}$$

$$= cf\left(n, B, P, a; \boldsymbol{x}\right) \ , \tag{2}$$

where n is the average density, a and R minor and major radius of the torus, B the magnetic field and P the deposited heating power. For later purposes the terms in parentheses in (1) are denoted as $S^{(1)}$, $S^{(2)}$ and $S^{(3)}$ with reference to the first, second and third term, respectively.

The particular values of x_1, x_2, x_3 specify the plasma kinetic model as shown in table I and may be represented by a vector $\boldsymbol{x}$. Depending on the model, $\boldsymbol{x}$ consists of either one (M_1: $\boldsymbol{x} = (x, 0, 0)$), two ($M_2$: $\boldsymbol{x} = (x, y, 0)$, M_3: $\boldsymbol{x} = (x, 0, z)$) or three ($M_4$: $\boldsymbol{x} = (x, y, z)$) elements. The number of degrees of freedom X_{dof} (fourth column in table I) therefore varies between one and three. The more complex a model becomes, the higher is the number of degrees of freedom, e.g. incorporating collisions in the low-β model increases X_{dof} from one to two. So if there is evidence in the data that collisions in a low-β plasma should be taken into account, model M_2 should be more likely than M_1. However the price to be payed is the use of a more complex model. So there exists a tradeoff between the goodness of fit (with more degrees of freedom it is easier to fit the data more closely) and the complexity of the theory we have to use. The principle that the simplest theory should be used which is still in accordance with the data is termed Occam's Razor and it is inherently included in

Bayesian probability theory. We will come back to this further down.

Since we concentrate in this paper on data from a single machine, variables which are constant within the examined data set (e.g. R) are absorbed in c. The function $f(n, B, P, a; \boldsymbol{x})$ now consists only of the prefactor na^4B^2 and of the one to three dimensionless terms in parentheses (depending on the model). One may wonder, whether e.g. the inverse of the line averaged density $1/na^2$ (last term in parentheses in (1)) has in fact the dimension [m]. Recall however, that factors consisting of fundamental constants have been omitted (in this particular case $\epsilon_0 m_e c^2/e^2$ with unit [1/m] is missing).

In order to construct a general form-free energy confinement scaling function we want to follow an idea already proposed by Connor and Taylor [4]. It consists of expressing the energy content not as one single term like in (2) but as a series of such dimensionally exact terms, which would still show the same invariances under similarity transformation as the basic plasma equations for a certain model. Consider a set of measurements of the plasma energy content for N different values of the experimental input variables (n, B, P, a). The theoretical prediction for the energy content may be represented by an N-dimensional vector $\boldsymbol{W^{theo}}$ where the i-th component corresponds to the i-th measurement. It may be described for all plasma models by

$$\boldsymbol{W^{theo}} = \sum_{k=1}^{N} c_k \boldsymbol{f}\left(\boldsymbol{x}_k\right) \ . \tag{3}$$

For each term k of the sum the vector $\boldsymbol{x}_k$ of the expansion function $\boldsymbol{f}$ (same vector notation as in $\boldsymbol{W^{theo}}$, with $f_i(\mathrm{x}_k) = f\left(n_i, B_i, P_i, a_i; \mathrm{x}_k\right)$) and the expansion coefficient c_k have to be properly chosen. This means, e.g., for model 2 three adjustable parameters for every expansion order: c_k, x_k, y_k. The functions (or rather vectors) in (3) are linearly independent for any two sets of exponents $\boldsymbol{x}_k$ and $\boldsymbol{x}_l$ if at least one of the exponents in the two sets differs. Therefore, in mathematical terms equation (3) is nothing but the expansion in a basis which is dimensionally exact. In general, N such linearly independent vectors form a complete basis in the N-dimensional data space and would allow a pointwise reconstruction of the data. But this is neither desirable, nor with respect to physics correct, since the corresponding vector of measured energy contents $\boldsymbol{W^{exp}}$ is corrupted by noise. What we really want is an expansion, truncated at some appropriate upper limit E and describing the physics, while the residual $N-E$ terms in the expansion (3) fit only noise:

$$\boldsymbol{W^{theo}} = \sum_{k=1}^{E} c_k \boldsymbol{f}\left(\boldsymbol{x}_k\right) = \mathbf{F}\boldsymbol{c} \; . \tag{4}$$

The expression on the right hand side is the mathematical notation for multiplying the $N \times E$ matrix $\mathbf{F}$ (with the E vectors $\boldsymbol{f}\left(\boldsymbol{x}_k\right)$ as columns) by the vector $\boldsymbol{c}$ (with elements c_k). In our case, due to the noise in the data, a few vectors will be shown already to be sufficient to describe the data. E.g., in model 2 for expansion order $E=3$, the theoretical prediction for energy content of measurement i is given by

$$\begin{aligned} W_i^{theo} &= \sum_{k=1}^{E=3} c_k f_i\left(\boldsymbol{x}_k\right) = c_1 f_i\left(\boldsymbol{x}_1\right) + c_2 f_i\left(\boldsymbol{x}_3\right) + c_3 f_i\left(\boldsymbol{x}_3\right) \\ &= c_1 n_i a_i^4 B_i^2 \left(\frac{P_i}{n_i a_i^4 B_i^3}\right)^{x_1} \left(\frac{a_i^3 B_i^4}{n_i}\right)^{y_1} \\ &+ c_2 n_i a_i^4 B_i^2 \left(\frac{P_i}{n_i a_i^4 B_i^3}\right)^{x_2} \left(\frac{a_i^3 B_i^4}{n_i}\right)^{y_2} \\ &+ c_3 n_i a_i^4 B_i^2 \left(\frac{P_i}{n_i a_i^4 B_i^3}\right)^{x_3} \left(\frac{a_i^3 B_i^4}{n_i}\right)^{y_3} \end{aligned} \tag{5}$$

Here we meet again the principle of Occam's Razor. Though a higher expansion order will describe the data progressively better, the ultimate fitting of noise does not make sense. So what we are looking for is the determination of the appropriate expansion order E in the plasma physics model which describes the data best. A comprehensive answer to these questions is provided by Bayesian probability theory.

III. BAYESIAN INFERENCE

Bayesian probability theory [12,13] rests on two basic rules, the sum rule and the product rule. The sum rule states that the probability of either of two mutually exclusive propositions A, B being true is given by

$$p(A+B|I) = p(A|I) + p(B|I) \; . \tag{6}$$

The plus sign in the argument stands for a logical "or" and I denotes the background information which gave rise to the propositions A and B. Eq. (6) may be generalized to more than two alternatives. If we have a set of mutually exclusive and exhaustive propositions, e.g. some models M_i, we obtain the normalization rule

$$p(\sum_i M_i|I) = \sum_i p(M_i|I) = 1 \ . \tag{7}$$

The product rule

$$p(M_i, D|I) = p(M_i|I)p(D|M_i, I) \tag{8}$$

relates the probability of M_i and D being true to the probability of M_i being true times the probability of D being true given that M_i is true, where D shall be some data acquired by an experiment. Due to symmetry in the arguments M_i and D the probability $p(M_i, D|I)$ may be expanded alternatively into

$$p(M_i, D|I) = p(D|I)p(M_i|D, I) \ , \tag{9}$$

and the combination of (8) and (9) constitutes Bayes' theorem

$$p(M_i|D, I) = \frac{p(M_i|I)p(D|M_i, I)}{p(D|I)} \ . \tag{10}$$

Bayes' theorem is an instruction how to learn: The prior probability $p(M_i|I)$ for the set of hypotheses M_i, i.e. what we know before carrying out an experiment, is updated to the posterior probability $p(M_i|D, I)$, i.e. what we learned from the experiment, by multiplication with $p(D|M_i, I)$. The latter probability function describes the theory of the experiment and is called the likelihood function. $p(D|I)$, the global likelihood for the entire class of models, which is sometimes also called evidence, is not independent of the other probabilities and is obtained via the sum and product rule

$$p(D|I) = \sum_i p(M_i, D|I) = \sum_i p(M_i|I)p(D|M_i, I) \ . \tag{11}$$

In this paper we shall deal with hypotheses which are specified by the possible values of continuously varying parameters. In such a case probabilities become probability densities.

If for example a hypothesis is the model M_1 which in turn is specified in terms of a set of E parameters $\{x_k\} = \boldsymbol{x}$, then $p(\boldsymbol{x}|M_1, I)d^E x$ is the probability that the true values of $\boldsymbol{x}$ lie in the range $[\{x_i, x_i + dx_i\}\ \forall\ i]$ in the Euclidean space. The product and sum rule hold for probability densities as well and upon replacing the discrete sum by an integral we have in analogy to (11)

$$p(D|M_1, I) = \int d^E x\ p(\boldsymbol{x}|M_1, I)p(D|\boldsymbol{x}, M_1, I)\ . \tag{12}$$

Eq. (12) is called marginalization (over $\boldsymbol{x}$) and is a very important ingredient of Bayesian probability theory. It will be applied repeatedly in the body of this paper.

The marginal likelihood of the data, assuming a certain model M_j is true, is the key quantity to perform the comparison of different models given a set of data D. From Bayes' theorem we have

$$p(M_j|D, I) = \frac{p(M_j|I)p(D|M_j, I)}{p(D|I)}\ . \tag{13}$$

In order to compare two models M_j and M_k given the set of data D we take the ratio of their posterior probabilities and obtain

$$\frac{p(M_j|D, I)}{p(M_k|D, I)} = \frac{p(M_j|I)}{p(M_k|I)} \frac{p(D|M_j, I)}{p(D|M_k, I)} = O_{jk}\ . \tag{14}$$

This quantity (14) is usually called the posterior odds ratio. Note that $p(D|I)$ has dropped out. In the usual case that we have no reason to favor the prior probability $p(M_j|I)$ over $p(M_k|I)$ (in fact, this is the driving force for carrying out a sensible experiment) the posterior odds simplifies to the ratio of the marginal likelihoods given either model M_j or model M_k. Our first goal will be to determine which of the four kinetic CT models is most probable given a set of measured plasma energy contents W_i^{exp} (the actual data which will further on stand for the previously used D) as a function of the machine variables n, B, P, a. The odds ratios from (14) will then answer the question whether transport for a certain data set of the W7-AS is best described in terms of the collisionless or collisional, low- or high-β kinetic models.

IV. CALCULATION

A. The likelihood function

We want to describe the experimental data $\boldsymbol{W^{exp}}$ corrupted by noise by a theoretical scaling $\boldsymbol{W^{theo}}$:

$$W_i^{exp} = W_i^{theo} + \varepsilon_i \ . \qquad (15)$$

$\boldsymbol{\varepsilon}$ is the noise, where $< \varepsilon_i >= 0$ and $< \varepsilon_i^2 >= \sigma_i^2 \cdot 1/\omega$. $\boldsymbol{\sigma}$ is the vector of experimental uncertainties associated with the measurement $\boldsymbol{W^{exp}}$. The uncertainty in the energy content $\boldsymbol{W^{exp}}$ contains the direct contribution from the diamagnetic measurement as well as indirect contributions from the finite precision in the input variables (n, B, P, a). Both contributions have been estimated to the best of our knowledge (see appendix A). However, to allow for possible deviations from the true errors we introduce an overall correction factor ω ($\omega = 1$ for exactly specified variances σ_i^2). With this approach we assume that we already describe the correct qualitative behavior of the error for different values of $\boldsymbol{W^{exp}}$ but may fail on the quantitative level by an overall factor ω. The Bayesian analysis yields $\omega^{-0.5} \approx 0.8$ for the most probable model and the optimum expansion order which indicates an overestimation of $\boldsymbol{\sigma}$ of merely 20% (the explanation how this quantity is obtained will be given in chapter IV E).

The mere fact that the experimental error of the data $\boldsymbol{W^{exp}}$ represents testable information leads by virtue of the principle of maximum entropy [14] to the normalized likelihood function which constitutes our starting point for the forthcoming analysis.

$$p\ (\boldsymbol{W^{exp}}|\omega, \boldsymbol{c}, \boldsymbol{x}, \boldsymbol{\sigma}, E, M_j, I) = \left(\frac{\omega}{2\pi}\right)^{\frac{N}{2}} \prod_i^N \sigma_i^{-1}$$
$$\times \exp\left\{-\frac{\omega}{2}\sum_i \left(W_i^{exp} - \sum_k^E c_k f_i(\boldsymbol{x}_k)\right)^2 / \sigma_i^2 \right\} \ . \qquad (16)$$

M_j denotes the plasma kinetic model and determines the vectors $\boldsymbol{x}_k$ used to generate the expansion vectors $\boldsymbol{f}\,(\boldsymbol{x}_k)$. $\boldsymbol{c}$ is as before the vector of the expansion coefficients and E the expansion order. Of course E, $\boldsymbol{c}$ and $\boldsymbol{f}(\boldsymbol{x}_k)$ will depend on model M_j. In order to simplify notation we shall add the model index j only in cases where ambiguity would arise.

If we define $\tilde{W}_i^{exp} = W_i^{exp}/\sigma_i$ and $\tilde{f}_i(\boldsymbol{x}_k) = f_i(\boldsymbol{x}_k)/\sigma_i$, and with the help of the matrix notation as on the r.h.s. in (4), the sum in the exponent of (16) may be written as

$$\chi^2 = \left(\tilde{\boldsymbol{W}}^{\boldsymbol{exp}} - \tilde{\mathbf{F}}\boldsymbol{c}\right)^T \left(\tilde{\boldsymbol{W}}^{\boldsymbol{exp}} - \tilde{\mathbf{F}}\boldsymbol{c}\right) \ . \tag{17}$$

Again $\tilde{\mathbf{F}}$ is a matrix with N rows and E column vectors $\tilde{\boldsymbol{f}}(\boldsymbol{x}_k)$. The maximum of the likelihood occurs at the minimum of (17) and is given by

$$\frac{\partial}{\partial \boldsymbol{c}^T} \left\{ \tilde{\boldsymbol{W}}^{\boldsymbol{exp}^T} \tilde{\boldsymbol{W}}^{\boldsymbol{exp}} - 2\boldsymbol{c}^T \tilde{\mathbf{F}}^T \tilde{\boldsymbol{W}}^{\boldsymbol{exp}} + \boldsymbol{c}^T \tilde{\mathbf{F}}^T \tilde{\mathbf{F}} \boldsymbol{c} \right\} = 0 \ . \tag{18}$$

The solution of (18) is straight forward and yields

$$\boldsymbol{c}_{\boldsymbol{ML}} = \left(\tilde{\mathbf{F}}^T \tilde{\mathbf{F}}\right)^{-1} \tilde{\mathbf{F}}^T \tilde{\boldsymbol{W}}^{\boldsymbol{exp}} \ . \tag{19}$$

The index ML stands for maximum likelihood. A numerical problem may arise in the calculation of the inverse matrix in (19) if some vectors in $\tilde{\mathbf{F}}$ are becoming collinear. The most efficient and stable algorithm to cope with this issue is the singular value decomposition (SVD). We employ the SVD in order to check just on the collinearity and to avoid numerical problems in dealing with column vectors of $\tilde{\mathbf{F}}$ which are collinear within machine accuracy. Let $\mathbf{F}'$ be the $N \times E$ matrix whose column vectors are normalized with respect to their maximum value so that $\tilde{\mathbf{F}} = \mathbf{F}'\mathbf{T}$ with $\mathbf{T}$ being an $E \times E$ diagonal matrix containing the values $(\mathbf{T})_{kk} = \mathrm{Max}_i \ f_i(\boldsymbol{x}_k)$.

$$\mathbf{F}' = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^T \tag{20}$$

is called the singular value decomposition of $\mathbf{F}'$. While $\boldsymbol{\Delta}$ and $\mathbf{V}^T$ are square matrices of dimension E, $\mathbf{U}$ is an $N \times E$ matrix with orthogonal column vectors. The diagonal matrix $\boldsymbol{\Delta}$ carries the scales. If elements $(\boldsymbol{\Delta})_{kk}$ are smaller than the computing accuracy indicating that vectors in $\mathbf{F}'$ are collinear, the corresponding elements in $\boldsymbol{\Delta}^{-1}$ are zeroed yielding $\tilde{\boldsymbol{\Delta}}^{-1}$ which is equivalent to eliminating linearly dependent vectors in the model equation (4). Use of (20) in (19) gives finally

$$\boldsymbol{c_{ML}} = \mathbf{T}^{-1}\mathbf{V}\tilde{\boldsymbol{\Delta}}^{-1}\mathbf{U}^T\tilde{\boldsymbol{W}}^{\boldsymbol{exp}} \ . \tag{21}$$

We are looking for the probability of a model given the data $\boldsymbol{W}^{exp}$. The odds ratio (14) reads

$$\frac{p(M_j|\boldsymbol{W^{exp}},\boldsymbol{\sigma},I)}{p(M_k|\boldsymbol{W^{exp}},\boldsymbol{\sigma},I)} = \frac{p(M_j|\boldsymbol{\sigma},I)}{p(M_k|\boldsymbol{\sigma},I)}\frac{p(\boldsymbol{W^{exp}}|M_j,\boldsymbol{\sigma},I)}{p(\boldsymbol{W^{exp}}|M_k,\boldsymbol{\sigma},I)} \ . \tag{22}$$

Compared to (14) we exchanged the common variable D with the data vector $\boldsymbol{W^{exp}}$ and inserted the further conditioning on our error estimate $\boldsymbol{\sigma}$. The first ratio on the r.h.s., the so-called prior odds, is set to unity since we do not want to favor one model over another. An expert may rule out certain models as improper for the plasma description in W7-AS and give them a lower model probability $p(M_j|\boldsymbol{\sigma},I)$, but since we want to handle this problem on the most common footing we do not intend to do so. We are left with the determination of the second ratio, the so-called Bayes factor, which consists of marginal likelihoods conditional on the models and the error estimate. In the following, by extensive use of the marginalization rule, $p(\boldsymbol{W^{exp}}|M_k,\boldsymbol{\sigma},I)$ is related to the likelihood function (16).

First we marginalize over the expansion order, which is given by a discrete sum over all possible E to obtain

$$p(\boldsymbol{W^{exp}}|M_j,\boldsymbol{\sigma},I) = \sum_E p(E|M_j,\boldsymbol{\sigma},I)p(\boldsymbol{W^{exp}}|E,M_j,\boldsymbol{\sigma},I) \tag{23}$$

Again there is no reason to favor a certain expansion order over some other and therefore $p(E|M_j,\boldsymbol{\sigma},I) = const$, which leaves us with $p(\boldsymbol{W^{exp}}|E,M_j,\boldsymbol{\sigma},I)$ the so-called prior predictive value for an expansion order E and a model M_j. The next three sections exclusively deal with the calculation of this quantity.

B. Invariant measure

The remaining marginal likelihood $p(\boldsymbol{W^{exp}}|E,M_j,\boldsymbol{\sigma},I)$ is obtained by marginalization over ω, $\boldsymbol{c}$ and $\boldsymbol{x}$. These are continuous quantities and have to be integrated:

$$p(\boldsymbol{W}^{exp}|E,M_j,\boldsymbol{\sigma},I)=\int p(\boldsymbol{W}^{exp}|\omega,\boldsymbol{c},\boldsymbol{x},E,M_j,\boldsymbol{\sigma},I) \cdot p(\omega,\boldsymbol{x},\boldsymbol{c}|E,M_j,\boldsymbol{\sigma},I)\;d\mu(\omega,\boldsymbol{x},\boldsymbol{c})\;. \tag{24}$$

The likelihood function shows up in (24), together with the so-called prior function $p(\omega,\boldsymbol{x},\boldsymbol{c}|E,M_j,\boldsymbol{\sigma},I)$ and the Riemannian metric of the problem $d\mu(\omega,\boldsymbol{x},\boldsymbol{c}) = \mu(\omega,\boldsymbol{x},\boldsymbol{c})d\omega d\boldsymbol{x}d\boldsymbol{c}$, which constitutes an invariant measure in the multidimensional marginalization integral. The latter stems from the requirement that the description of a problem should be the same regardless in which coordinate system it is considered [15]. It is defined by

$$\mu(\omega,\boldsymbol{x},\boldsymbol{c})=\sqrt{\det\left[\mathbf{g}\right]}\;. \tag{25}$$

$|\mathbf{g}|$ is the determinant of the Fisher information matrix. The elements are given by the second derivative of the log-likelihood function with respect to the parameters $\boldsymbol{\theta}=(\omega,\boldsymbol{x},\boldsymbol{c})$ and read [1]

$$g_{ij}=-\left\langle\frac{\partial^2\ln p(\boldsymbol{W}^{exp}|\boldsymbol{\theta},E,M_j,\boldsymbol{\sigma},I)}{\partial\theta_i\partial\theta_j}\right\rangle\;. \tag{26}$$

Performing the derivatives and calculation of the expectation value (26) yields the invariant measure for expansion order E and a model with certain number of degrees of freedom X_{dof} (as shown in the fourth column of table I)

$$\mu(\omega,\boldsymbol{x},\boldsymbol{c})=\sqrt{\frac{N}{2}}\;\omega^{\frac{E\cdot X_{dof}}{2}-1}|\mathbf{C}|^{X_{dof}}|\tilde{\boldsymbol{\Delta}}|^{X_{dof}+1}\Xi(\boldsymbol{x})\;, \tag{27}$$

with

$$\Xi(\boldsymbol{x})=\left|\left\{\left[\mathbf{U}^T\mathbf{L}^{(i)}\left(\mathbb{1}-\mathbf{U}\mathbf{U}^T\right)\mathbf{L}^{(i')}\mathbf{U}\right]_{kk'}\right\}\right|^{X_{dof}}\;. \tag{28}$$

The notation is as follows: $\mathbf{C}$ is an $E\times E$ diagonal matrix consisting of the expansion coefficients c_k only. $\tilde{\boldsymbol{\Delta}}$ and $\mathbf{U}$ stem from the singular value decomposition (20). The matrix

---

[1]The brackets in (26) denote the expectation value which is defined for an arbitrary quantity A by $\langle A\rangle=\int A\;p(\boldsymbol{W}^{\boldsymbol{exp}}|\omega,\boldsymbol{c},\boldsymbol{x},E,M_j,\boldsymbol{\sigma},I)\;d\boldsymbol{W}^{\boldsymbol{exp}}/\int p(\boldsymbol{W}^{\boldsymbol{exp}}|\omega,\boldsymbol{c},\boldsymbol{x},E,M_j,\boldsymbol{\sigma},I)\;d\boldsymbol{W}^{\boldsymbol{exp}}$.

$\mathbf{L^{(i)}}$ is an $N \times N$ diagonal matrix consisting of the logarithms of the i-th term in parentheses in (1), $\mathbf{L^{(i)}} = \mathbf{diag}\left(\ln S_\nu^{(i)}\right)$. The column (row) element of the complete $(E \cdot X_{dof}) \times (E \cdot X_{dof})$ matrix in the curly brackets in (28) is obtained by running over all possible i (i') for each k (k') of the expansion in (3).

C. Prior probabilities

The prior function in (24) decomposes into

$$p(\omega, \boldsymbol{x}, \boldsymbol{c}|E, M_j, \boldsymbol{\sigma}, I) = p(\omega, \boldsymbol{x}|E, M_j, \boldsymbol{\sigma}, I)p(\boldsymbol{c}|\boldsymbol{x}, E, M_j, \boldsymbol{\sigma}, I) \ . \tag{29}$$

Since we have no information about $\boldsymbol{x}$ and ω, a constant value is assigned to $p(\omega, \boldsymbol{x}|E, M_j, \boldsymbol{\sigma}, I) = const.$ As far as the coefficients $\boldsymbol{c}$ are concerned we are not totally ignorant. First let us have a look at χ^2 in the likelihood function (16). Its minimum value is

$$\chi^2_{min} = \tilde{\boldsymbol{W}}^{\boldsymbol{exp}^T} \tilde{\boldsymbol{W}}^{\boldsymbol{exp}} - \boldsymbol{c_{ML}}^T \tilde{\mathbf{F}}^T \tilde{\mathbf{F}} \boldsymbol{c_{ML}} \ . \tag{30}$$

According to (17) χ^2_{min} is a positive definite function. Consequently it cannot drop below zero and we have

$$\tilde{\boldsymbol{W}}^{\boldsymbol{exp}^T} \tilde{\boldsymbol{W}}^{\boldsymbol{exp}} \geq \boldsymbol{c_{ML}}^T \tilde{\mathbf{F}}^T \tilde{\mathbf{F}} \boldsymbol{c_{ML}} \ . \tag{31}$$

While (31) is valid for $\boldsymbol{c} = \boldsymbol{c_{ML}}$ only, we can extend its form to an estimation for arbitrary coefficients $\boldsymbol{c}$. For those the right hand side of (31) has to allow for the uncertainties in the data. In order to establish a new upper limit we add $\boldsymbol{\Sigma}^T \boldsymbol{\Sigma}$ to the left hand side of (31)

$$\tilde{\boldsymbol{W}}^{\boldsymbol{exp}^T} \tilde{\boldsymbol{W}}^{\boldsymbol{exp}} + \boldsymbol{\Sigma}^T \boldsymbol{\Sigma} \geq \boldsymbol{c}^T \tilde{\mathbf{F}}^T \tilde{\mathbf{F}} \boldsymbol{c} \ . \tag{32}$$

A conservative approximation of $\boldsymbol{\Sigma}$ is to assume that the deviation of the expansion from the measured data shall not be larger than the data value itself, which means $\boldsymbol{\Sigma} = \tilde{\boldsymbol{W}}^{\boldsymbol{exp}}$ and results in

$$2\tilde{\boldsymbol{W}}^{\boldsymbol{exp}^T}\tilde{\boldsymbol{W}}^{\boldsymbol{exp}} \geq \boldsymbol{c}^T\tilde{\mathbf{F}}^T\tilde{\mathbf{F}}\boldsymbol{c} \ . \tag{33}$$

We name this Bessel prior because (31) is nothing but the Bessel inequality if $\tilde{\mathbf{F}}$ were a complete basis. This just imposes an upper boundary on the choice of possible coefficients and is equivalent to an entropic prior proposed by Rodriguez [15], where the hyper-parameter α has been marginalized by using the Bretthorst approximation that the prior may be taken out of the integral at the maximum of the likelihood [12]. We write the Bessel prior as a θ-function which allows only those values for $\boldsymbol{c}$ which fulfill (31)

$$p(\boldsymbol{c}|\boldsymbol{x}, E, M_j, \boldsymbol{\sigma}, I) \propto \theta\left(\frac{\boldsymbol{c}^T\tilde{\mathbf{F}}^T\tilde{\mathbf{F}}\boldsymbol{c}}{2\tilde{\boldsymbol{W}}^{\boldsymbol{exp}^T}\tilde{\boldsymbol{W}}^{\boldsymbol{exp}}} \leq 1\right) \ . \tag{34}$$

After stating the complete prior function, the next task is the evaluation of its normalization constant Z given by

$$\begin{aligned} Z &= \int p(\omega, \boldsymbol{x}, \boldsymbol{c}|E, M_j, \boldsymbol{\sigma}, I)\ d\mu(\omega, \boldsymbol{x}, \boldsymbol{c}) \\ &= \sqrt{\frac{N}{2}} \int d\boldsymbol{x}\ \Xi(\boldsymbol{x}) \cdot Z_{Bessel} \cdot Z_\omega \ . \end{aligned} \tag{35}$$

with

$$Z_{Bessel} = \int_{-\infty}^{\infty} d\boldsymbol{c}\ |\tilde{\boldsymbol{\Delta}}|^{X_{dof}+1}|\mathbf{C}|^{X_{dof}}\theta\left(\frac{\boldsymbol{c}^T\tilde{\mathbf{F}}^T\tilde{\mathbf{F}}\boldsymbol{c}}{2\tilde{\boldsymbol{W}}^{\boldsymbol{exp}^T}\tilde{\boldsymbol{W}}^{\boldsymbol{exp}}} \leq 1\right) \tag{36}$$

and

$$Z_\omega = \int_{\omega_0}^{\omega_1} d\omega\ \omega^{\frac{E\cdot X_{dof}}{2}-1} \ . \tag{37}$$

For the contribution from the ω integration Z_ω (37) we employ the conservative approximation that our estimation of the experimental error is correct in quantitative respect at least by a factor of two. Since the error enters the problem in a quadratic manner this means that the overall correction factor is something between $\omega_0 = 1/2^2$ and $\omega_1 = 2^2$. Inserting these values in the integration limits of (35) results in

$$Z_\omega = \frac{2^{EX_{dof}+1}}{EX_{dof}}\left(1 - 4^{-EX_{dof}}\right) \ . \tag{38}$$

The integral in $\boldsymbol{c}$ (36) covers $-\infty$ to ∞ and the upper limit established by the Bessel prior becomes effective. The term $\boldsymbol{c}^T\tilde{\mathbf{F}}^T\tilde{\mathbf{F}}\boldsymbol{c}$ in the θ function constitutes an ellipsoidal sphere in phase space. In order to calculate the volume of this hyper sphere we perform a transformation of the principle axes and require $|\tilde{\mathbf{F}}^T\tilde{\mathbf{F}}| \approx |\tilde{\boldsymbol{\Delta}}|^2$. With this approximation we get from the integration over $\boldsymbol{c}$

$$Z_{Bessel} = \left(2\tilde{\boldsymbol{W}}^{\boldsymbol{exp}^T}\tilde{\boldsymbol{W}}^{\boldsymbol{exp}}\right)^{-\frac{E(X_{dof}+1)}{2}} \frac{\Gamma\left(\frac{E(X_{dof}+1)}{2}\right)}{\Gamma\left(\frac{X_{dof}+1}{2}\right)^E} . \tag{39}$$

The integration over $\boldsymbol{x}$ has to be postponed to the final integration of the posterior function.

Knowing the normalization of the prior function we are finally in the position to perform the marginalizations over $\boldsymbol{c}$ and ω in (24).

$$p(\boldsymbol{W}^{exp}|E, M_j, \boldsymbol{\sigma}, I) = \frac{1}{\prod \sigma_i}\pi^{-\frac{N-E}{2}}2^{\frac{EX_{dof}}{2}}\frac{1}{Z_\omega Z_{Bessel}}\Gamma\left(\frac{N-E(X_{dof}+1)}{2}\right)$$
$$\cdot \int d\boldsymbol{x}\ \xi(\boldsymbol{x})\ \frac{|\mathbf{C^{ML}}|^{X_{dof}}|\tilde{\boldsymbol{\Delta}}|^{X_{dof}}}{\left(\tilde{\boldsymbol{W}}^{\boldsymbol{exp}^T}\tilde{\boldsymbol{W}}^{\boldsymbol{exp}} - \boldsymbol{c_{ML}}^T\tilde{\mathbf{F}}^T\tilde{\mathbf{F}}\boldsymbol{c_{ML}}\right)^{\frac{N-E(X_{dof}+1)}{2}}} , \tag{40}$$

with $\xi(\boldsymbol{x}) = \Xi(\boldsymbol{x})/\int d\boldsymbol{x'}\ \Xi(\boldsymbol{x'})$. $\mathbf{C^{ML}}$ is still the $E \times E$ diagonal matrix from (27) but now with the maximum likelihood values of (21) as elements.

D. Prior predictive value

While the $\boldsymbol{c}$ and ω integrations are done analytically the remaining $\boldsymbol{x}$ integration is carried out numerically by computing the prior predictive value [16] with Markov Chain Monte Carlo (MCMC). This is given by

$$p(\boldsymbol{W}^{exp}|E, M_j, \boldsymbol{\sigma}, I) = \int \Lambda(\boldsymbol{x})\Pi(\boldsymbol{x})\ d\boldsymbol{x}\ . \tag{41}$$

In (41) we combined the factors in the integral (40) to new functions Λ and Π, where

$$\Lambda(\boldsymbol{x}) \propto \left(\tilde{\boldsymbol{W}}^{\boldsymbol{exp}^T}\tilde{\boldsymbol{W}}^{\boldsymbol{exp}} - \boldsymbol{c_{ML}}^T\tilde{\mathbf{F}}^T\tilde{\mathbf{F}}\boldsymbol{c_{ML}}\right)^{-\frac{N-E(X_{dof}+1)}{2}} \tag{42}$$

contains the normalization constants and the term originating from the likelihood, while

$$\Pi(\boldsymbol{x}) = \frac{1}{Z_\Pi}\xi(\boldsymbol{x})\ |\mathbf{C}^{\mathbf{ML}}|^{X_{dof}}|\tilde{\mathbf{\Delta}}|^{X_{dof}} \tag{43}$$

comprises the prior and the terms stemming from the Riemannian measure. Z_Π is given by the requirement $\int \Pi(\boldsymbol{x})\ d\boldsymbol{x} = 1$. The only way to evaluate (41) is to use $\Pi(\boldsymbol{x})$ as a probability density. Since the likelihood originating term $\Lambda(\boldsymbol{x})$ is generally much more structured than $\Pi(\boldsymbol{x})$ the variance will be large and extremely long Markov chains are needed to obtain the desired accuracy. However, one can solve this problem by defining a new function

$$\Omega(\nu) = \int \Lambda^\nu(\boldsymbol{x})\Pi(\boldsymbol{x})\ d\boldsymbol{x} \tag{44}$$

with $\Omega(\nu = 0) = 1$. $\Omega(\nu = 1)$ is the sought after quantity in (41). The derivative with respect to ν gives

$$\begin{aligned}\frac{\partial \ln \Omega(\nu)}{\partial \nu} &= \int \ln \Lambda(\boldsymbol{x})\rho_\nu(\boldsymbol{x})\ d\boldsymbol{x} \\ &= \langle \ln \Lambda(\boldsymbol{x})\rangle_\nu\ ,\end{aligned} \tag{45}$$

with

$$\rho_\nu(\boldsymbol{x}) = \frac{\Lambda^\nu(\boldsymbol{x})\Pi(\boldsymbol{x})}{\int \Lambda^\nu(\boldsymbol{x}')\Pi(\boldsymbol{x}')\ d\boldsymbol{x}'} \tag{46}$$

as the new sampling density. If we integrate both sides of (45),

$$\int_0^1 \langle \ln \Lambda(\boldsymbol{x})\rangle_\nu\ d\nu = \int_0^1 \frac{\partial \ln \Omega(\nu)}{\partial \nu}\ d\nu \tag{47}$$

$$= \ln \Omega(\nu = 1) - \ln \Omega(\nu = 0) \tag{48}$$

$$= \ln p(\boldsymbol{W}^{exp}|E, M_j, \boldsymbol{\sigma}, I)\ , \tag{49}$$

we are back where we started in (41). To obtain the marginal likelihood one therefore has to calculate the integral on the l.h.s. in (47) where the expectation value $\langle \ln \Lambda(\boldsymbol{x})\rangle_\nu$ is accessible by MCMC techniques. The proof that this is feasible may be found in [16].

Now everything is ready to perform the model comparison. For a test of our analysis with a simulated data set please see appendix B.

E. Expectation values

Within the Bayesian approach it is not only possible to perform model comparisons but also to ask for the expectation value of quantities like the energy content given a certain model. Let $\boldsymbol{v}$ be the vector of some independent new variables with $\boldsymbol{v} = (\hat{n}, \hat{B}, \hat{P}, \hat{a})$. The expectation value of the energy content for this given set of $\boldsymbol{v}$ is then

$$\langle W \rangle_v = \frac{\int dW \; W \; p(W|\boldsymbol{W^{exp}}, \boldsymbol{v}, M_j, \boldsymbol{\sigma}, I)}{\int dW \; p(W|\boldsymbol{W^{exp}}, \boldsymbol{v}, M_j, \boldsymbol{\sigma}, I)} \; . \tag{50}$$

To obtain $p(W|\boldsymbol{W^{exp}}, \boldsymbol{v}, M_j, \boldsymbol{\sigma}, I)$ we have again to marginalize over all quantities entering the problem

$$\begin{aligned} p(W|\boldsymbol{W^{exp}}, \boldsymbol{v}, M_j, \boldsymbol{\sigma}, I) &= \sum_E \int d\boldsymbol{x} \; d\boldsymbol{c} \; d\omega \; \mu(\omega, \boldsymbol{x}, \boldsymbol{c}) p(E, \boldsymbol{x}, \boldsymbol{c}, \omega, W|\boldsymbol{W^{exp}}, \boldsymbol{v}, M_j, \boldsymbol{\sigma}, I) \\ &= \sum_E \int d\boldsymbol{x} \; d\boldsymbol{c} \; d\omega \; \mu(\omega, \boldsymbol{x}, \boldsymbol{c}) p(W|E, \boldsymbol{x}, \boldsymbol{c}, \boldsymbol{v}, M_j, \boldsymbol{\sigma}, I) p(E, \boldsymbol{x}, \boldsymbol{c}, \omega|\boldsymbol{W^{exp}}, M_j, \boldsymbol{\sigma}, I) \; . \end{aligned} \tag{51}$$

Knowing the vector $\boldsymbol{x}$, coefficients $\boldsymbol{c}$, expansion order E and the variables $\boldsymbol{v}$, the first term in (51) is completely specified by

$$p(W|E, \boldsymbol{x}, \boldsymbol{c}, \boldsymbol{v}, M_j, \boldsymbol{\sigma}, I) = \delta\left(W - \boldsymbol{g}^T(\boldsymbol{x}, \boldsymbol{v}) \cdot \boldsymbol{c}\right) \tag{52}$$

with an E-dimensional vector $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{v})$ with elements $f(\hat{n}, \hat{B}, \hat{P}, \hat{a}; \boldsymbol{x}_k)$. The second term is, with the help of the product rule,

$$p(E, \boldsymbol{x}, \boldsymbol{c}, \omega|\boldsymbol{W^{exp}}, M_j, \boldsymbol{\sigma}, I) = \frac{p(\boldsymbol{W^{exp}}, E, \boldsymbol{x}, \boldsymbol{c}, \omega|M_j, \boldsymbol{\sigma}, I)}{p(\boldsymbol{W^{exp}}|M_j, I)} \; . \tag{53}$$

Inserting (52) and (53) in (51) gives after some straightforward algebra

$$\langle W \rangle_v = \frac{\sum_E p(\boldsymbol{W^{exp}}|E, M_j, \boldsymbol{\sigma}, I) \int d^E\boldsymbol{x} \; \rho(\boldsymbol{x}) \; \boldsymbol{g}^T(\boldsymbol{x}, \boldsymbol{v}) \boldsymbol{c_{ML}}}{\sum_E p(\boldsymbol{W^{exp}}|E, M_j, \boldsymbol{\sigma}, I)} \; . \tag{54}$$

$\rho(\boldsymbol{x})$ is the same probability density as in (46) for $\nu = 1$ and can therefore be obtained from the same MC run already performed for model comparison.

Furthermore we may want to know the confidence interval of this prediction for the energy content. It is given by

$$\sigma_{\langle W \rangle_v} = \sqrt{\langle W^2 \rangle_v - \langle W \rangle_v^2} \; . \tag{55}$$

The expectation value $\langle W^2 \rangle_v$ reads

$$\langle W^2 \rangle_v = \frac{\int dW \;\; W^2 p(W|\boldsymbol{W^{exp}}, \boldsymbol{v}, M_j, \boldsymbol{\sigma}, I)}{\int dW \;\; p(W|\boldsymbol{W^{exp}}, \boldsymbol{v}, M_j, \boldsymbol{\sigma}, I)} \; . \tag{56}$$

Similar algebraic steps as above lead to

$$\langle W^2 \rangle = \frac{\sum_E p(\boldsymbol{W^{exp}}|E, M_j, \boldsymbol{\sigma}, I) \int d^E \boldsymbol{x} \;\, \rho(\boldsymbol{x}) \;\, W^2(\boldsymbol{x})}{\sum_E p(\boldsymbol{W^{exp}}|E, M_j, \boldsymbol{\sigma}, I)} \; , \tag{57}$$

with

$$\begin{aligned} W^2(\boldsymbol{x}) = \mathrm{Tr}\left[\boldsymbol{g}^T(\boldsymbol{x}, \boldsymbol{v}) \left(\tilde{\mathbf{F}}^T \tilde{\mathbf{F}}\right)^{-1} \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{v})\right] & \frac{\boldsymbol{\tilde{W}^{exp}}^T \boldsymbol{\tilde{W}^{exp}} - \boldsymbol{c_{ML}}^T \tilde{\mathbf{F}}^T \tilde{\mathbf{F}} \boldsymbol{c_{ML}}}{N - E - 2} \\ + \boldsymbol{c_{ML}}^T \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{v}) \boldsymbol{g}^T(\boldsymbol{x}, \boldsymbol{v}) \boldsymbol{c_{ML}} \; . & \end{aligned} \tag{58}$$

In chapter IV A we mentioned that it is possible to infer an estimate for the unknown error level ω given a certain model. This is done by evaluating the expectation value $\langle \omega \rangle$ given by

$$\langle \omega \rangle = \frac{\int d\omega \;\, \omega \;\, p(\omega|\boldsymbol{W^{exp}}, M_j, \boldsymbol{\sigma}, I)}{\int d\omega \;\, p(\omega|\boldsymbol{W^{exp}}, M_j, \boldsymbol{\sigma}, I)} \tag{59}$$

and results in

$$\langle \omega \rangle = \frac{\sum_E p(\boldsymbol{W^{exp}}|E, M_j, \boldsymbol{\sigma}, I) \int d^E \boldsymbol{x} \;\, \rho(\boldsymbol{x}) \;\, \frac{N-E}{\boldsymbol{\tilde{W}^{exp}}^T \boldsymbol{\tilde{W}^{exp}} - \boldsymbol{c_{ML}}^T \tilde{\mathbf{F}}^T \tilde{\mathbf{F}} \boldsymbol{c_{ML}}}}{\sum_E p(\boldsymbol{W^{exp}}|E, M_j, \boldsymbol{\sigma}, I)} \; . \tag{60}$$

For all the four models we get values of order one, in particular the winning collisional low-β model has $\langle \omega \rangle = 1.5$. Although the measurement error is therefore a little bit overestimated (factor $\langle \omega \rangle^{-0.5} \approx 0.8$) this still means that the error estimation in appendix A is already a reasonable quantitative approximation to the true error $\boldsymbol{\sigma}$.

V. RESULTS

The data which we have used in our calculations are the 153 $\iota \approx 1/3$ W7-AS data documented in the international stellarator energy confinement data base [3] (ι: rotational transform). The data range is shown in table II. We have selected the $\iota \approx 1/3$ data only since the single variable scans which we present further down have been performed at this value of the rotational transform. Although this limitation establishes no fundamental restriction to the power of our approach it has the further advantage that we get rid of an additional, dimensionless and ill defined variable which would otherwise have to be considered in the ansatz (3).

The $N = 153$ data originate from both ECRH and NBI heating. While most of the data was generated with ECRH ($N_{ECRH} = 110$), the data available for NBI heating ($N_{NBI} = 43$) removes the correlation between density and magnetic field which would be very strong for ECRH only [17].

A. Model comparison

Having determined $p(M_j|\boldsymbol{W}^{\boldsymbol{exp}}, \boldsymbol{\sigma}, I)$ for each of the four CT models M_j we finally return to the odds ratio (22). With the help of the normalization rule (7), $\sum_j p(M_j|\boldsymbol{W}^{\boldsymbol{exp}}, \boldsymbol{\sigma}, I) = 1$, odds ratios are converted back to probabilities.

The resulting model probabilities are depicted in the last but one column of table I. We see that the $\iota \approx 1/3$ W7-AS data are best described by the collisional low beta Connor Taylor model. The high beta models follow in second and third place with much lower probability. The collisionless low beta model is clearly inappropriate to describe the transport physics in W7-AS.

Our result is well in agreement with the experimental outcome from W7-AS. While the measured collision frequencies suggest a collisional model, the absence of strong magnetic turbulence in the plasma edge and the very small value of the toroidal beta (0.14% for the examined data set) point to the low beta case.

In considering the four Connor Taylor models we made the implicit assumption of a fully ionized pure plasma which has no interactions with the wall affecting the plasma. Of course these are idealized conditions which are not realized in a real world device like W7-AS.

However, the Connor Taylor models were used only to get a dimensionally exact form of the scaling relation, which then was applied to experimental data. After all the present theory is not constrained to a special selection of plasma models or idealized assumptions. Whenever it is possible to describe a real world plasma by more sophisticated models, it is only a minor step to incorporate these models in the present theory.

B. Expansion order

The terms $p(\boldsymbol{W^{exp}}|E, M_j, \boldsymbol{\sigma}, I)$ arising in (23) can be inverted using Bayes theorem to obtain the probability for the expansion order $p(E|\boldsymbol{W^{exp}}, M_j, \boldsymbol{\sigma}, I)$ in the light of the data and a particular model:

$$\begin{aligned} p(E|\boldsymbol{W^{exp}}, M_j, \boldsymbol{\sigma}, I) &= \frac{p(E|M_j, I)p(\boldsymbol{W^{exp}}|E, M_j, \boldsymbol{\sigma}, I)}{p(\boldsymbol{W^{exp}}|M_j, \boldsymbol{\sigma}, I)} \\ &= \frac{p(\boldsymbol{W^{exp}}|E, M_j, \boldsymbol{\sigma}, I)}{\sum_{E'} p(\boldsymbol{W^{exp}}|E', M_j, \boldsymbol{\sigma}, I)} \ . \end{aligned} \tag{61}$$

The expansion order with the highest probability is also shown for each model in table I. In the sum over all expansion orders in (23) it has the most important contribution as may be seen in Fig. 1 where $p(E|\boldsymbol{W^{exp}}, M_j, \boldsymbol{\sigma}, I)$ is displayed as full circles for the most probable plasma model, the collisional low beta case. The error bars indicate Monte Carlo integration uncertainties. The open squares, with associated error bars, depict the misfit between data and model prediction (3) as function of the expansion order. We observe quite a typical behavior. Up to three terms in the expansion (3) lead to a rapid decrease of the misfit. Though it does decrease further monotonically with increasing expansion order, the probability for a given E decreases rapidly (note the logarithmic scale!), so that contributions of higher expansion orders become very small. This is a demonstration of Occam's razor automatically included in Bayesian theory. This principle dictates that a simpler model should be preferred unless a more complicated one leads to a substantially better fit to the data. Note that the present optimum three term expansion reduces the misfit to about 65% of its initial value. The possibility for such an effect has previously been pointed out by

Kaye et al. [5].

Since already an expansion order of E=3 shows to be sufficient to describe the data leading to a W^{theo} given by (5), one is tempted to ask for the respective scaling exponents and coefficients. However, to obtain this and the following results it is necessary to integrate over all parameters entering the problem (i.e. $\omega, \boldsymbol{c}, \boldsymbol{x}$). Therefore it is not possible to describe the scaling parameters in the traditional sense giving a list of their values.

C. Energy content prediction

After having obtained the most probable model we can proceed and ask for the prediction of the present theory for the energy content. The latter is the expectation value $\langle W \rangle_v$ (50) obtained from averaging W over $p(W|\boldsymbol{W^{exp}}, \boldsymbol{v}, M_j, \boldsymbol{\sigma}, I)$. The additional condition $\boldsymbol{v}$ specifies the "input data vector" $\boldsymbol{v} = (n, B, P, a)$, where (n, B, P, a) are the parameters for which one wants to predict the energy content. The analytic expressions were given in chapter IV E. Fig. 2 shows a scatter plot of the predicted energy content versus the experimental data.

Even more, it is possible to test the assertion regarding the most probable model. Such a test is provided by a comparison of measured single variable scans to the predictions from the present theory. In the following we will show the result for density and power scans obtained on the one hand from the present theory and on the other hand from experiments on W7-AS. Because these data are not included in the stellarator confinement data base $\boldsymbol{W^{exp}}$, this test shows the predictive power of our approach.

The full circles in Fig. 3 represent experimental results for the density scan. Representative error bars signify the precision level of these data. The continuous curve depicts the result of the present semi-empirical theory along with the confidence range indicated by the gray shaded area. The stellarator energy confinement data base is represented by the open circles which are spread all over since they were obtained for various settings of the variables (B, P, a). The histogram at the base line indicates the number of shots in $\boldsymbol{W^{exp}}$ at the respective density and gives an impression of the range over which our result for the single variable scan is best supported by the data base. Last but not least the dashed curve represents the density dependence as inferred from an unrestricted single power law conventional least squares fit resting on the same data $\boldsymbol{W}^{exp}$ as our Bayesian result. It only hits the progression of the data at two points, while staying out of the data scatter (of the full circles) most of the time. Within the density range of the single variable scan the prediction of the semi-empirical theory runs straight through the data and exhibits clearly the previously supposed density saturation [3,18], which can never be obtained by a single power law term at all. Outside this range the data set $\boldsymbol{W}^{exp}$ is too sparse, which is reflected in the rapidly widening error band. In contrast to the robust but erroneous power law scaling the present theory indicates where the extrapolation becomes unreliable and additional experiments are inevitable. It might be an unfortunate but honest conclusion that an extrapolation beyond the parameter regime supported by the data base is not possible. Note that the comparison between the single variable scan and the prediction of our analysis holds on absolute scales! Neither in Fig. 3 nor in the subsequent Fig. 4 are adjustable scale parameters necessary. This means that experiments in W7-AS have an impressive reproducibility.

Finally, Fig. 4 displays a similar comparison for a power scan in W7-AS. Again the semi-empirical theory shown as the continuous line predicts the measured energy content - on an absolute scale - within experimental error and corroborates the experimentally observed power degradation. The dashed curve is from a power law fit and is, as for the density scan, concave while the present model shows convex dependence in both cases.

VI. CONCLUSION

Based upon the rules of probability theory a set of form-free dimensionally exact energy confinement functions was analyzed for the 153 $\iota \approx 1/3$ W7-AS data of the international stellarator data base. In a first step, out of four Connor Taylor models the collisional low beta one was identified to be the most probable plasma physics model for W7-AS. Secondly, single variable scans were reproduced in quantitative agreement with experiments. The result of a single variable scan is therefore already hidden in the data obtained for arbitrary variable choices and can be extracted from the latter by a proper data analysis. Even more, the present approach has two major advantages over the traditional best fit approach: Besides its predictive power even on a quantitative level it indicates honestly where the prediction can be trusted and where additional experiments are required.

The results obtained in the present work differ in particular from one conclusion drawn in our earlier paper [7]. Our previous approach was a dimensionally constrained energy confinement analysis of an enlarged data set (including also the $\iota \approx 1/2$ measurement in addition to the $\iota \approx 1/3$ subset used in this paper). The confinement function was restricted to a single power law term with varying exponents. The exponents were determined from a Bayesian analysis trading between best fit and dimensional correctness. All conclusions drawn from this approach are conditional on the assumption that the confinement function is of the power law type. The result of the model comparison was that the highest model probability was associated with the collisionless high beta case. Caveats on this conclusion were however also given. In particular, the selectivity in model probabilities was considered to be too small by far. The simplest collisionless low beta model differed previously only by a factor of seven from the collisionless high beta case. In the present work, which does not assume anything about the scaling function, the "distance" between the least probable and the most probable model surmounts thirteen orders of magnitude. Furthermore the reproduction of single variable scans as in Figs. 3 and 4 is at the heart of the present work and could never have been achieved by the previous treatment.

Acknowledgments

The authors acknowledge discussions of errors associated with the energy content measurements with M. Anton, V. Erckmann, R. Jänicke, F.P. Penningsfeld, U. Stroth and F. Wagner. In addition we would like to thank H. Maaßberg for instructive discussions. One of us (V. Dose) is indebted to D. Pfirsch for discussions on the Connor/Taylor paper.

APPENDIX A: ESTIMATION OF THE EXPERIMENTAL ERROR OF THE ENERGY CONTENT

For our estimation we consider the energy content in the one term power law ansatz:

$$W = cn^{\alpha_n} B^{\alpha_B} P^{\alpha_P} a^{\alpha_a} \ . \tag{A1}$$

We use this simplified approach rather than the full expansion in dimensionally exact scaling functions because it is sufficient to have an overall description of the error dependence on the input variables. The magnetic field B is measured very accurately, so we assume its error to be negligible ($\sigma_B = 0$). An upper limit of the measurement error of the energy content itself is $\sigma_W = 0.5kJ$. The minor radius a is determined with an error of about $\sigma_a = 0.5cm$. Since the line integral over the electron density n_e

$$\int n_e dl = 2an \ , \tag{A2}$$

is obtained with very high accuracy (within 2 %), the error in the density n depends mainly on the error in a.

$$\frac{\partial n}{\partial a} = -\frac{\int n_e dl}{2a^2} = -\frac{n}{a} \ . \tag{A3}$$

To determine the error of the heating power P we have to take a closer look at the constituents of $P = P_{abs,ECR} + P_{abs,NBI}$. The error of the ECR-heating is negligible for densities smaller than $n < 8 \cdot 10^{19}/m^3$ and can be set to zero for our data set. The total absorbed NBI-heating power depends on the density via the global heating efficiency $\eta(n)$ [19]

$$P_{abs,NBI} = \eta(n) P_{NBI} \ , \tag{A4}$$

with P_{NBI} as the power of the injected particles and

$$\begin{aligned} 4\eta(n) = {} & [1.075 - 1.3\exp(-10.8n \cdot 2a)]\exp(-0.18n \cdot 2a) \\ & + [1.075 - 1.3\exp(-6.5n \cdot 2a)]\exp(-0.18n \cdot 2a) \\ & + [0.85 - \exp(-8.5n \cdot 2a)]\exp(-0.7n \cdot 2a) \\ & + [1.03 - 1.25\exp(-6n \cdot 2a)]\exp(-0.75n \cdot 2a) \ , \end{aligned} \tag{A5}$$

where n is in $[10^{20}/m^3]$ and a is in $[m]$. The error in P_{NBI} is again negligible, but $\eta(n)$ has the error dependence on the minor radius as stated above. Therefore

$$\frac{\partial P}{\partial a} = \frac{\partial P}{\partial n}\frac{\partial n}{\partial a} = -\frac{P_{abs,NBI}}{\eta(n)}\frac{\partial \eta}{\partial n}\frac{n}{a} \ . \tag{A6}$$

Since n and P are finally dependent on the error in a the error of the energy content is given by

$$\sigma^2 = \sigma_W^2 + \left(\frac{\partial W}{\partial a}\sigma_a\right)^2 \ , \tag{A7}$$

with

$$\frac{\partial W}{\partial a} = \alpha_n \frac{W}{n}\frac{\partial n}{\partial a} + \alpha_a \frac{W}{a} + \alpha_P \frac{W}{P}\frac{\partial P}{\partial a} \ . \tag{A8}$$

Inserting the partial derivatives from (A3) and (A6) gives the final result for our error estimate

$$\sigma^2 = \sigma_W^2 + W^2 \left(\frac{\sigma_a}{a}\right)^2 \times \left(\alpha_a - \alpha_n - \alpha_P \frac{P_{abs,NBI}}{P_{abs,NBI} + P_{abs,ECR}} \frac{1}{\eta}\frac{\partial \eta}{\partial n} n\right)^2 \ . \tag{A9}$$

In Fig. 5 the actual data is shown with the error bars assigned. The inset shows the value of the error consisting of a constant part (σ_W term in (A9)) and a contribution originating from the uncertainties in the machine variables a, n and P (comprised by the second term in (A9)). The latter varies roughly like the energy content W as expected from (A9).

APPENDIX B: NUMERICAL TESTS

Let us check our analysis and simulate a data set in order to examine the performance. The data set $\hat{\boldsymbol{W}}^{\boldsymbol{exp}}$ shall be composed in equal parts of two vectors, the first dimensionally exact in the collisional low-β (M_2), the other in the collisionless high-β CT-model (M_3). Both vectors consist of a single power law term only (expansion order $E = 1$).

$$\hat{\boldsymbol{W}}^{\boldsymbol{exp}}(x, y, z) = c\left[\boldsymbol{f}_{M_2}(x, y) + \boldsymbol{f}_{M_3}(x, z)\right] \ . \tag{B1}$$

The coefficient c is adjusted such that $||\boldsymbol{W}^{\boldsymbol{exp}} - \hat{\boldsymbol{W}}^{\boldsymbol{exp}}||^2$ is a minimum so that $\hat{\boldsymbol{W}}^{\boldsymbol{exp}}$ is as close as possible to the actual experimental data. Only the highest model M_4 with three degrees of freedom should be capable of explaining a data set $\hat{\boldsymbol{W}}^{\boldsymbol{exp}}(x, y, z)$ constructed in this special way. However, if the data is noisy simpler models may be sufficient to explain the data (Occam's razor).

To simulate similar conditions as in the experiment Gaussian noise of width according to the error estimate in appendix A is added to (B1). We start with a data set for those (x_0, y_0, z_0) which correspond to scaling exponents obtained by a least squares fit for a single power law term (Fig. 6a). Here the simplest model M_1 (which depends only on one variable x, see table I) wins, indicating that either the values for y_0 and z_0 are not significant enough to justify the choice of a more complex model or that the noise is too large. Let us first examine what happens if y and z are increased by a factor of 4 (Fig. 6b). While the probability for M_1 vanishes ($p(M_1|\hat{\boldsymbol{W}}^{\boldsymbol{exp}}, I) = 10^{-22}$) the third model now wins, followed by the more complex M_4. Doubling y and z again gives nearly the same probabilities for M_3 and M_4, but the latter is still less probable than M_3 (Fig. 6c). A further increase in y and z would lead to negative and therefore, according to general experience, unrealistic exponents. But we can still test our above stated assumption that if the data is not affected by noise the highest model has to succeed. And indeed, in the second case $(x_0, 4y_0, 4z_0)$ a reduction of the added Gaussian noise by a factor of ten already shifts the model probability to the collisional high-β model and clearly shows the hierarchy in the complexity of the models (Fig. 6d).

REFERENCES

[1] DE BOO, J. C. et al., Sov. J. Plasma Phys. **11** (1985) 40.

[2] REBUT, P. H., LALLIA, P. P., and WATKINS, M. L., Plasma Physics and Controlled Nucl. Fusion Research **2** (1988) 191.

[3] STROTH, U. et al., Nucl. Fusion **36** (1996) 1063.

[4] CONNOR, J. W. and TAYLOR, J. B., Nucl. Fusion **17** (1977) 1047.

[5] KAYE, S. M. et al., Phys. Fluids B **2** (1990) 2926.

[6] CHRISTIANSEN, J. P., CORDEY, J. G., KARDAUN, O. J. W. F., and THOMSEN, K., Nucl. Fusion **31** (1991) 2117.

[7] DOSE, V., PREUSS, R., and VON DER LINDEN, W., Nucl. Fusion **38** (1998) 571.

[8] DOSE, V., PREUSS, R., and VON DER LINDEN, W., Phys. Rev. Lett. **81** (1998) 3407.

[9] HUGILL, J. and SHEFFIELD, J., Nucl. Fusion **18** (1978) 15.

[10] KADOMTSEV, B. B., Sov. Phys.-J. Plasma Phys. **1** (1975) 295.

[11] CHRISTIANSEN, J. P., CORDEY, J. G., and THOMSEN, K., Nucl. Fusion **30** (1990) 1183.

[12] BRETTHORST, G. L., *Bayesian Spectrum Analysis and Parameter Estimation*, Springer Press, Berlin, Heidelberg, 1988.

[13] O'HAGAN, A., *Kendall's advanced theory of statistics, Bayesian Inference*, John Wiley & Sons, New York, 1994.

[14] JAYNES, E. T., Prior probabilities, in *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics*, edited by ROSENKRANTZ, R. D., page 114, Reidel, Dordrecht, 1983.

[15] RODRIGUEZ, C., From euclid to entropy, in *Maximum Entropy and Bayesian Methods*, edited by W. T. GRANDY, J., Kluwer Academic, Dordrecht, 1991.

[16] VON DER LINDEN, W., PREUSS, R., and DOSE, V., The prior predictive value, in *Maximum Entropy and Bayesian Methods*, edited by ET AL., V. D., Kluwer Academic, Dordrecht, 1998.

[17] DOSE, V., VON DER LINDEN, W., and GARRETT, A., Nucl. Fusion **36** (1996) 735.

[18] STROTH, U., Plasma Phys. Control. Fusion **40** (1998) 9.

[19] PENNINGSFELD, F. P., Fast access to neutral injection heating profiles in W7-AS, Technical Report IPP 4/265, Max-Planck-Institut für Plasmaphysik, 1993.

TABLES

TABLE I. Parameters of CT models, model probability and maximum expansion order

CT model	x_1	x_2	x_3	X_{dof}	$p(M_j \vert \boldsymbol{W^{exp}}, \boldsymbol{\sigma}, I)$	E_{max}
1. collisionless low-β	x	0	0	1	$4 \times 10^{-12}\%$	3
2. collisional low-β	x	y	0	2	99.7%	3
3. collisionless high-β	x	0	z	2	0.25%	3
4. collisional high-β	x	y	z	3	0.025%	2

TABLE II. Range for the 153 $\iota \approx 1/3$ data of W7-AS taken from [3].

W $[kJ]$	n $[10^{19}m^{-3}]$	B $[T]$	P $[MW]$	a $[m]$
0.475 – 12.1	0.831 – 18.5	1.24 – 2.54	0.119 – 1.18	0.112 – 0.176

FIGURES

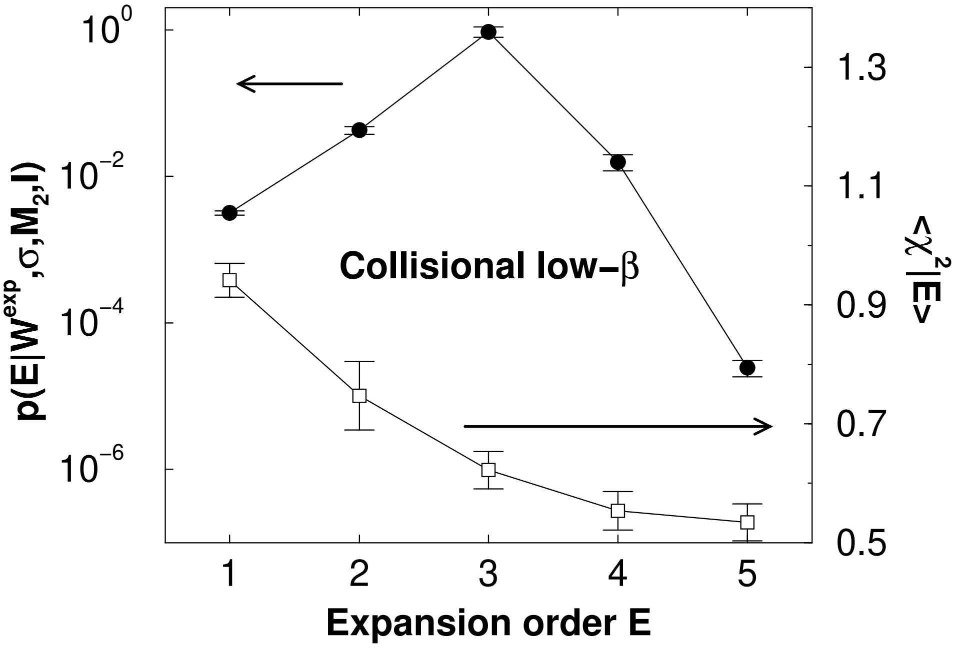

FIG. 1. Probability of the expansion order (lower graph) and the decrease in χ^2 (upper graph) in the collisional low-β model.

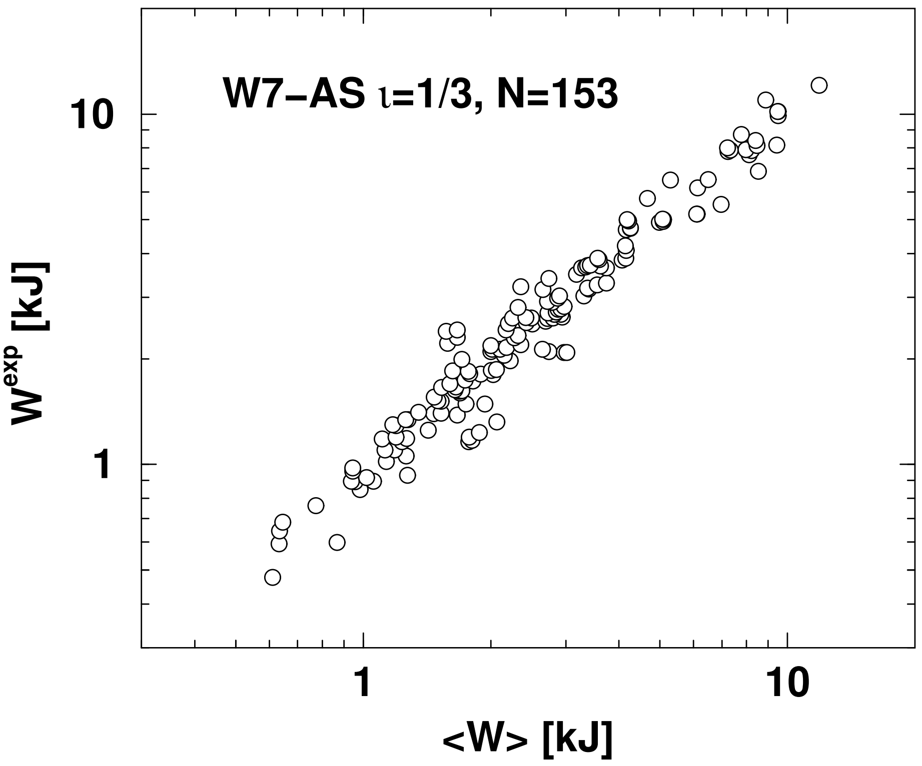

FIG. 2. Scatter plot showing the energy content of the plasma predicted by the present approach $\langle W \rangle$ versus the experimental data W^{exp}.

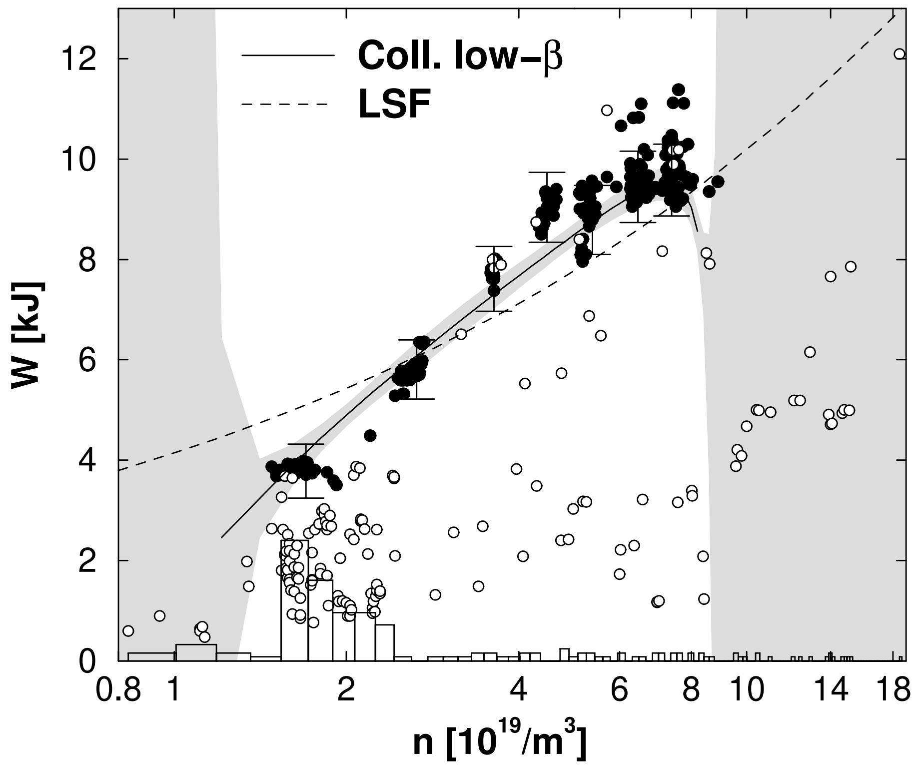

FIG. 3. Energy content of the plasma as function of the density. Open circles: input data. Notice that these data are obtained for choices of (B, P, a) which differ from point to point and are therefore spread all over. The histogram accounts for the distribution over the density axis. Full circles: Single variable scan for $B = 2.5$T, $P = 0.45$MW, $a = 0.176$m. Only some data points carry error bars to give an impression of the uncertainty in the measurement. The result of our method is given by the solid line, where the shaded area is the error. A least squares fit (LSF) of the input data would yield $W \sim n^{0.39}$ (dashed line).

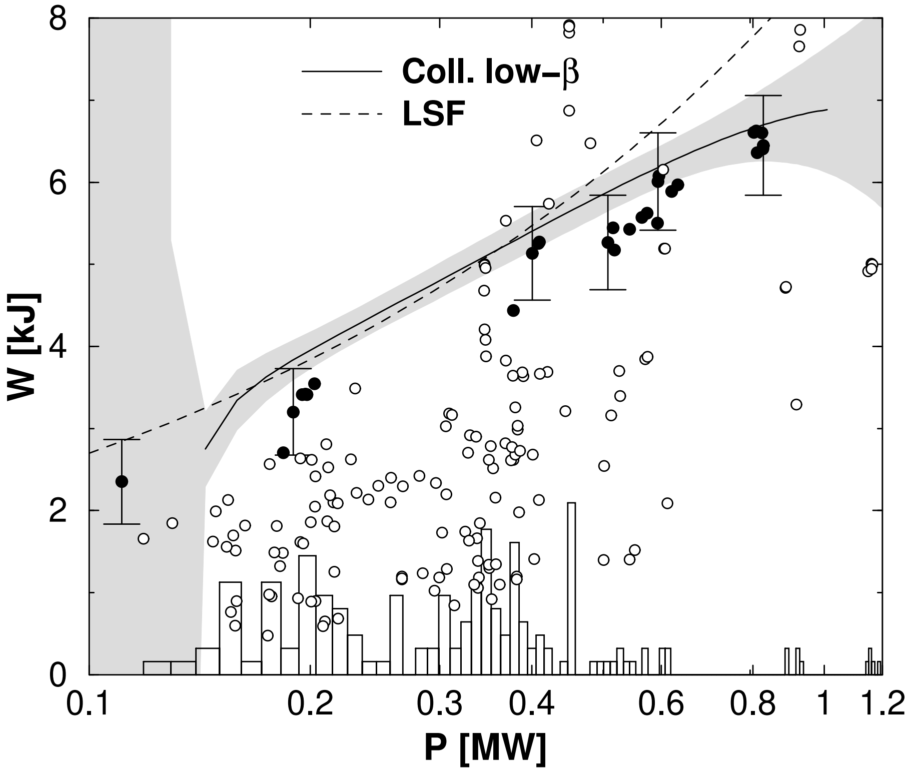

FIG. 4. Power dependence of the energy content. Full circles: Single variable scan for $n = 2.4 \cdot 10^{19}\mathrm{m}^{-3}$, $B = 2.5\mathrm{T}$, $a = 0.176\mathrm{m}$. The histogram is again with respect to the input data. Our result is represented by the solid line with the shaded area as the error region. A least squares fit of the input data would yield $W \sim P^{0.5}$ (dashed line).

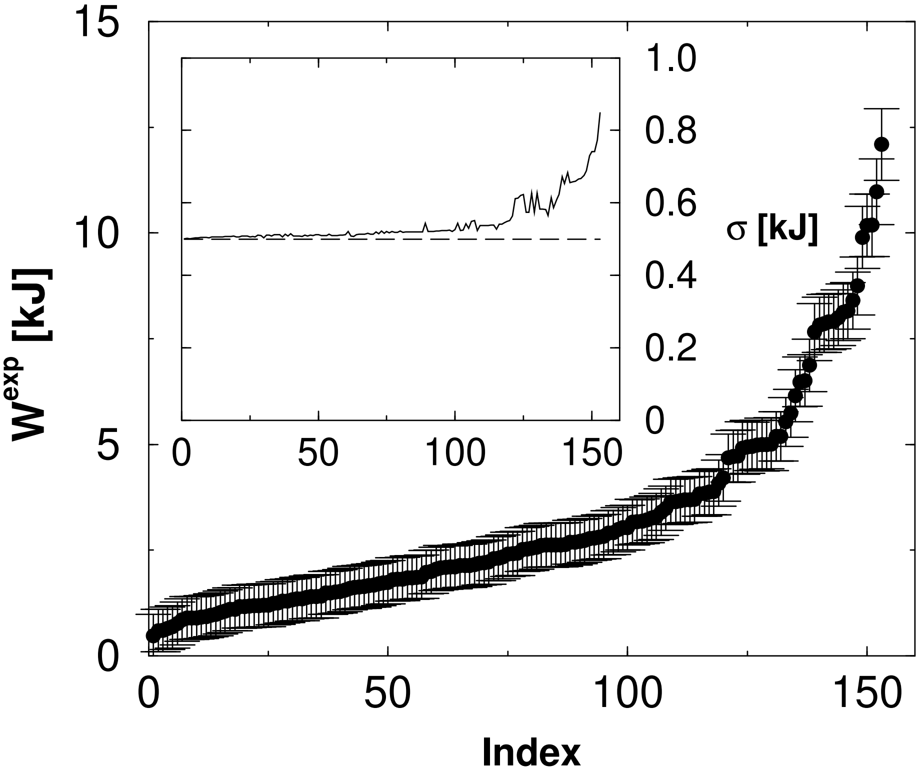

FIG. 5. Error bars assigned to the experimental energy content W^{exp}. The data was arranged with respect to size. Inset: Value of the error in same order as the data. The dashed line is the error of the diamagnetic measurement: $\sigma_W = 0.5kJ$.

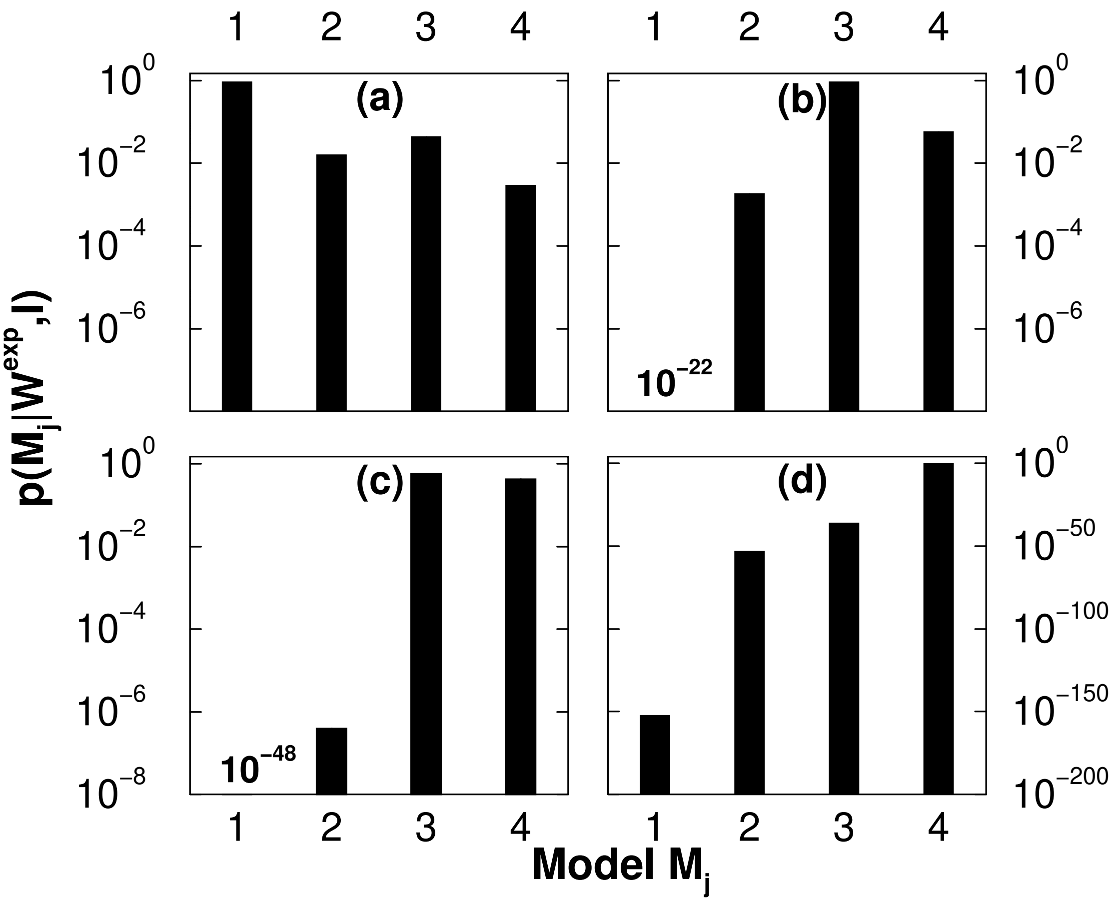

FIG. 6. Model probabilities for a simulated data set according to Eq. (B1) (see text): (a) The chosen (x, y, z) correspond to scaling exponents obtained by a least squares fit for a single power law term: x_0=0.485, y_0=0.032, z_0=0.063; (b) x=x_0, y=4y_0, z=4z_0; (c) x=x_0, y=8y_0, z=8z_0; (d) x=x_0, y=4y_0, z=4z_0, σ=σ_0/10. The data in (a), (b), and (c) was corrupted by Gaussian noise in the range of our error estimate σ_0, for (d) the added noise was reduced by a factor of 10.