

Computer Simulations, Analysis and Transfer of PDMS-Spectra

St. Harsdorf^a, E.R. Hilf^{a,b}, B. Nitzschmann^a,
W. Schlez^a, J. Tomaschewski^a, K. Vogt^a, and P. Wagner^a

^a *Fachbereich Physik,*

^b *Institut für Chemie und Biologie des Meeres (ICBM)
Universität Oldenburg, FRG*

Abstract

The PDMS process was simulated by a MD (Molecular Dynamics) code for variable incident-ion angles and using a 3-dimensional array of molecules. The intermolecular Lennard-Jones potential is assumed to depend on the electronic temperature. The resulting cluster spectra and their internal excitations are studied.

Some tools have been developed to extract information from PDMS spectra by numerical evaluation. The program ABUNDANCE extracts the relevant information on the spectral lines by *far*-moments instead of the more used to *close*-moments, with the advantage of better automatization. SYNTHESIS extracts chemical structure information; BREAK calculates the fragment spectrum for a pure substance of known chemical structure with entropic-estimated bond breaking probabilities, — even if the structure is of netted geometry; COMPARE allows the comparison of a spectrum to others of a data base, but using a linear instead of the usual Euclidean norm.

Finally EMS-NET is presented, a tool to communicate with other laboratories by automated search and transfer of spectra including the necessary side-information. Its tools are BITNET, ORACLE, ISPF, and REXX, packages served by many main-frame computers.

1. Simulation of PDMS Spectra by Molecular Dynamics Calculation

A Molecular Dynamics simulation was set up to learn about the mechanical part of the desorption process. As a start point the initial thermal distribution of electronic excitation, developed by the mutual interaction of the secondary electrons in flight, induces a momentary change of the atomic Lennard Jones potentials. These then serve as driving forces for the full set of microscopic mechanical equations of motion of an array of up to $100 \cdot 100 \cdot 10$ molecules.

*supported by the German Ministry for Science and Technology, BMFT, division of Marine Research, grant No. 8578 MF.

Such idealized simulation calculations may be useful for the understanding of a complex physical process because of its possibility of systematic variation of physically relevant parameters, before setting up phenomenological models which are easy to evaluate, such as shock wave explosion, thermal spike, hydrodynamic expansion.

The hope is to connect two sets of quantities, the ones determining the desorption such as the angle, energy, charge and mass of the incoming primary ion, and the ones defining the mechanical properties of the impinged solid, such as intermolecular forces, sound velocity, molecular weight, shape, and size, etc. with the velocity and arrival time profile of the eventually counted desorbed particles. By this, one may serve their hope for experimental optimization of yields, resolution, high mass-end detection, etc. .

Here we calculated for various incident-ion angles and dE/dx of the incoming heavy ion the start velocities and internal excitation of the desorbed particles.

The MD process is emulated by setting the electronic temperature along the impact track to a level of about 10 eV . The developing electronic temperature distribution, $T(R, t)$, drops by diffusion. We use the simple scheme $T_{i,t+1} = T_{i,t} + C \sum_{j=1}^{NN} (T_{j,t} - T_{i,t})$.

The attractive part of the atomic potentials is then reduced responding to the momentary electronic temperature,

$$U = U_0 \left(\frac{A}{r^{12}} - \frac{B}{r^6} e^{-\frac{G}{k_B T}} \right) f(r)$$

with G being the electronic band gap¹, and k_B the Boltzmann constant. The attenuation function $f(r)$ smoothes the singularity at $r = 0$.

The attributes of the mass points were adapted here from ordinary ice, as a simple small molecule of wellknown parameters. The iteration scheme used here is the leapfrog scheme.

We neglect the Coulomb repulsion between the desorbed particles, thus setting aside the relevant findings of Macfarlane², Moshhammer and Wien^{3,4} here.

As an example out of the numerous⁵ calculable quantities we give in Fig. 1 the yield of desorbed particles as a function of their time of desorption. Emerging in a bulk, they seem to tell of an explosive type of process rather than thermal emission. The initial internal excitation of the desorbed particles decreases with their time of leaving the surface, see Fig. 2 , which is supported by the increasing average distance from the primary ion track. The present status of all the simulation calculations may be seen as support for the present qualitative understanding of the PDMS process as explosive triggered by a primary hot ion track. However a lot of work has still to be done to gain a realistic quantitative corroboration of the respective known experimental data.

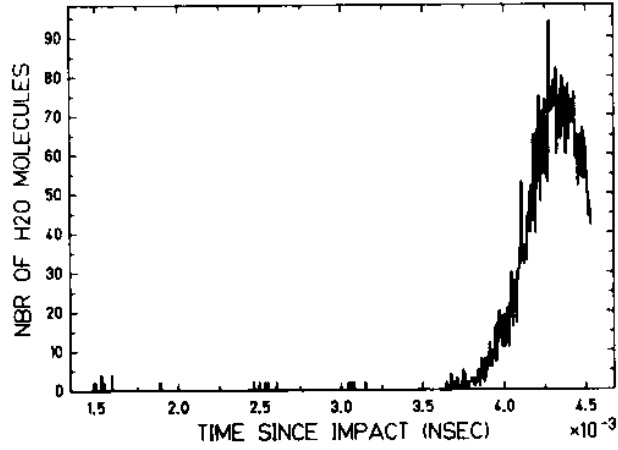


Figure 1: Yield of desorbed particles as a function of their time of desorption.

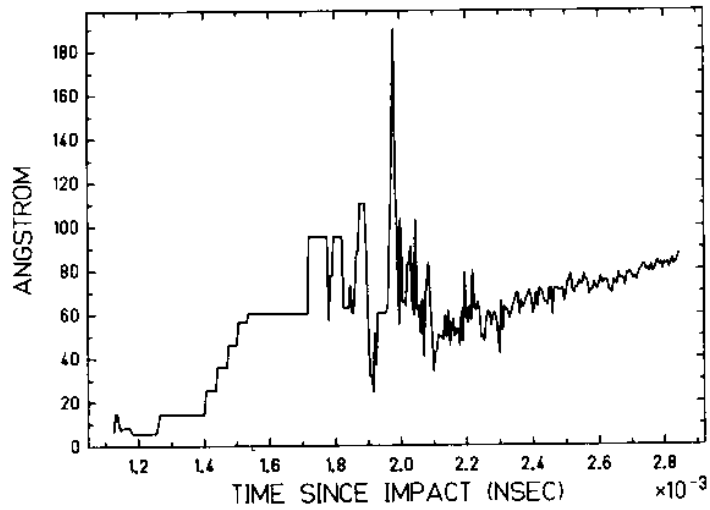


Figure 2: Mean distance to impact track of particles leaving the surface as a function of their time of desorption. The data beyond $2.2 \cdot 10^{-3}$ are irrelevant sublimations, since the bulk of the desorbed particles had left in this run. Parameters are different from Fig. 1 .

2. Why Numeric Evaluation of PDMS-Spectra?

MeV-ion per nucleon induced desorbed secondary ions are measured by their time of flight unto their arrival at the stop detector. Typically, with the recent advent of the ORSAY-TDC, 65,000 adjacent time bins incoming ions are used. Due to the joint acceleration voltage the time of flight is assumed to be roughly proportional to the square root of the mass. Thus PDMS is used as a universal mass spectrometer. The information wanted is typically either the presence of a known molecule e.g. in a mixture, or chemical structure from an unknown molecule, hopefully of a pure sample.

The most efficient way up to now in the overwhelming majority of cases is the use of a well tuned instrument in combination with an expertly prepared sample and a highly trained eye of the experimentalist to integrate over the presented information by inspecting a spectrum.

The use of computers has been invading this creative process in several steps. The already used-to step is the use of software for calibration, display with various tools such as zoom or scaling, search for peaks, printout, integration into texts, etc.

Further steps are called for if one wants or if there is a need to broaden the scope of the up to now possible applications. Examples are the analysis of more complicated cluster spectra as e.g. Mn-Oxides, more complex mixtures as samples from some applied fields⁶, more information pieces for the chemical structure⁷ of a given molecule, large numbers of similar samples in industry, where it is financially paying, to fully exploit every possible automatization, and retrieve only the minor part of information, one really asked for. The idea here is that computer cranking is cheaper than chemistry or paper floods of output. Thus it pays, not to unearth any information more than once, and not more than really wanted. The second to final step, what has been called an expert system, is, when the experience of the experimenter is followed as it builds up during a series of experiments. Recalling his decisions, propositions are then made by the software program what a judgement by a person could give. For these AI (Artificial Intelligence)-programs several examples are at present developed for different applications in spectroscopy.

Although we confess that we are inclined to work along these lines we present here only some basic software tools, at present under development and offered to any research laboratory as they become available (Just contact us at 091853@doluni1). Finally they may be integrated into a larger package. We will restrict ourselves here to give some basic operational ideas.

3. Background Subtraction and Moment Analysis

The disentangling of the relevant part of a spectrum from its background is known to be problematic. The original experimental information is stored here as a time of flight channel spectrum of, numbers, giving the counts, registered in successive time bins of e.g. 1 *nsec*. As a usual first step one may inspect some graphical display and pinpoint in most cases the more conspicuous lines which are then calibrated with some knowledge on a few known lines.

The background may be, as a first guide line, accounted for by subtracting a constant from the yields, to help ease the inspection of the spectrum. However by this the resulting net spectrum may contain some negative *counts*, which have no measurable counterpart.

There are several attempts to overcome the intricate intervovenness of fragment counts and background, at least partially. All skip the desire to learn local, channel-wise information and pass over to summing over some part of the channel spectrum with some weight functions, with the advantage that from these integral quantities which will be called *moments*^{8,9}, one can safely subtract analogously calculated sums over the background (or an assumed, or even added one)¹⁰.

Here we will just give a simple example, with the aim to yield quantities close to those, which most inspection-experts would say, characterize a mass line in decreasing order of interest, such as line position, line count content, line width, line asymmetry. But it should be kept in mind, that thus even the line position is the result of a weighted sum or 'integral'.

However we hope that such evaluation methods may be especially useful for automatizing the computer analysis of a given spectrum, with a somewhat known bias, aimed to be useful also in cases of a large background. The desired observables assumed to be of decreasing interest are now defined.

A simple wellknown definition of a set of power series moments¹¹ is

$$P_n(j) := \sum_{t_i \in C_j} (t_i - t_0)^n Y(t_i)$$

summed over a range C_j of time bins which hopefully contains just one relevant line j . These wellknown *close*-moments, where t_0 is the line position calculated with a linear weight,

$$t_0 := \sum_{t_i \in C_j} t_i Y(t_i) / \sum_{t_i \in C_j} Y(t_i) \quad .$$

By this explicit calculation of t_0 an effective computer automatization of the line scanning of large spectra is hindered. Some further slight disadvantages will be discussed by comparing to another set of moments, named *far*-moments,

$$\hat{P}_n(j) := \sum_{t_i \in \hat{C}_j} (t_i - t_1)^n Y(t_i - t_1) \quad ,$$

where here t_1 is arbitrary and because it may be even well below the assumed position of the line in question, assumed to be in \mathcal{C} .

The desired observables assumed to be of decreasing interest are then defined by

$$\begin{aligned}
 \text{Amplitude } A &:= P_0 b / \sqrt{2\pi} & \hat{A} &:= P_0 \hat{b} / \sqrt{2\pi} \\
 \text{Position } t_0 &:= \hat{P}_1 / \hat{P}_0 |_{t_1:=0} & \hat{t}_0 &:= \hat{P}_1 / \hat{P}_0 + t_1 \\
 \text{Width } b &:= \sqrt{P_2 / P_0} & \hat{b} &:= \sqrt{\hat{P}_2 / \hat{P}_0 - \hat{t}_0^2} \\
 \text{Asymmetry } \gamma &:= (P_3 / P_0) / b^3 & \hat{\gamma} &:= (\hat{P}_3 / \hat{P}_0 - \hat{t}_0 \hat{P}_2 (3 - \hat{t}_0^2 \hat{P}_0 / \hat{P}_2) / \hat{P}_0) / \hat{b}^3 \\
 & & & \vdots
 \end{aligned}$$

The expressions using the *far*-moments look more clumsy but can be easily translated using the known conversion

$$\hat{P}_n = \sum_{m=0}^n \binom{n}{m} t_0^{n-m} P_m .$$

There are some reasons to prefer the *far*-moments. They are good for automatic scanning of spectra, since no positions of lines have to be known in advance. There is a better noise escamoting power. The numeric evaluation using the Horner scheme up to level n is faster, with only n instead of $n + 1$ multiplications. Data are to be picked only once instead of twice with a CPU time saving of $1/(n + 1)$ for the n^{th} moment, as long as the number of channels involved is much larger than n .

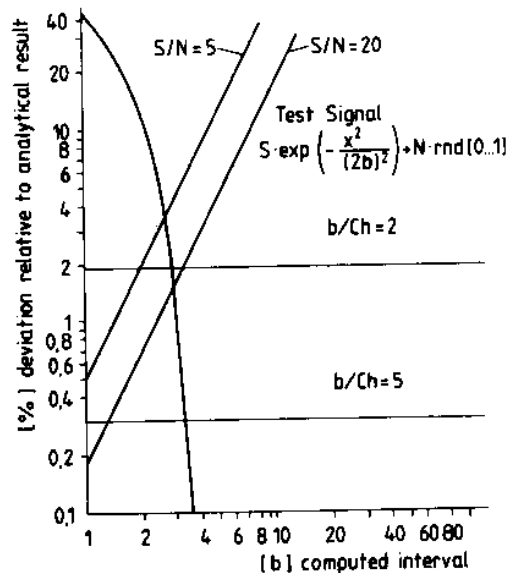
The *far*-moments, in addition, are somewhat more accurate, see Fig. 3., deviations stemming from the width of \mathcal{C} , the signal to noise level, and the discretization of \hat{t}_0 . The last one is important for narrow signals using *close*-moments. Noise-effects and the size of \mathcal{C} are somewhat counteractive.

Any moment analysis is much faster than the scheme used at present by us in daily work: For each channel t_j of the raw spectrum we calculate the (linear) scalar product of the experimental spectrum with an ideally assumed line shape $Y_o(t_i) = Y_{\infty} (1 + \tilde{\gamma}_j) \exp((t_i - t_j)^2 / \tilde{b}_j)$,

$$\hat{Y}(t_j) := \sum_{i=0}^{65000} (Y_o(t_i) Y_{\text{exp}}(t_i))^{1/2} . \quad (1)$$

These scalar products form a new spectrum, $\hat{Y}(t_j)$, where the background appears reduced and independent of the time resolution. It is known for the original time bin spectrum that by increasing the number of time bins per given flight time interval, one *decreases* the yield-resolution. In the averaged spectrum, lines are more conspicuous. In a next step the highest peak of the spectrum is picked. For its peak position $t_{j'}$ a fit to the original spectrum is done by minimizing the scalar (linear) product of the above defined deformed Gaussian, now adjusting its parameters (number of counts $Y_{\infty}(t_{j'})$, width $b_{j'}$, and asymmetry $\gamma_{j'}$) and storing them as results

Figure 3: Deviations of *close-* and *far-*moments from the exact values for different signal to noise ratio (S/N) and signal to channel width ratio (b/Ch).



for the first identified line. Then the adjusted Gaussian is subtracted from the original data, negative counts put to zero, and the whole procedure is started for the next line. Though this procedure works rather well gaining lines in successive order of magnitude, and works fairly well even for partially overlapping lines, and lines in the high mass region, where a large number of channels contribute, and the signal to noise ratio is rather small, this procedure is rather CPU time consuming in proportion to the number of properties wished to gain for any line. With at least as many sums as the number of channels times the number of lines times the number of properties times the number of steps for the fit...

In contrast, analyzing with the *far-*moments needs just one sum over a small range of channels per desired line property, and can, in addition, be organized such as to scan the spectrum successively and once with only a few loops in areas where the set of moments for an assumed line hints by unusual large b , γ , that a redo with one more assumed line may work better.

The information stored in many equally important channels is transferred by some kind of weighted summation (integration) into a series of observables of *decreasing degree of interest*. Using a set of moments an automatic analysis of many spectra and their comparison becomes possible.

4. Chemical structure and PDMS-Spectrum

For application, especially such a wide field as the application of PDMS to marine research, some tools may turn out useful, aimed to support chemical analysis or substance identification in a mixed sample.

4.1. Synthesis of Structural Information

For a given sample of an assumed to be pure substance often the information asked for are hints on or pieces of the chemical structure of the possibly unknown molecule. We have developed a tool, SYNTHESIS, to gain stepwise chemical information.

The first step, just automatizing what is routinely done in discussing a spectrum, is intended to look for a chemical bond, that is finding two fragments and their superstructure, identified by the sum of their mass. Thus all triplet of lines are scanned through, and are kept, if the sum of the mass of two of them gives the mass of the third,

$$Y(m_i) \neq 0, \quad Y(m_j) \neq 0, \quad Y(m_k) \neq 0, \\ m_i + m_j = m_k, \quad i, j, k = 1, \dots, N.$$

This triplet is then kept as a possible chemical bond $i - j$ between the fragments i and j forming the superstructure k . Because of the many cases in PDMS that the charge of the molecular or even fragment ion is an adduct with some electron donor or acceptor such as Na, K, H, . . . , we include these into the search.

In the second step now the bondlist is scanned for any two bonds with the same superstructure. Thus we have to check, scanning the bond list, whether

$$Y(m_l) \neq 0 \quad \text{for } l = a, b, c, d, e, g, \quad \text{and } Y(m_f) = 0, \\ m_a + m_b = m_c = m_d + m_e, \quad m_a + m_e = m_f, \quad m_a + m_g = m_d.$$

for $a, b, c, d, e, f, g = 1, 2, 3, \dots, N$. This establishes the proposition of the structural piece $a - g - e$. Thus *not-found* masslines are explicitly included to establish structural findings.

The third and higher steps can in principle be designed following these lines.

We realized, that for the higher steps of SYNTHESIS it pays at present, reducing the CPU-time and rising greatly the final chance of success, to interweave the expertness, intuition and knowledge of the prospective user, so as to drop unrealistic solutions in an early stage, reduce thus the number of cases to follow, and pick more plausible ones. This helps especially for the intrinsic ambiguities of the program such as to distinguish cluster series from polymers. The tool will be developed to become an instrument, being helpful in proportion to the expertness of the user.

The up to now purely 'upwards' strategy is supported testing systematically assumptions of chemical connections of found substructures by a routine BREAK.

For a given chemical structure of a known molecule with assumed, estimated by an entropic desorption model¹², breaking probabilities for each chemical bond type it calculates the to be expected PDMS-spectrum. BREAK uses an algorithm, which allows for any interconnected (netted) chemical structure, not just chain-like ones. BREAK is at present trained by learning about breaking probabilities for known substances under fixed preparation conditions.

For any given chemical structure assumption the resulting spectral prediction is then compared with the experimental spectrum using the subroutine SIMILARITY. It compares spectra using a vector space.

For this vector space we used a *linear norm* (see equ.(1)) to calculate the scalar product. Although this introduces some mathematical intricacies since it is non-Euclidean, one gains two advantages: The norm, or *length* of a spectrum becomes its total number of counts, an informative, observable, experimentally measurable quantity. The search for the fingerprint of a specific substance in a mixed spectrum or a series of spectra becomes more sensitive. For example, in the set of spectra¹³ of Chlorophyll, searching for the tiny occurrence of β -Carotene, the derivative of Chlorophyll, indicating the age of the sediment sample, is reported linearly in accordance to its yield, and not quadratic, as in the Euclidean metric, tending to suppress small fingerprints.

5. Interlaboratory Data Base

In the past, large central Data Bases in many fields have been set up with the aim to provide laboratories worldwide with information to be accessed by postal computer-network. In Mass Spectrometry the most wellknown is the WILEY base, with some 100,000 mass spectra provided to some central hosts by many different laboratories. By applying for a password, and paying for each search in accordance, one can, with some (SQL type) retrieval language retrieve per spectrum the about 10 most conspicuous lines, the names, chemical structure, and the author of the respective spectrum. For intense users one may buy the full data set by means of discettes or CD-ROM, getting regular updates by mail. WILEY contains mostly EI spectra of light substances.

Contemplating first to set out analogously for a PDMS data base, by discussing with those active in the field, and because of the large sets due to larger molecules, better detection of large mass lines and for the often less clearly defined substance nomenclature, we had to revise the strategy coming up with the following concept, for which the programs are finished now.

Each laboratory, active in producing PDMS spectra, transfers those spectra of its own, it wishes eventually to get known to some other outside groups, first to its own main frame host on its user-id there. The information stored for each spectrum is the position, width, yield, asymmetry,.. (as far as known) of each relevant line,

data being extracted by the above mentioned analysis programs). In addition a fairly large amount of side information is added, about the sample, its preparation, the method of adsorption, or the machine type. Finally it is added, who is the author of the spectrum, and to what subset of laboratories does the author free the spectrum to be handed out in case of request. The method of storing is provided by a program of us, using the ORACLE Data Base, which runs on many different operating systems such as MVS, VM, UNIX or even DOS, and hardware, let it be a main frame, workstation or PC.

Each user of this decentralized distributed net of PDMS data can now search for spectra or its side information by using the now public research computer network, *bitnet*. Although this is somewhat slow, sometimes, it is free of charge for scientific laboratories. The search is achieved in that the program sends out queries to the respective laboratories, prepares answers, if the data are found and freed, transfers the data into the standard data format JCAMPS-DX, which is proposed internationally as a future standard for MS data transmission between laboratories, and sends the messages off. Due to the pre-evaluation of the channel spectrum for the relevant line information and packing of the data before transmission the network lines will not be overloaded. The eventually incoming answers come into the user number owner's mail box only. Thus by no means the external sender can enter the user number itself and do something unwanted there, even not by mistake. The searcher, when scanning his readerlist or mail box can then pick the data which came in, read them, before deciding whether to incorporate them into his user number. But then he can incorporate it easily into his own ORACLE stored data and do e.g. internal SIMILARITY searches or display the imported data on his own usual means.

For the system operator, programs have been written to allow easy updates and changes of the data base frame to keep up with the changing demands of the community.

The local machine and software requirements are at the moment, besides an access to the BITNET network, a host, capable to run ISPF and REXX such as all machines with either VM or MVS operating systems. Locally, the same working surface can be achieved with the PC-versions SPFPC, Personal REXX, and PC-ORACLE. Thus after installation the customer has to have no knowledge of these tools but just to follow the menu driven requirements using the SQL retrieval language for his questions to other laboratories.

The EMS-NET package writing has been finished, trial runs are planned for spring between ORSAY and the T. H. Darmstadt, with Y. Le Beyec and K. Wien being engaged from the beginning in getting the tools to become useful for future users.

We are very thankful for the many laboratories who have encouraged us by supporting the idea of setting up a concept for a PDMS data base.

We hope that with this new PDMS Transfer System the scientific communica-

tion between the PDMS laboratories will be served. The transfer system presented here may be useful in many other fields as well because of its independency of the structure and type of the transferred data itself.

The discussions with participants at this workshop resulted in the insight, that a transfer system would be much preferred as compared to a centralized big data base, due to the higher flexibility, the control for each laboratory on its own data, and preferably a direct communication between the laboratories instead of via a mammoth central data base, — a net adds a dimension of communication.

6. Acknowledgements

The large scale molecular dynamics calculations were mainly done at the Computer Centre of the GSI Darmstadt, known for its excellent service, using a fixed line paid by the German State of Niedersachsen. The calculations were started when one of us, (E. R. H.) was at Orsay and it profited by the continuous fruitful collaboration with Y. Le Beyec, S. Della Negra and K. Wien.

This work was stimulated by the continuous pointing to the wishes of an eventual practical user by our experimental colleague W. Tuszynski and his collaborators, R. Angermann and F. Hillmann.

The State of Niedersachsen decided to found the proposed ICBM, the Institute of Chemistry and Biology of the Marine Environment at our University, which could serve as a future frame and host to our group.

References

1. H.F. Kammer, *Continuum Mechanical Model for Heavy Ion Induced Desorption*, in: *PDMS and Clusters*, Eds.: E.R. Hilf, H.F. Kammer, and K. Wien; (Publ.: Springer); *Lecture Notes in Physics* 269 (1987), 107.
2. R.D. Macfarlane, B. Wolf and D. Bunk; this volume, page 10.
3. R. Moshhammer; Dissertation, T.H.Darmstadt, in preparation (1990).
4. K. Wien; this volume, page 82.
5. B. Nitzschmann; Diploma Thesis, (Universität Oldenburg, FRG) in preparation (1990).
6. K. Varmuza, H. Lohninger; *Principal Component Analysis of Chemical Data*, in *PCs for Chemists*, Eds.: D. Hazdi, J. Zupan; Elsevier, Amsterdam (1990).
7. D. Henneberg, et al.; in *Software Entwicklung in der Chemie 2*; *Proceedings Computer in der Chemie* Ed. J. Gasteiger; Springer (1988).

8. E.R. Hilf, R. Wolff; *Gross Properties of Nuclear Density Distributions*, report CERN-81-09 (1981), 805.
9. M.H. Wendel, E.R. Hilf; *Journ. Math. Phys.* 26 (1985), 1570.
10. D.J. Gingras; *Use of Homomorphic Signal Processing Techniques for the Estimation of Absorbance Spectra as encountered in Fourier Transform Spectroscopy* SPIE, 975 (1988), 154.
11. E. Härtter; *Wahrscheinlichkeitsrechnung, Statistik und mathematische Grundlagen* (Vandenhoeck u. Ruprecht) (1987), 57.
12. E.R. Hilf, F.Kammer, B.Nitzschmann; in *The Analysis of Peptides and Proteins by Mass Spectrometry*, Ed.: C. McNeal; Wiley (1988).
13. W. Tuszynski, R. Angermann, F. Hillmann, K. Maier-Schwartz; this volume, page 131.