To be submitted to to
Nuovo Cimento

ISTITUTO NAZIONALE DI FISICA NUCLEARE
Laboratori Nazionali di Frascati

LNF-82/52(P)
10 Luglio 1982

Lucia Incoccia and Settimio Mobilio:

THE EFFECT OF STATISTICAL NOISE ON STRUCTURAL PARAMETERS IN EXAFS DATA ANALYSIS

Servizio Documentazione
dei Laboratori Nazionali di Frascati

INFN - Laboratori Nazionali di Frascati
Servizio Documentazione

LNF-82/52(P)
10 Luglio 1982

THE EFFECT OF STATISTICAL NOISE ON STRUCTURAL PARAMETERS IN EXAFS DATA ANALYSIS

Lucia Incoccia and Settimio Mobilio
PULS - Laboratori Nazionali di Frascati, Frascati, Italy

ABSTRACT

We analyze the effect of the statistical noise on the physical parameters derived from an EXAFS spectrum. Two separate classes of experimental cases are discussed, namely heavy and light backscatterers. The weight of the noise in the two cases is in fact quite different, depending on the k range useful for the analysis of the data. We find that for heavy backscatterers the statistical noise can usually be neglected, while for light backscatterers it can be one of the main limitations in obtaining reliable results from EXAFS.

I. - INTRODUCTION

The consideration of the reliability of an experimental techniques is one of the major concerns of every experimentalist. In particular, as any measurement is affected by a certain error, it is very important to have a clear idea of how those errors reflect upon the physical quantities derived from that measurement. This is especially true when the process of extracting information from the data is long and intricate and requires a great amount of computer work. This is just the case of EXAFS, a technique widely and intensively used to obtain structural information on a variety of systems. The rapid growth of interest toward this spectroscopy in the recent years was due mainly to the development of the single electron theory that explained successfully the major characteristics of the spectra by means of a simple formulation and the use of Synchrotron Radiation, that renders the experiments very easy and quick in most cases. All this caused a great enthusiasm towards EXAFS and an enormous amount of experimental work in the field, but also a tendency to understimate the intricacies of the data analysis and the importance of the experimental errors. Consequently, many experiments were done in the past without taking the proper care to avoid high errors. Recently an increasing attention has been devoted to this problem and some critical discussion has appeared in the literature, examining the most important of the so called systematic errors in an EXAFS experiment: thickness effect[3], sample inhomogeneities[4], monochromator's resolution[5], glitches[6].

All these works demonstrate that the experimental accuracy is vital to extract meaningful information from the data, especially if one is interested in coordination numbers or thermal factors. There is however another aspect of the problem that has not been considered in detail until now, whose effect can be dramatic, that is the

statistical noise. The criteria to optimize the signal-to-noise ratio[2], in fact, by calibrating the sample thickness and the detection system are not always applicable, and one is often forced to be satisfied with noisy data. The question is then: what informations can be drawn from an experiment affected by a certain noise and how reliable are they? The purpose of the present work is to answer that question and to evaluate the errors made in the determination of distances, coordination numbers, and thermal factors.

For the sake of completeness, we examined two typical examples of inorganic and organic systems; as we will discuss in what follows, the two cases are quite different and require a separate analysis.

In Section II we describe the method followed in our work and we report the results in Sect. III. Sect. IV is dedicated to the discussion of the results and to the consequences they have in the analysis of the data.

2. - DESCRIPTION OF THE PROCEDURE

Broadly speaking, the experimental cases one usually faces can be roughly divided into two classes: heavy backscatterer (atomic number $\geq$ 20) and light bascatterer (atomic number $\leq$ 10). This somewhat artificial division, although neglecting the intermediate cases, is the most appropriate to discuss the effect of the statistical errors, and is well representative of a large part of the EXAFS work done in the last years. In fact the weight of the error on the physical quantities derived from an EXAFS spectrum is very different in the two cases, depending on the k range available. For example, when studying biological samples, the most common atoms surrounding the absorbing one are O, N, C, and the k range useful usually very short ($\Delta k = 3 \div 8$ Å^{-1}). The other important point is the concentration of the atomic species under study in the sample. In an absorption spectrum, the percentage of the total signal due to the one atom of interest is related to its concentration; this means that, all other experimental conditions being the same, a less concentrated sample has an EXAFS spectrum with higher noise. For all these reasons, we will discuss separately the two cases of low and high atomic numbers, and with error magnitudes typical of concentrated and dilute samples respectively.

The basic procedure adopted was as follows: a theoretical spectrum was generated by using the EXAFS formula:

$$\chi(k) = \sum_j N_j f_j(k, \pi) \frac{e^{-2\sigma_j^2 k^2}}{k R_j^2} \sin(2 k R_j + \delta_j(k)) \qquad (1)$$

To each point of the above spectrum a statistical error was added, given by the random generator routine of a PDP 11/34. Two distributions of the error around the "true" value of the data were considered: uniform and gaussian (Fig. 1). We anticipate here that the distribution of the error doesn't affect at all the results: what is

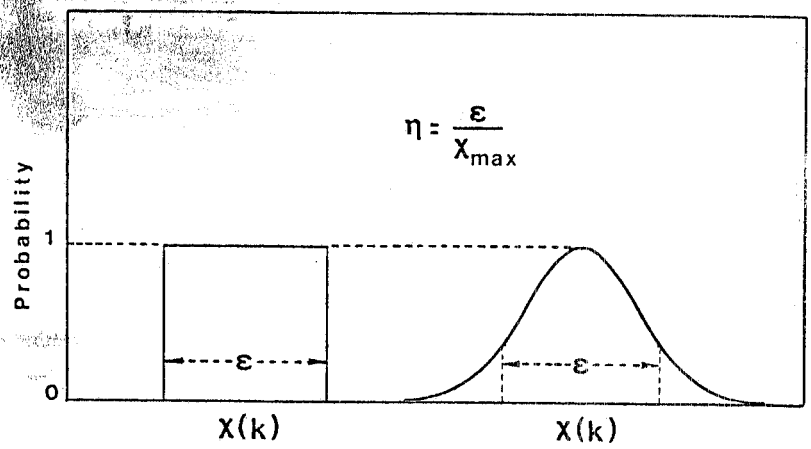

FIG. 1 - Error distribution around the true value of $\chi(k)$: a) uniform; b) gaussian.

important is only the absolute value of the error with respect to the data. For the sake of clarity, then, we report in this paper only the results obtained using the uniform distribution.

The spectrum was then Fourier transformed, and the position and intensity of the peaks in the F.T. magnitude recorded. The first peak in the F.T. was back-transformed into the k space, and the resulting amplitude compared with the one of a model compound. The logarithm of the ratio between the two amplitudes, given by:

$$\ln\left[A(k)/A_m(k)\right] = \ln\left[\frac{N R_m^2}{N_m R^2}\right] - 2(\sigma^2 - \sigma_m^2)k^2 \qquad (2)$$

was fitted to a straight line. From the slope and the intercept at $k^2=0$ of this line N and $\Delta\sigma^2$ were determined. The model compound was built in the same way as the "unknown" but with a smaller value of σ^2 and very low noise. In this way all the uncertainties connected with the proper choice of the model[4] (amplitude transferability) were avoided, as well as the errors due to different numerical treatment of the two spectra, and it was possible to study the effect of the statistical error alone.

For each error magnitude considered, this procedure was repeated 1000 times, thus obtaining a certain distribution of the results. The mean, $\overline{X}$, of the distribution was then determined and the fractional standard deviation $d = \sqrt{v^2}/\overline{X}$ evaluated, where v^2 is the variance of the distribution.

a) Heavy backscatterer

The theoretical $\chi(k)$ of Eq. (1) was generated in the interval k=3÷14 $\mathring{A}^{-1}$ with a spacing of $\delta k=0.025\ \mathring{A}^{-1}$. The material chosen was pure fcc Cu, and the spectrum was built adding the contributions of 6 shells with the known coordination numbers and distances; amplitudes and phases were from ref. (8), and Debye-Waller factors at room temperature from ref. (9).

In Fig. 2 we show the $\chi(k)$, obtained in this way, together with the magnitude of its F.T. Starting from the theoretical $\delta(k)$, several "experimental" spectra were generated, with percentage errors η ranging from 5% to 40% of the maximum of $\chi(k)$: some examples are shown in Fig.3. They represent almost all the experimental cases usually treated, from the higly concentrate to the very dilute or ill-dimensioned samples. The F.T. was performed on the function $\chi(k).k$ in the interval 3-14 $\mathring{A}^{-1}$, using a Hanning-type window and the resulting F(r) was calculated in R space with a spacing of $\delta R=0.02\ \mathring{A}$.

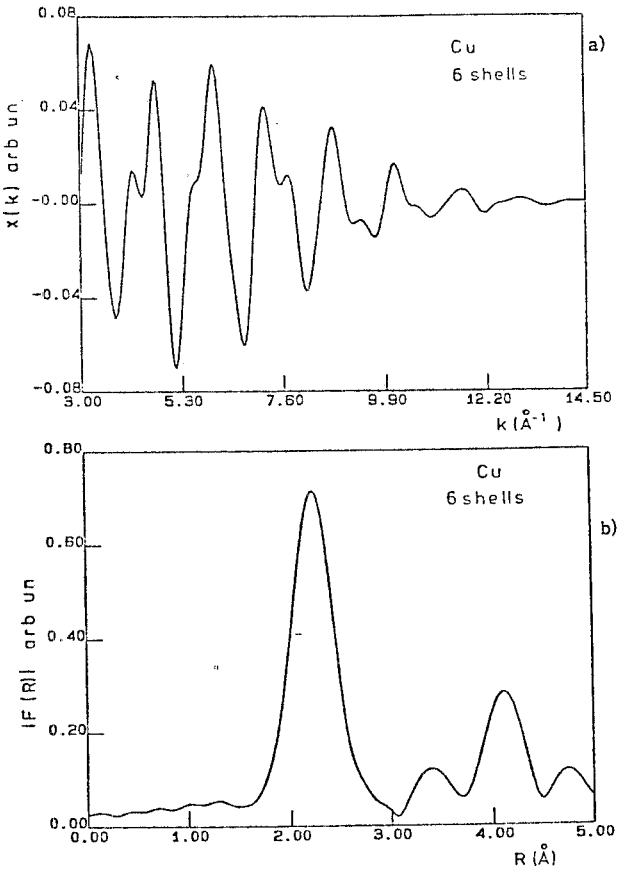

FIG. 2 - a) Theoretical $\chi(k)$ for 6 shells of fcc Cu; b) Fourier transform of the above spectrum in the range 3-14 $\mathring{A}^{-1}$.

The peaks whose position R and intensity Y was monitored were the first and the third ones, placed at R=2.24 Å and R=4.10 Å respectively in the F.T. of the theoretical spectrum. They are both well separated from the other peaks and rather prominent from the background.

In Fig. 4 we report the R and Y distributions obtained for each error η considered, and we summarize the results in Table I.

Beside the above analysis in R space, we investigated also the effect of the statistical error on N and $\Delta\sigma^2$ determination in k space, by comparison with a model compound. The latter was pure fcc Cu at 77 K, with σ^2 factors after ref. (9).

The analysis was made on the first shell to avoid errors due to the side truncation of the peak in R space. As is well known, in fact, the shape of the amplitude of the inverse F.T. is strongly affected by the choice of the range in R space to be back-transformed, if the peak is not well separated from the others. For this same reason, we decided to generate the "experimental" amplitude function A(k) in the following way: for a given error magnitude, η, we Fourier-transformed the error alone, and added the resulting real and imaginary parts to the ones of the theoretical $\chi(k)$ at room temperature. The so obtained F.T. was then back-transformed, keeping fixed in each run the range of inverse tranformation.

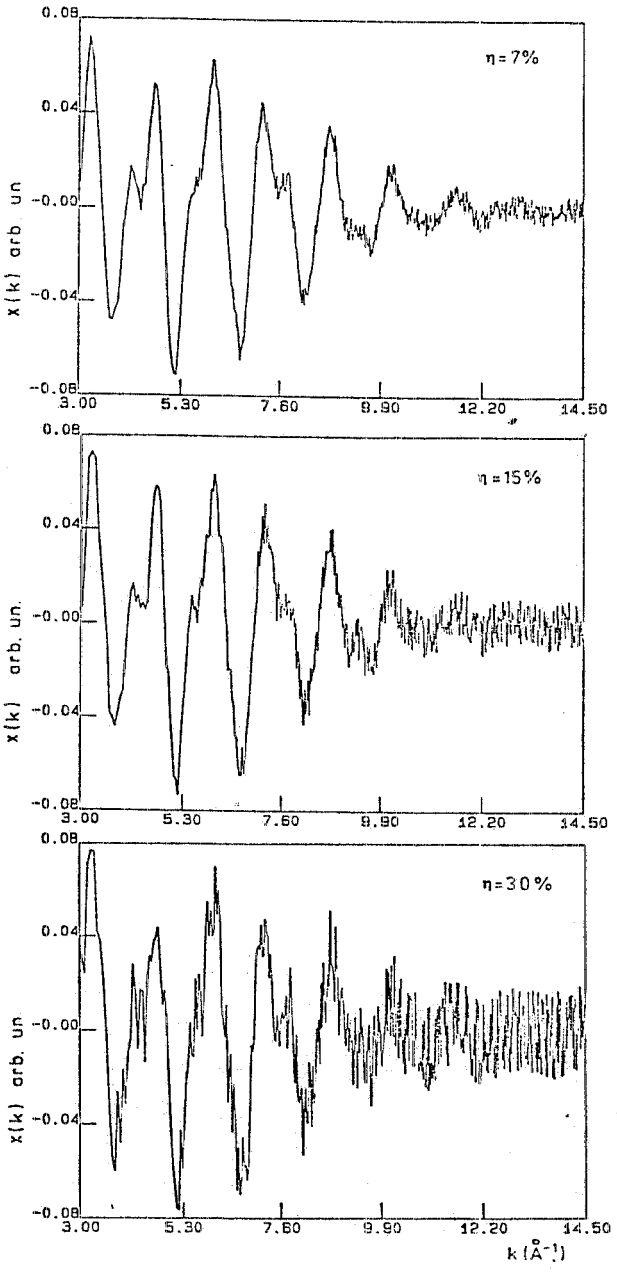

FIG. 3 - Examples of "experimental" $\chi(k)$ of Cu with different noise values.

In Fig. 5 we report an example of the logarithm of the ratio of the amplitudes. From a linear fit in the range 20-140 $Å^{-2}$ we deduced N and $\Delta\sigma^2$ whose distributions are shown in Fig. 6. In Table III the numerical results are reported. For this k-space analysis the maximum percentage error considered was 30%. For higher noise, in fact, the behaviour of ln(A1/A2) was no more linear over a k^2 range large enough to give reliable results.

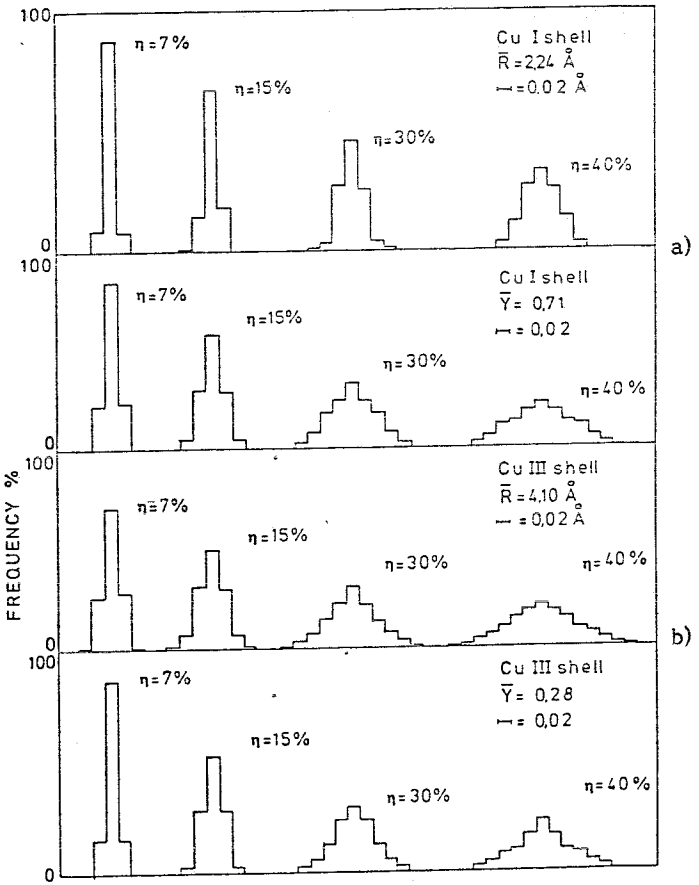

FIG. 4 - Distribution of peak position, R, and peak intensity, Y, for the first (a) and third shell (b) of Cu at different noise values; each distribution was obtained from a set of 1000 iterations.

TABLE I - Fractional standard deviation of the distributions obtained for R and Y for the first and third shell of Cu at different values of the percentage noise.

I shell	$\bar{R}$ = 2.24 Å $\bar{y}$ = 0.71		III shell	$\bar{R}$ = 4.10 Å $\bar{y}$ = 2.28
η(%)	d_R	d_y	d_R	d_y
7	0.003	0.014	0.003	0.032
15	0.005	0.024	0.005	0.058
30	0.007	0.046	0.010	0.105
40	0.010	0.067	0.015	0.140

TABLE II - Fractional standard deviation of the distributions obtained for R and Y for the first and second shell of FeN_6C_8 at different values of the percentage noise.

I shell	$\bar{R}$ = 1.57 Å $\bar{y}$ = 2.54		II shell	$\bar{R}$ = 2.27 Å $\bar{y}$ = 0.690
η(%)	d_R	d_y	d_R	d_y
10	0.020	0.060	0.025	0.223
20	0.039	0.11		
30	0.054	0.140		

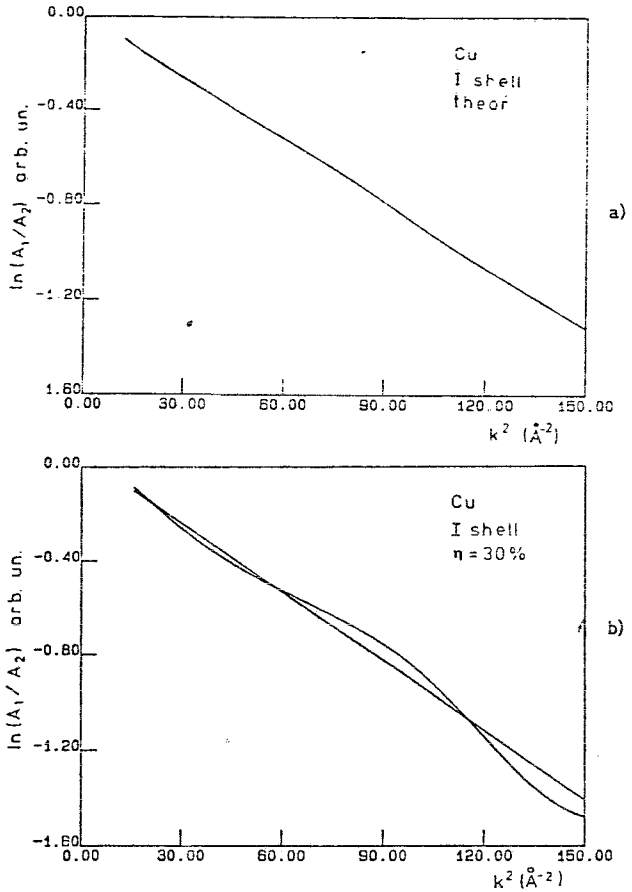

FIG. 5 - Logarithmic plot of the ratio of the amplitudes $\ln(A_1/A_2)$ for the first shell of Cu; the model compound is Cu at 77 K. a) theoretical, b) experimental, obtained with a value of η of 30%. The straight line is the result of a fit in the range k^2=20-140 $Å^{-2}$.

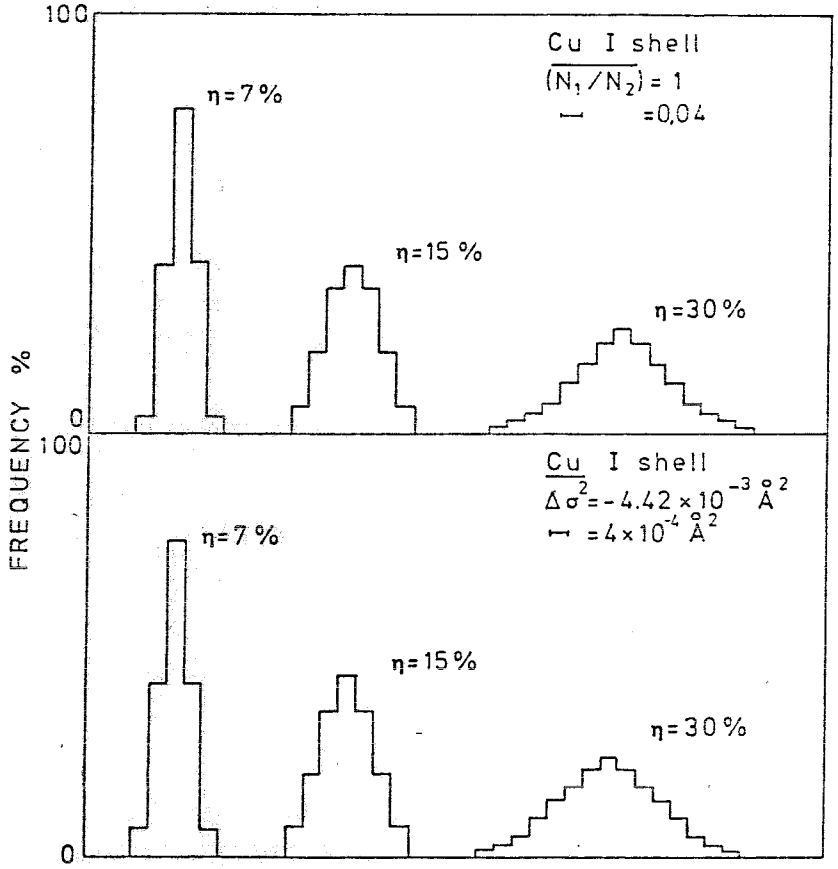

FIG. 6 - Distributions of coordination numbers, N, and thermal factor, $\Delta\sigma^2$ obtained in the k-space analysis of the first shell of Cu for different values of the noise.

b Light backscatterer

The theoretical $\chi(k)$ used in this case was the one of a fictitious molecule, FeN_6C_8, with a Fe atom surrounded by 6 N atoms at R=2.05 Å (I shell) and 8 C atoms at R=2.78 Å (II shell). The amplitudes and phases were from ref. (8), and the Debye-Waller factors were σ^2=0.005 Å^2 at 300 K and σ^2=0.001 at 77 K for both shells. The theoretical spectrum (Fig. 7) was generated in the range 2.5-12 Å^{-1}, with a spacing k=0.025 Å^{-1}. The

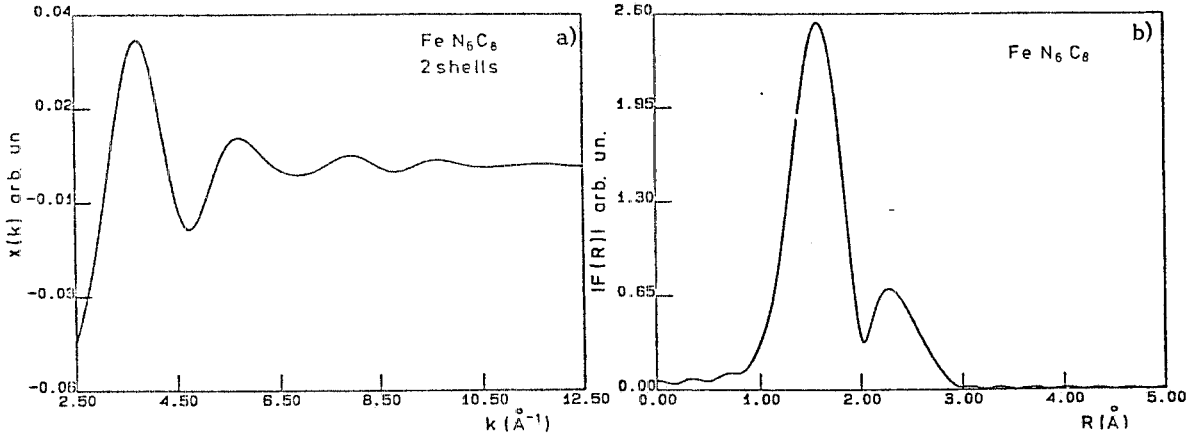

FIG. 7 - a) Theoretical $\chi(k)$ of the FeN_6C_8 molecule; b) Fourier transform of the above spectrum in the range 2.5-12 Å^{-1}, obtained using a gaussian window and a k^3 weighting factor.

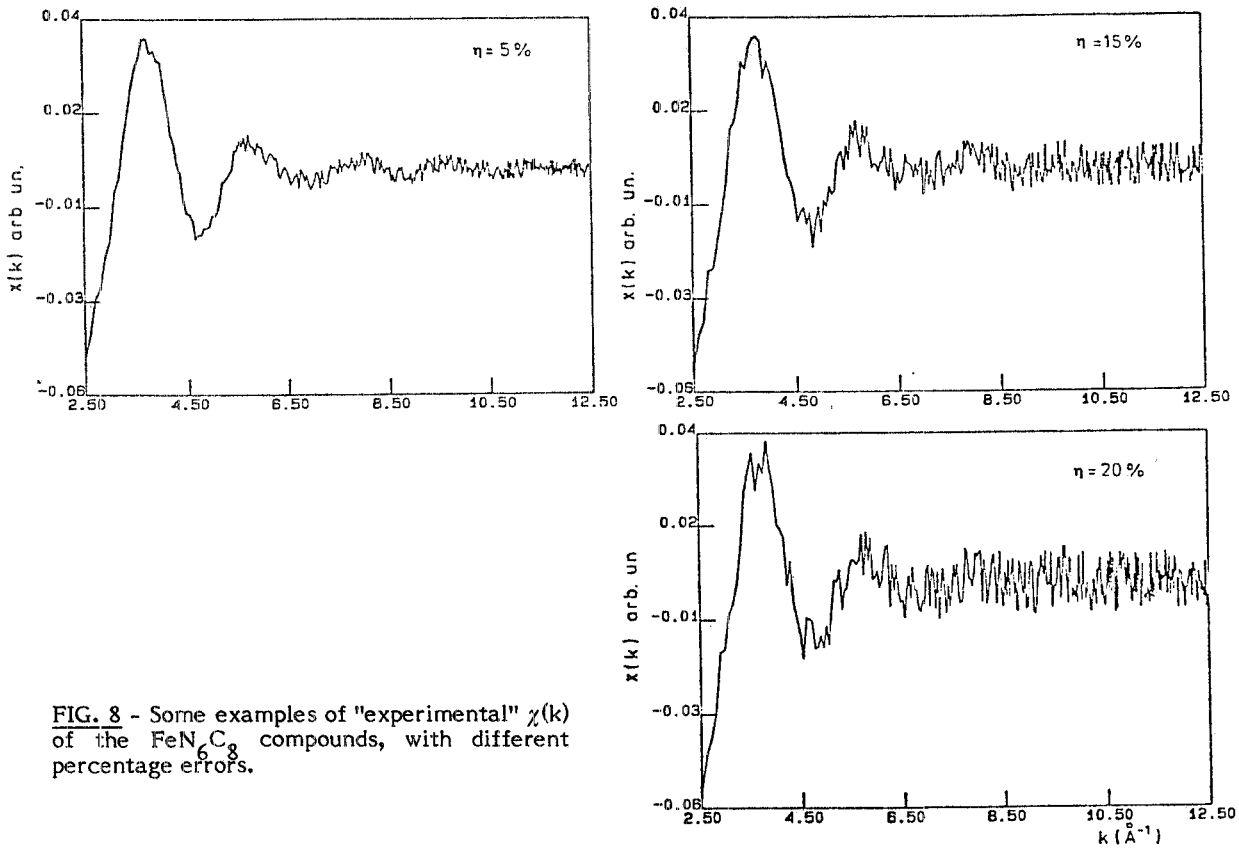

FIG. 8 - Some examples of "experimental" $\chi(k)$ of the FeN_6C_8 compounds, with different percentage errors.

procedure followed to simulate the experimental spectra (Fig. 8) and to process them was the same as in the case of the heavy backscatterer; some differences, like the use of a gaussian window and a k^3 weighting factor in the F.T. were introduced as more appropriate in the analysis of a light backscatterer. The error magnitudes considered were of course smaller in this case than in the preceding one, due to the rapid decay of the backscattering amplitude of the N and C atoms. An example of the logarithmic plots obtained in k space is shown in Fig. 9; the linear fit was made in the range 20-100 $\AA^{-2}$. Fig. 10 and Fig. 11 illustrated the R and k space analysis, while Tables II and IV summarize the results.

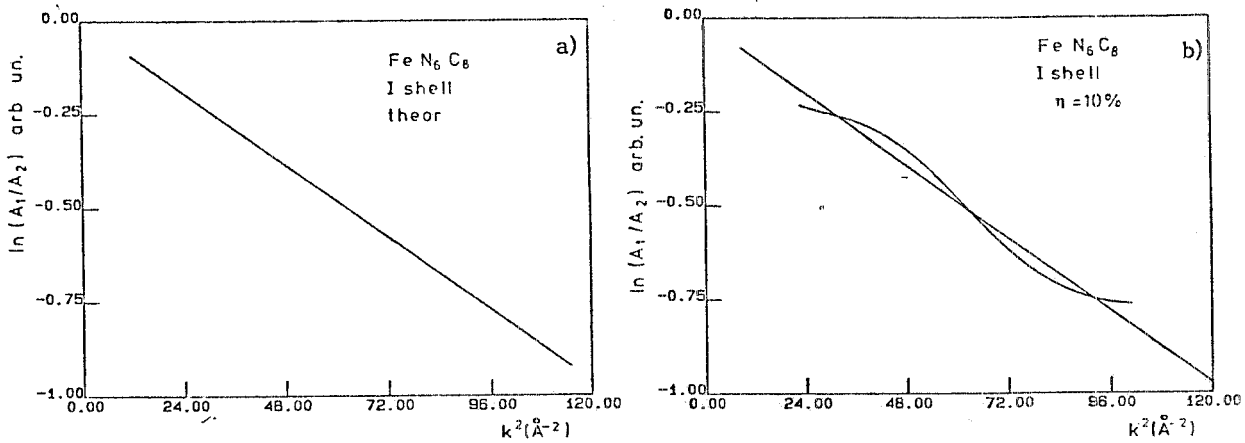

FIG. 9 - Logarithmic plot of the ratio of the amplitudes, $\ln(A_1/A_2)$ for the first shell of FeN_6C_8; the model compound was the same molecule with σ^2=0.001 $\AA^2$. a) theoretical; b) experimental, obtained with a noise value η_o=10%; the straight line is the result of a fit in the range Δk^2=20-100 $\AA^{-2}$.

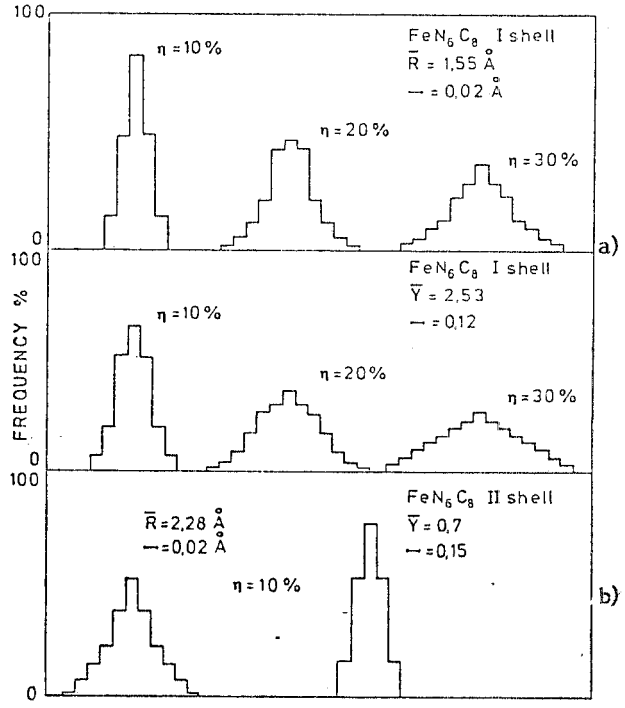

FIG. 10 - Distributions of peak position R, and peak intensity Y of the first (a) and second (b) shell of the FeN_6C_8 molecule.

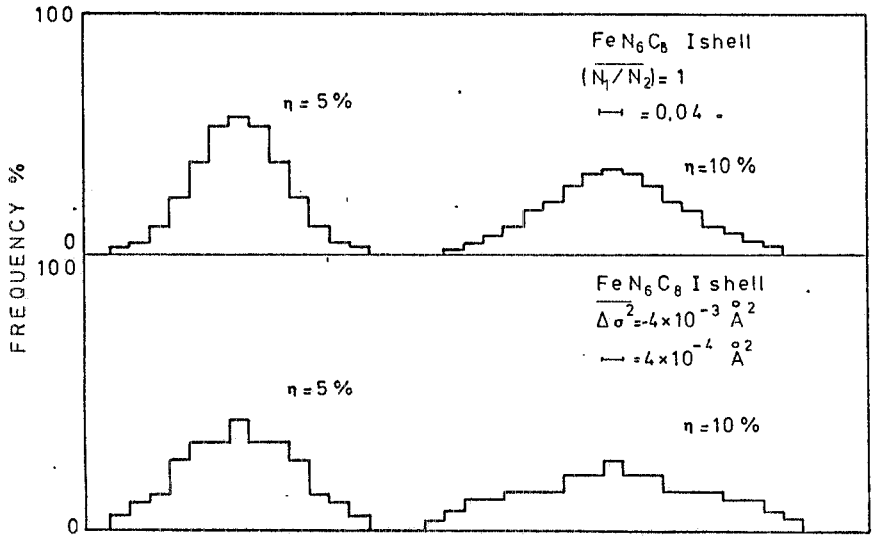

FIG. 11 - Distributions of coordination number N, and thermal factors σ^2 for the first shell of FeN_6C_8.

TABLE III - Fractional standard deviation of the distributions obtained for N and $\Delta\sigma^2$ for the first shell of Cu at different values of the percentage noise.

Cu I shell	$\overline{N_1/N_2} = 1$	$\overline{\Delta\sigma^2} = 0.0042$ Å^2
η(%)	d_N	d_σ
7	0.032	0.071
15	0.063	0.138
30	0.123	0.264

TABLE IV - Fractional standard deviations of the distributions obtained for N and $\Delta\sigma^2$ for the first shell of FeN_6C_8 at different values of the percentage noise.

FeN_6C_8 I shell	$\overline{N_1/N_2} = 1$	$\overline{\Delta\sigma^2} = 0.004$ Å^2
η(%)	d_N	d_σ
5	0.10	0.30
10	0.18	0.54

III. RESULTS

a) R-space analysis

Heavy backscatterers

In Table I are reported the results obtained on the first and third shell of Cu in different cases. As can be seen, even for errors as high as forty per cent, the determination of the distances is as good as 0.02 Å for the first shell and 0.05 Å for the third. This is a very encouraging result showing that also for very noise measurements the determination of distances is fairly accurate. This is related to the large k-range useful for the analysis and consequently the large number of zero crossings. In fact, the value of the distances are mainly determined by the zeroes of the oscillations: the errors on the positions of these zeroes due to the noise are statistically well averaged. It is also to be underlined that the greater effect of the error on the position of the third shell is essentially due to the loss of resolution of the corresponding peak in the F.T. This can be understood from Fig. 12a, where the F.T. of the errors, shown on the same scale of the F.T. of the theoretical spectrum, is negligible with respect to the first peak but not to the third one. Also the intensities of the peaks are well determined in these cases of heavy backscatterers, as long as the peaks are well resolved. This implies that it could be possible to extract reliable informations also from the width, the area and the height of the peaks.

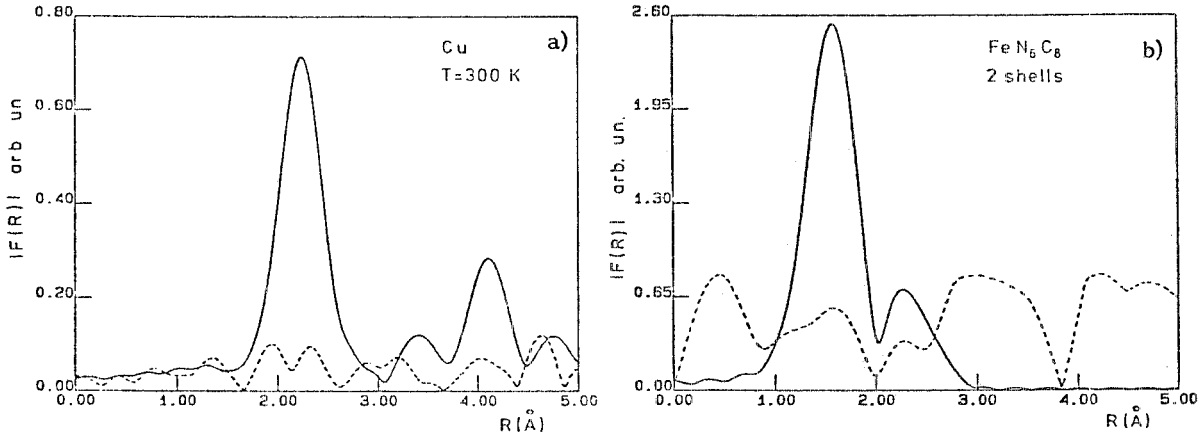

FIG. 12 - a) Comparison between the F.T. of the theoretical spectrum of Cu and of the noise (dashed line) at a noise value of 20%; b) Comparison between the F.T. of the theoretical spectrum of FeN_6C_8 and of the noise (dashed line) at a noise value of 20%.

Light backscatterers

Table II reports the results obtained on the simulated spectrum of light backscatteres. In this case one has to be more careful since an error of 15% is enough to produce errors of the order of 0.05 Å in the first shell distance and of 15% in the peak height. Indeed for the light backscatterers the amplitude of the EXAFS spectrum dies very soon ($k_{max}=6\div7 Å^{-1}$) causing a small number of zeroes in the oscillations. The larger effect of the noise in this case, especially on the peaks amplitude, is due to the fact that the maximum of the EXAFS spectrum is at the beginning of the k range transformed, where one is forced to use a window function that smooth the data to zero. Another point is that for a light backscatterer one usually weights the data with a k^3 factor, which enhances the noise at high k values.

As for the second shell, a noise greater than 10% worsens dramatically the resolution in R-space making it impossible to get reliable informations from F.T. since the intensity of the F.T. of the second peak and of the noise are comparable (Fig. 12b). One could think of backtransforming both peaks together in k-space and fitting

the resulting spectrum with two shells. Even with this method, the results are invalidated by the noise since one has transformed also the frequencies of the noice under the two peaks.

b) k-Space Analysis

Heavy backscatterer

Table III shows the results obtained on N and $\Delta\sigma^2$ by the analysis in k-space of the inverse transform of the first peak of Cu. The first thing to note is that the fractional mean deviations on N and $\Delta\sigma^2$ are acceptable even for highly noisy spectra. The inference that can be made, consequently, is that in the case of heavy backscatterers the statistical noise is not the main reason for the high errors in coordination numbers and Debye-Waller factors often reported in the literature. More probably, these are due to a bad choice of the model compound or to truncation effects both in the F.T. and in the inverse F.T.

The results of Table III show also that the fractional standard deviation (f.s.d.) of $\Delta\sigma^2$ is of the order of the percentage noise, while the f.s.d. of N is roughly one half of this value.

A point to stress is that, in any case, they are both considerably greater than the errors made in the determination of the height of the peaks in the F.T. This implies that a derivation of N and σ^2 directly from the F.T. should be preferable, since it is less affected by the noise.

Light backscatterer

Table IV shows the result in the case of the FeN_6C_8 molecule. Here the situation is much worse than in the preceding case since even with a noise as low as 5%, which means a fairly good spectrum when working with biological samples, one risks errors of the order of 50% on $\Delta\sigma^2$ and of 20% on N. It is evident, then, that the statistical noise is the main source of errors for light backscatterers.

The reason is again the rapid fall of the backscattering amplitude as a function of k: a great part of the physical information lies in the region $k \lesssim 3$ $\mathring{A}^{-1}$, which is lost in the EXAFS analysis. This is more so as the values of σ^2 of the shell examined increases, causing a further "compression" of the photoelectron wave amplitude towards low k values. Consequently, the evaluation of the effect of the noise on the quantities one measures from EXAFS must take into account also an estimate of the actual σ^2 factor of the system. On the other hand, a comparison between Tables II and IV shows that again the measure of the peak height in R space is more accurate. This can be understood from the fact that, while in the first Fourier-transforming process the frequencies of the noise are spread over the whole k spectrum, in the inverse one only the frequencies of the noise underlying the shell of interest are selected.

IV - CONCLUSIONS

We have discussed in this paper the effect of the statistical noise on the analysis of EXAFS spectra both in R and k space. It has been shown that when the physical content of the spectrum is spread over a large k-range ($\Delta k \gtrsim 10$ $\mathring{A}^{-1}$), as it is for a heavy backscatterer, bond-length informations can be obtained with excellent accuracy even with high noise; as for coordination numbers and Debye-Waller factors, the errors induced by the noise are quite low: this means that systematic experimental errors, a bad choice of models, or problems connected with numerical processing of the data dominate and should be the main concern of the experimentalist. The situation is different with light backscatterers, where the physical information of a spectrum is concentrated at very low k values. As discussed in the preceding sections, the consequence is a strong sensitivity of the EXAFS results to the statistical noise. This means that besides the usual cautions with model compounds or data processing one has to additionally be aware of the effect of the noise. This effect is

large in bond length determination, and becomes dramatic when measuring coordination numbers and $\Delta\sigma^2$ factors by the logarithm method in k-space.

A possible alternative could be to extract the informations from the F.T. alone, which we showed to be much less affected by the statistics. This approach has some inherent difficulties: indeed, since the spectrum in R space is the convolution of the pair distribution function with the F.T. of the backscattering amplitude, there is no general way to relate the width and intensity of the peaks with N and σ^2[10]. To apply this method then it is necessary to know the backscattering amplitude to deconvolute the pair distribution function from the F.T., as in the usual k space analysis. Efforts to extract N and σ^2 from the spectrum in R space are in progress[11].

In conclusion results of the present work strongly indicate that great care must be taken to keep the noise low if one is to measure the physical quantities from EXAFS with any significance.

REFERENCES

(1) P.A. Lee, P.H. Citrin, P. Eisenberger and B.M. Kincaid; Rev. Mod. Phys. 53, 769 (1981) and references therein.
(2) See, for ex.: "Synchrotron Radiation Research" Edited by W. Winick and S. Doniach, Plenum Press, New York (1980)
(3) E.A. Stern and K. Kim, Phys. Rev. B23, 3781 (1981).
(4) E.A. Stern in: Proc. of the Daresbury Study Week-end, 28-29 March (1981), DL/SCI/R17; J. Goulon, C. Goulon-Ginet, R. Cortes and J.M. Dubois, to be published.
(5) B. Lengeler and P. Eisenberger, Phys. Rev. B21, 4507 (1980).
(6) F. Comin, L. Incoccia and S. Mobilio; Frascati report LNF-82/11 (1982), to be published on J. Phys. E.; E.A. Stern and Kun-quan Lu in: Proc. of the Second Int. Conf. on Synchrotron Radiation Instrumentation, Ithaca, New York, July 15-17 (1981).
(7) E.A. Stern, D.E. Sayers and F.W. Lytle; Phys. Rev. B11, 4836 (1975).
(8) B.K. Teo and P.A. Lee, J. Am. Chem. Soc. 101, 2015 (1979).
(9) E.Sevillano, H. Meuth, J.J. Rehr, Phys. Rev. B20, 4908 (1979).
(10) R.B. Greegor and F.W. Lytle, Phys. Rev. B20, 4902 (1979).
(11) S. Mobilio and L. Incoccia, these Proceedings.