

A FLEXIBLE PROCEDURE FOR ANALYZING IMPEDANCE SPECTROSCOPY RESULTS: DESCRIPTION AND ILLUSTRATIONS

J. Ross MACDONALD and Larry D. POTTER, JR.

Department of Physics and Astronomy, University of North Carolina, Chapel Hill, NC 27514, USA

Received 30 January 1987; accepted for publication 3 March 1987

General considerations in nonlinear least squares fitting of small-signal ac frequency response data (conductive or dielectric systems) are first discussed. A very general and flexible complex nonlinear least squares fitting (CNLS) program (LOMFP) is described which runs on microcomputers of the PC and AT type. Advantages and disadvantages of LOMFP are discussed and compared to those of a recent CNLS program developed by Boukamp. LOMFP incorporates many thousands of equivalent circuits fitting possibilities and ten different distributed circuit elements which can be used in these circuits. Its fitting is both speedy and accurate, and it provides considerable insurance against confusing local minima fits with true least squares fits. It is optimized for fitting real, imaginary, or complex data of either partly conducting or purely dielectric character, and it incorporates great flexibility in handling and fitting data in different forms (different immittance levels and rectangular or polar representation). Seven different data weighting choices are provided. Exact synthetic data were generated from a typical Impedance Spectroscopy equivalent circuit, and the effects on CNLS fitting were investigated of rounding the exact data to 4, 3, and 2 decimal places. We present the results of an extensive study of the effects of transforming exact data to different immittance levels and representations and then rounding to two places or of transforming rounded 2-place data directly. We conclude that real data should not be transformed from its original measured form before carrying out CNLS fitting. Finally, the effects were investigated of various kinds of weightings on CNLS fitting of 2-place data, leading to the conclusion that in spite of its advocacy and use by Zoltowski and Boukamp, modulus weighting generally yields misleading and appreciably worse CNLS fitting results than does proportional weighting. The LOMFP program (source and executable files) is available from the first author.

1. Introduction

Impedance spectroscopy (IS) is a fancy modern name for the small-signal ac measurement and analysis of electrical response data over an appreciable span of frequency, a span currently ranging from as low as about 10^{-5} Hz to as high as 10^7 Hz or even higher. By this broad definition it includes all such measurements on either dielectric or somewhat conducting materials and so has a rich history extending back for more than fifty years.

In the last thirty years, the use of IS for the analysis of both liquid and solid electrolytes has grown substantially. And, in the last 10 to 15 years, it has come to be even more widely employed because, we believe, of the following factors: (a) first, because of the development and availability of computer-based automatic measuring equipment; (b) second, because of an increasing number of demonstrations of its usefulness in analyzing data for a wide variety of

materials; and (c) last, because of the development of powerful and time-saving methods of data presentation and analysis. These developments have made it much easier to obtain IS data and to analyze it adequately. Here we shall consider only factor (c), the most recent of the three; the one whose use is still not as completely widespread as it deserves to be; and the one which can contribute most to the understanding of the electrical properties and behavior of material-electrode systems. As Socrates might have said if he were a modern experimentalist: "The inadequately examined data set is not worth generating". The ultimate purpose of nearly any experimental procedure is "insight, not numbers" [1].

The designation "impedance spectroscopy" should really be replaced by "immittance spectroscopy", because IS can and should deal with all four common immittance functions: impedance, $Z = Z' + iZ''$; admittance, $Y = Z^{-1} = Y' + iY''$; complex capaci-

tance or dielectric constant, $C^* = Y/i\omega = C' - iC''$ (or $\epsilon^* = C^*/C_c = \epsilon' - i\epsilon''$, where C_c is the capacitance of the empty cell); and (complex) modulus, $M_c = C^*{}^{-1} = i\omega Z = M'_c + iM''_c$ (or $M = \epsilon^*{}^{-1} = M' + iM''$, a dimensionless complex quantity.) As usual, $i \equiv \sqrt{-1}$. IS should deal with all these quantities, even when measurements are made at only one of these four levels, because they weight the data differently and much can usually be learned by examining two and/or three dimensional plots [2] of a given data set for each of these levels [3,4]. Because there have been several recent reviews that deal wholly or in part with IS [5–7], it is unnecessary to discuss it in general any further here. Instead, we shall deal only with the analysis component of IS, specifically with complex nonlinear least squares (CNLS) data fitting.

2. CNLS fitting – general

When a single, simple Debye relaxation process is all that appears in experimental IS results, graphical analysis of the imaginary part of the immittance function involved, plotted versus frequency, or the logarithm of frequency, suffices to allow approximate estimates of the values of the relaxation parameters to be obtained. Usually, however, more than one process, none of them necessarily simple, combine to yield the experimental response. Graphical and subtractive methods have been used in the past to analyze such data, but they fail when the effects of the processes overlap appreciably in time and frequency. Further, such approaches, even when applicable, yield no information on the uncertainties of the estimated parameters.

An appreciable advance was the use of nonlinear least squares (NLLS) to analyze the imaginary (or possibly real) part of the frequency response. But when such analysis is done independently for the two parts of the response, one obtains separate parameter sets whose values often differ appreciably for some or all of the parameters involved. Although the real and imaginary parts are related by the holistic Kronig–Kramers relations when no errors are present, there will always be random errors involved, as well as possibly some systematic ones. Analysis of the imaginary response alone has been much

employed in the dielectric area and is still used for conducting systems as well, such as ion-conducting polymers [8]. Unfortunately, it is often not stated how the fitting was actually accomplished in such work.

By simultaneous least squares fitting of the real and imaginary parts (or modulus and phase) of data to a suitable model, one uses all the data, tends to average out independent errors, usually reduces statistical uncertainties, and obtains a single consistent set of model parameter estimates, say Q_j , for the system, as well as estimates of their standard deviations σ_{Q_j} . Such CNLS fitting, which adjusts all parameter values simultaneously, generally yields very high resolution and can thus resolve processes with strongly overlapping time constants. Further, estimates of the σ_{Q_j} s (here designated as S_{Q_j} or simply S_j) allow one to evaluate which parameters are important and which, if any, are unimportant to the fitting.

The first complete CNLS approach for the present area was that of Macdonald and Garber [9]. It included various weighting possibilities, an important feature, as we shall demonstrate later. This approach was further developed and discussed in [10], and the virtues of CNLS fitting and 3-D plotting were demonstrated there. Next, Tsai and Whitmore [11] discussed a CNLS approach very similar to those above but involving only the option of unity weighting. Finally, Boukamp has recently described a simple, approximate subtractive analysis approach [12] as well as a full CNLS one which involves modulus weighting (MWT) [13]. This last approach (the EQIVCT program) incorporates many useful features. It will be compared with the program described below, one which is a substantial generalization and modernization of the earlier work of refs. [9] and [10].

3. Iteration and convergence

Because the parameters to be estimated enter fitting models nonlinearly in NLLS, fitting of such models requires an iterative approach, one which may either fail to converge or converge to a local (not absolute) minimum of the sum of squares. The larger the number of free parameters to be determined, the more likely is such inadequate convergence, but of

course the better the initial guesses for the parameter values the more likely is convergence to the proper *least squares* solution (LSS). Because of the ubiquitous possibility of inadequate convergence in NLLS, it is always desirable to carry out the fitting with two or more separate and quite different initial parameter values sets. If two or more of them converge to the same final parameter set, it is usually reasonable to assume that the least squares solution has been obtained. We shall discuss a further safeguard later in this section. When one is dealing with a CNLS fit, all the above considerations again apply.

In the CNLS situation, the basic problem is to minimize the weighted sum of squares cost or objective function [7,9,10]

$$S_w = \sum_{i=1}^N W_i^R [A_e' - A_t'(\omega_i; \mathbf{Q})]^2 + W_i^I [A_e'' - A_t''(\omega_i; \mathbf{Q})]^2, \quad (1)$$

here written for an arbitrary immittance function $A = A' + iA''$. Here $\omega \equiv 2\pi f$, where f is the frequency. The e and t subscripts denote experimental and theoretical quantities, and the R and I superscripts allow one to discriminate between the weighting used for the real and for the imaginary squared deviations or residuals. Here there are N frequency values, and \mathbf{Q} again represents the set of M free and/or fixed parameters present in the theoretical model, $A_t(\omega, \mathbf{Q})$.

The weights appearing in eq. (1), which are all unity in the "unweighted" or unity-weighted (UWT) situation, are usually defined by $W_i^a = (S_i^a)^{-2}$, where ideally the S_i^a are the experimentally determined, estimated standard deviations of the individual measurements; then W_i^a is the inverse variance. Here $a = R$ or I . Since the S_i^a 's are rarely known accurately and are often unavailable, other choices for them are usually made. Such choices and their consequences will be discussed in detail later. It is worth mentioning, however, that when very accurate data are available, the actual choice of weighting used affects the parameter estimates very little. Unfortunately very accurate data are a scarce commodity.

Nearly all available CNLS fitting programs use the general Levenberg-Marquardt (LM) approach for carrying out the actual minimization of S_w . But various generalizations and modifications of the LM

basic algorithm have been produced. Here we mention briefly the particular features of our fitting program (FP), the LOMFP, with the letters referring to LM, Olson, and Macdonald. More details, and the provenance of the actual minimizers used in LOMFP, are provided in ref. [10].

Because of the convergence problem mentioned above, we generally use two separate CNLS programs in series, making up the total LOMFP. Both operate in double precision mode. the first, LEV, always converges for any reasonable (and often unreasonable!) set of data, model, and initial choices. Although it usually yields the LSS, especially for simple situations, it does not always do so. Its parameter estimates, reduced to eight decimal places, are used as inputs to the second part of the program, a more "delicate" minimizer named OLSON, one which also provides further statistical error estimates when it converges. When the first minimization achieves the LSS or obtains results near it, the second minimizer usually requires only one to three iterations to achieve convergence, taken as the condition for which the relative changes of all parameters from one iteration to the next are each less than XTOL or the relative change in the weighted sum of squares is less than FTOL. These tolerances are often taken as small as 10^{-11} since such small choices add little to the program run time.

Because the second minimizer must invert the system matrix, it does not always converge, even when the first one does. It incorporates various automatic procedures to obtain a converged solution, but we find that it nevertheless usually fails when the first minimization yields a result appreciably different from the LSS. In favorable cases, it does continue to iterate, however, until it achieves an appreciably smaller weighted sum of squares than does the first minimizer. It is likely that this converged result is a LSS, but it should nevertheless always be checked by using a different input parameter estimate set. When the second minimizer does not converge, it is direct evidence that the result of the first minimizer is very likely not a LSS; again different input choices should be used until both minimizers converge, if they do, to essentially the same result, hopefully and likely a LSS.

The LOMFP uses numerically calculated derivative values for both minimizers, unlike the EQVCT

of Boukamp. There are advantages and disadvantages to both approaches. That of Boukamp, using calculated analytical derivatives, possibly speeds up convergence slightly and may lead to a small decrease in execution time. In our program, which incorporates many more distributed circuit element possibilities than does EQIVCT, not all derivatives are available in analytic form. Further, it is a convenience to be able to use the computer to calculate derivative values as needed. When the step size used in the numerical derivative calculations (an input choice) is selected to fall between too large values, which yield inaccurate derivatives, or too small values, where round-off becomes important, accurate derivative values are obtained and little or no penalty in convergence time (and none in the accuracy of the final parameter estimates) is incurred.

4. Strengths and weaknesses of LOMFP and EQIVCT

Although a very important safety and accuracy feature of the LOMFP is its serial use of two quite different minimizers, as discussed in the last section, the heart of its utility is in its flexibility and generality for fitting complicated equivalent circuits which may contain a wide variety of distributed circuit elements [3] (DCE's) (see later detailed discussion). Here again EQIVCT and LOMFP take quite different approaches. EQIVCT provides three distinct DCE's, and as many of them and as many R, C, and L's as needed can be arranged in any geometrical structure by proper input choice. Here ultimate circuit flexibility has been gained at the sacrifice of having available only a limited set of DCE's.

LOMFP involves somewhat the opposite choice. A much larger number of DCE's is available to be used in its circuits, but although these circuits are exceedingly flexible and general and include thousands of possibilities selected by simple input choices, they only allow completely arbitrary interconnection choices of all available circuit elements by adding additional circuit structures and recompiling and/or relinking the program. The details of this flexibility, and its limitations, which appear to be of little practical importance, are discussed later in this work.

There are several other important differences

between the two programs. First, EQIVCT apparently now runs only on a large mainframe computer, while different versions of LOMFP are available for mainframe use, for a Unix-based minicomputer, and even for PC's. The execution-times of typical LOMFP runs, all in double precision, are often less than a minute on AT-type PC's with a 80287 numerical co-processor chip and are slow, but not intolerably so, even on ordinary PC's without an 8087 chip. The total LOMFP source and executable programs, as well as detailed instructions, and, for PC's, many useful batch and transformation files, are all available from one of us (J.R.M.).

Another very important difference is in input and fitting possibilities. Only one EQIVCT choice has been discussed [12,13]. For this choice, the data are in rectangular admittance form, MWT is used, and the fitting is apparently carried out for the data in the same form as that used as input. By contrast, LOMFP allows the input immittance data to be in Z , Y , ϵ (or C^*), or M form, either rectangular or polar. The input frequency values may be in either f or ω form. Actual fitting can involve any of several very flexible weighting choices (including MWT), and the fitting itself may be carried out at the Z , Y , ϵ (or C^*), or M level in either rectangular or polar form, independent of the level and form of the actual input data.

Finally, although CNLS is usually superior to separate fitting of the real or imaginary parts of the data, one of these may sometimes be missing, or, even if not missing, one might like to compare the results of separate CNLS, RNLS, and/or INLS fits. Any of these three types of fits may be readily selected by changing a single character in the LOMFP input file. For some complex data, some of the fit parameters may, in fact, sometimes turn out to be considerably better estimated by NLSS fitting of either the real or the imaginary parts of the data rather than by the full CNLS fit. Thus, the ability to carry out all three types of fitting very easily is a useful feature of LOMFP which can help one determine the best estimates of all the free parameters of the model. The current version of LOMFP is limited to a maximum of $N=100$ data points. Another minor limitation is that currently a fitting circuit may contain up to a maximum of 30 circuit element parameters, with any or all of them being free to vary. Of course in practical cases

one would usually not need so many parameters and would usually have only five to ten of them free. By means of a single input choice, each parameter entering into a given circuit may be specified to be fixed or free. If free, it may be constrained to be positive only or may be either positive or negative.

5. Circuit possibilities

Fig. 1 presents the five basic circuit structures which are currently implemented in LOMFP. The elements marked "DE" are DCE's and will be discussed in detail in the next section. The DAE element in fig. 1d is a distribution of activation energies DCE and will be also discussed later. The L element which appears in all the circuits is included to account

for possible uncompensated wiring inductance in the experimental measuring apparatus. Ionic response in the usual IS frequency range does not include any inductive inertial effects of the charge carriers. We believe that the appearance of inductive-type phase shifts in IS data (often associated with specific adsorption and appearing at low frequencies) does not indicate the presence of true inductance (energy storage in a magnetic field), and it is best represented by the use of negative resistances and capacitances [14], a built-in capability of LOMFP.

Next, note that all circuits include a geometrical capacitance element, C_g or C_1 , which spans the electrodes. Such a connection, rather than one taking C_g across only part of the circuit, is physically necessary [15-17]. It must be noted, however, that many IS measurements, especially those on low impedance

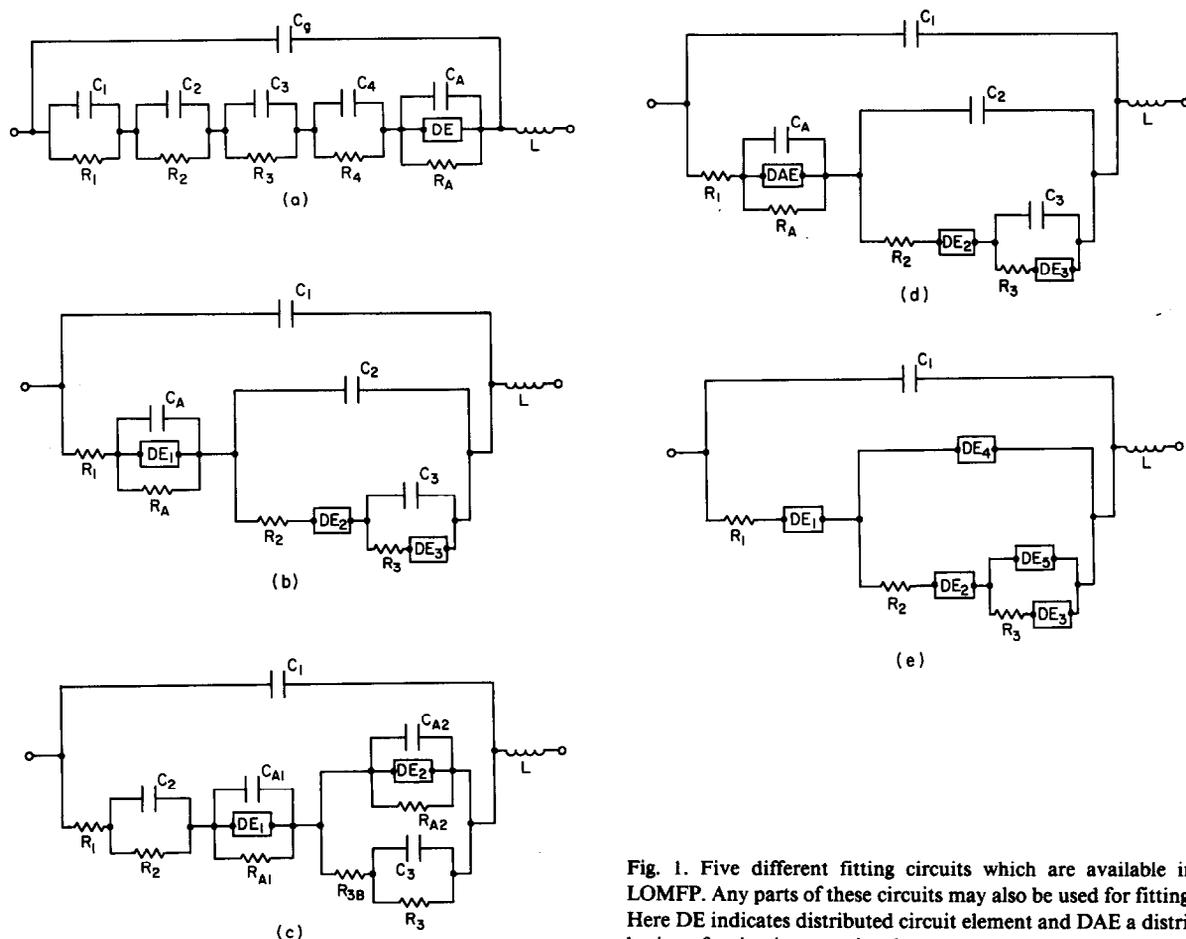


Fig. 1. Five different fitting circuits which are available in LOMFP. Any parts of these circuits may also be used for fitting. Here DE indicates distributed circuit element and DAE a distribution of activation energies element.

materials, do not involve a high enough frequency for the effects of C_g to be apparent in the data. Then even CNLS fitting yields no estimate of C_g , and it should be ignored. A summary of many equivalent circuits which have been proposed for solid and liquid electrolyte situations has been presented elsewhere [18]. Here we shall discuss only the capabilities of the fig. 1 circuits.

It may be wondered why LOMFP currently incorporates only five different circuit structures. There is, in fact, much more here than meets the eye. First, the DE and DAE circuit blocks or elements indicate that any of a variety of such elements may be incorporated in the circuit, providing great versatility. Second, the actual instantiation of the fig. 1 structures in LOMFP includes a very powerful device. In the input file which chooses a given circuit, any elements of the circuit (up to 30 in the fig. 1 circuit) which are ignored in the input (i.e. set fixed to zero) do not appear in the circuit. Only those elements which are to be incorporated into a given circuit need be specified (and designated as free to vary or fixed).

For example, in fig. 1a one might specify C_g , R_1 and a specific DE (up to four individual parameters) as free to vary. Then only these elements would appear in the model which is compared with the data by CNLS. This feature is provided by the use of logical variables in the program. It ensures that if a resistor, capacitor, inductance, or DCE (at the impedance level) is ignored, the impedance of the element will be set to zero if it appears in a series branch, while it will be set infinite if it appears in a parallel branch. Boukamp [13] has stated that an elimination approach of this type leads to the need to carry out many unproductive calculations. This is not at all the case with LOMFP. Unused elements do not appear at all in the calculations and thus do not increase its run time.

It will be noted that most of the circuits incorporate a hierarchical (or ladder network) type of structure. When only resistors and capacitors occur in an equivalent circuit, hierarchical, series, and parallel structure can all be made to show exactly the same frequency response at all frequencies [19], but this is not the case when DCE's appear in the circuits [17]. Even in the pure RC case, results of a detailed semi-microscopic theoretical calculation for con-

ducting situations, such as superionic materials, indicate that the hierarchical structure is more basic (in terms of the microscopic parameters, such as mobility and charge carrier concentration) than the other possibilities [14,16,20].

It should be emphasized that the simple input choice conventions used in LOMFP allow any or all of the elements shown in the fig. 1 circuits to be incorporated or ignored, with no significant speed penalties. Thus, the five circuit structures included, along with the DCE choices, allow a vast possible number of individual circuits to be built and used with simple input choices. Room has been provided in LOMFP, however, to add other structures if they should ever be needed. Further, additional circuits can be added as simple separate subroutines outside of LOMFP and linked with it during compiling, yielding infinite expansibility.

There is an electrochemical situation where not all parameters of a fitting model should be taken independent [20]. A LOMFP approach to fitting several interdependent parameters will be discussed in the next section. Finally, one "circuit" choice incorporated into LOMFP is not for direct IS CNLS fitting. Instead, it allows NLLS fitting of a wide variety of $y(x)$ functions, ones where the parameters may appear nonlinearly. Up to sixteen parameters are currently available. Since such IS parameters as relaxation times and resistances may often be thermally activated, it is worthwhile being able to use NLLS to obtain parameter estimates from such functions as [4,8,21]

$$\tau(T) = AT^n \exp[E_a/k_B(T-T_0)] \quad (2)$$

essentially the Vogel [22] (or VTF) equation. Here T is the absolute temperature and k_B is the Boltzmann constant. This function is directly available in LOMFP. Here A , n , E_a , and T_0 might all be free, but usually n will be known ab initio. The present T_0 is related to a glass transition temperature, and E_a is an activation energy (enthalpy). In simple cases where n is taken zero and $T_0 \equiv 0$ (ordinary thermal activation), it is customary to obtain estimates of A and E_a by using ordinary linear least squares fitting with $y(x) = \ln(\tau)$ and $x = T^{-1}$. But such a transformation will introduce bias of its own, and it is generally better to avoid such bias by fitting the original equation, one involving $\tau(T)$ data directly determined from

experiments and data fitting (such as CNLS fitting of IS data).

6. Distributed circuit element choices

In this section, we identify and briefly discuss the various distributed circuit elements which can be incorporated in the DE blocks of the various available equivalent circuits discussed earlier. Functions of these types are nearly always needed in fitting real IS data to an equivalent circuit. Any of the 11 available circuit elements can be incorporated into a given circuit by the selection of a single number in the input to LOMFP. Because most of the available DCE's have been discussed in some detail recently [3,5], we shall omit detailed discussion of them here.

Table 1 lists the available functions and formulas (where practical). Pertinent references are also included. In this table s is a normalized frequency: the product of a time constant or relaxation time τ ($=RC$) and ω . Thus the ωRC term in line 1 of the table may be replaced by s . The parallel RC (line 1) is not, of course, a distributed element, but it has been included in the table for comparison with the other elements, and because all DE blocks in the equivalent circuits can be replaced by the parallel "element" as well as by any of the other distributed elements listed. It has been shown that most DCE's of interest can be defined in normalized form either at the impedance level or at the complex capacitance/complex dielectric constant level [3,5]. Although they represent different responses and physical pro-

cesses, their normalized form is the same at these levels. We earlier used A to designate any one of the four immittance functions. To be more specific, consider A_k , where $k=M, Z, Y$, or ϵ ; then $A_Z \equiv Z$, $A_\epsilon \equiv \epsilon$, etc. Finally, define the dimensionless, normalized form of A_k as

$$I_k = (A_k - A_{k\infty}) / (A_{k0} - A_{k\infty}), \quad (3)$$

where A_{k0} and $A_{k\infty}$ are the low and high frequency limiting values of the relaxation process described by A_k . For example, at the impedance level,

$$I_Z(\omega) \equiv [Z_Z(\omega) - R_{Z\infty}] / (R_{Z0} - R_{Z\infty}), \quad (4)$$

and $R_{Z0} = Z'_Z(0)$, $R_{Z\infty} = Z'_Z(\infty)$. Of course the limiting frequencies need not be zero and infinity but merely far enough from the center relaxation frequency for $Z_Z(\omega)$ to reach its real limiting values. Here $Z_Z(\omega)$ is an expression for a particular theoretical DCE at the impedance level. With the above definitions, the normalized immittance function I_k satisfies $I_k(0) = 1$ and $I_k(\infty) = 0$.

For added generality, where possible the formulas for the DCE's in table 1 are given in I_k form with $i = Z$ or ϵ (or C^*). Because the CPE is not fully physically realizable, $Z_Z(0)$ does not exist for it, and it cannot be expressed in I_Z or I_ϵ form. The A_0 in the table is a frequency-independent constant. Those DCE's listed in table 1 without I_k expressions are generally too complicated to be given conveniently here but are discussed in detail in the references cited. It will be noted that in the circuits of fig. 1a-d several of the DCE's have a resistor R_A and capacitor C_A in parallel with them. Although the DCE can be used

Table 1
Summary of available distributed circuit elements.

No.	Physical process/name	Refs.	Acronym	I_k formula
1	parallel RC		PRC	$[1 + i\omega RC]^{-1}$
2	constant phase element	[5,23-26]	CPE	$A_k = [A_0(i\omega)^{w_k}]^{-1}$
3	ZC or ZARC	[5,24,26,27]	ZC	$[1 + (is)^{w_k}]^{-1}$
4	Havriliak-Negami	[28,29]	HN	$[1 + (is)^{w_k}]^{-\phi_k}$
5	generalized finite-length Warburg	[3,7]	GFW	$\tanh(is)^{w_k} / (is)^{w_k}$
6	Williams-Watts; stretched exponential	[29-31]	WW	-
7	generalized Jonscher element	[5,32]	GJ2	-
8	asymmetric exponential DAE	[5,33]	EDAE1	-
9	symmetric exponential DAE	[5,33]	EDAE2	-
10	symmetric gaussian DAE	[5,33]	GDAE2	-
11	general uniform diffusion	[34,35]	GUD	-

without these elements, omitting just the DCE automatically eliminates them as well.

Several of the DCE's are available with different parameterizations. The ZC is one of these. At the impedance level it has been shown [25,26] that the ZC may be either considered as a DE in its own right (involving the parameters $(R_0 - R_\infty)$, τ , and ψ) as in line 3 of table 1 or as a CPE in parallel with a resistor R_D (involving the parameters R_D (or $R_0 - R_\infty$), A_0 , and ψ). The goodness of fit of data using either parameterization is exactly the same for a CNLS solution, but the estimated relative standard deviations, S_r , of τ or A_0 will generally differ, as will the intercorrelations between the parameters. Which parameterization to use should thus depend on which choice yields the smallest intercorrelations and S_r 's. Since the ZC may be considered to involve a CPE, which is physically unrealizable over the full frequency range, the ZC is also unrealizable and does not reduce to limiting single-time-constant behavior at both frequency extremes [3,26]. Perhaps for this reason, there seems to be no plausible physically based theories which lead to full ZC response. It is worth mentioning, however, that theories of fractal interfaces [36,37] and of bulk hopping relaxation [38] can lead to response very similar to that of the ZC. Further, only at the extremes of frequency, often outside the usual range of measurement, does one expect the non-physical characteristics of the ZC to be of importance. Because of its simplicity, it thus is often found to be a useful fitting element.

There is another powerful general choice built into LOMFP which is of particular interest for the analysis of dielectric measurements. When the input number which specifies which of the DCE's is to be incorporated at a particular position in a circuit is taken negative, the DCE is defined at the complex dielectric constant level, rather than at the impedance level as is otherwise the case.

This transformation possibility will be made clearer by an example. Consider the fig. 1a circuit with only C_A , R_A , and DE active and take DE as a ZC. Then in the normal impedance level situation, $Z_{ZC} = R_D / [1 + (i\omega\tau)^\psi]^{-1}$, where $R_D \equiv R_0 - R_\infty$. The total impedance of the circuit, involving five parameters, is then

$$Z_T = Z_{ZC} / [1 + \{G_A + i\omega C_A\} Z_{ZC}], \quad (5)$$

where $G_A \equiv 1/R_A$. When the dielectric transformation is invoked, we obtain the new total impedance $Z_{TNEW} = [i\omega(C_C/R_U)Z_{TOLD}]^{-1}$, where R_U is a unit resistance included to produce the correct units. In the present case, the result is

$$Z_{TNEW} = (C_A/C_C)R_U + [i\omega\{C_C R_A/R_U\}]^{-1} + [1 + (i\omega\tau)^\psi] / [i\omega C_C R_D/R_U], \quad (6)$$

or

$$Z_{TNEW} = R_S + [i\omega C_S]^{-1} + [i\omega C_{ZC}]^{-1}, \quad (7)$$

where the series resistance $R_S \equiv (C_A/C_C)R_U$ and the series capacitance $C_S \equiv (R_A/R_U)C_C$. The complex capacitance of the ZC (defined at the complex capacitance or dielectric constant level) is

$$C_{ZC} = \frac{(\epsilon_0 - \epsilon_\infty)C_C}{1 + (i\omega\tau)^\psi} \equiv \frac{(C_0 - C_\infty)}{1 + (i\omega\tau)^\psi}, \quad (8)$$

with $C_0 \equiv (R_0/R_U)C_C$ and $C_\infty \equiv (R_\infty/R_U)C_C$, or $C_D \equiv C_0 - C_\infty \equiv (R_D/R_U)C_C$. Thus a parallel circuit has been transformed into a series one involving the proper ZC DCE at the dielectric level. It can be considered as either a unified DCE (the Cole-Cole element [24]) or as a CPE in series with a resistor. From the fitting estimates of R_A , C_A , R_D , and the known value of C_C one can immediately obtain the corresponding estimates of the series parameters R_S , C_S , and C_D using the above relations. Matters are even simpler when a single DCE is to be fitted to data. Then with Z data a Z form of the DCE is used, while with complex ϵ data an ϵ form is used directly with no transformation but with the proper interpretation of the parameters involved. Some of the DCE's in table 1 are shown without I_k expressions because they are too long to fit in the table. Further, all the DAE's listed in the table 1 require integration to obtain I_k ; they are calculated in LOMFP using accurate built-in integration procedures. There is ambiguity present for some of the DCE's in the table. Thus, for example, the HN element reduces to the ZC when ϕ_k is fixed at unity and to the Davidson-Cole form when ψ_k is fixed at unity. We include both DCE possibilities, both for simplicity and because of the different parameterization choices provided for the ZC. Another overlap occurs between the GUD element and the GFW. The I_k for the GUD represents the

normalized immittance of a general, finite-length, uniform transmission line with its termination impedance determined by a disposable input parameter. Thus the GUD here represents diffusion in a homogeneous, finite-length region with an arbitrary reaction (or disappearance) rate of the diffusing entity (charged or uncharged) at the boundary of the region. When the line is shorted (infinite reaction rate), the GUD reduces to just ordinary finite-length Warburg response [3], that of the GFW with ψ_k fixed at 0.5. In turn, this reduces to ordinary infinite-length Warburg response for $s > 3$. The GUD represents diffusion with no reaction when the termination parameter is taken zero: an open-circuited transmission line. The DCE's available in the Boukamp EQIVCT program are the CPE and the open short circuited limits of the GUD.

Often some of the fitting parameters in an equivalent circuit are known to be correlated. Consider the fig. 1b circuit with only the elements R_2 , C_2 , R_3 , C_3 , and DE_3 present and take DE_3 as the GUD. These elements and structure can then represent a typical reaction-adsorption-diffusion process, with R_2 and C_2 being the reaction resistance and capacitance (double layer capacitance), R_3 and C_3 the adsorption resistance and capacitance, and GUD representing the diffusion process. It turns out that all these elements but C_2 are interrelated [34]; for example, they depend directly or inversely on R_2 , the reaction resistance. Although one can fit the data taking all these parameters and elements independent, it is important to be able to take their interdependencies into account directly. To do so when their interrelations are known, we reparameterize the problem with new parameters which do not depend directly on each other. This may reduce the intercorrelations between the elements of the parameter set, a desirable result. More important theoretically, it eliminates a paradox which can otherwise occur. If the parameters are taken independent, then in principle C_3 might become very large, essentially wiping out any diffusion effects except at the lowest frequencies. Now C_3 represents a property of the system associated with the electrode-material interface and so is intensive. But at low frequencies GUD becomes extensive when the diffusion length becomes comparable with the electrode separation. It is not proper for a parallel intensive variable to be able to wipe out

the almost complete effect of an extensive one since they are associated with quite different regions in the material. Although this can happen with the original parameter set with all parameters free to vary independently during the CNLS fit, it cannot happen for the reparameterized set, where, for example, if $R_2=0$, then $R_3=0$, $Z_{GUD}=0$, and $C_3=\infty$. The LOMFP allows a choice of either type of parameterization for situations of the present type.

7. Weighting and weighting comparisons

7.1. Weighting possibilities and discussion

Ideally, every data value in a LS fit should be weighted using σ_i , the standard deviation associated with that value. Although an estimate of σ_i , S_i , may be obtained by replicating the measurements many times, this is often impractical and other weighting choices must then be used. Nevertheless, LOMFP allows individual values of S_i (when available) to be entered along with every (real, imaginary, or both) data point. The actual individual weights used are then given by $W_i \equiv S_i^{-2}$, the inverse (estimated) variance of the i th value.

The simplest weighting assumption to make when S_i values calculated from replicated measurements are unavailable is unity weighting (UWT), that where all S_i 's are taken unity (or any other constant non-zero value). This is often an appropriate choice when the immittance data vary only over a total range of 3 or so. But immittance data magnitudes often vary over three or more decades, and then UWT is inappropriate since it will lead to parameter estimates determined primarily by only the largest data values.

When the data range magnitude is large, it is often appropriate to assume that the random errors in the measurements are proportional to the magnitudes of the data values themselves (constant percentage error magnitudes over the entire range). This natural assumption leads to what has been termed proportional weighting (PWT). For a given data point with $A_i = A_i' + iA_i''$, then $S_i^R = |A_i'|$ and $S_i^I = |A_i''|$. Similarly, for data in polar form, $A_i = A_{im} \angle \theta_i$, $S_i^m = A_{im}$ and $S_i^\theta = |\theta_i|$.

Another related weighting, modulus weighting (MWT), has been discussed and advocated by Zol-

towski [39] and is that used in EQIVCT. In this weighting, which is only meaningful for data in rectangular form, $S_i^R = S_i^I = [(A_i^R)^2 + (A_i^I)^2]^{1/2}$, the modulus of A_i . Although Zoltowski [39] has stated that there is no rational basis for carrying out the fitting of data at one immittance level or form as compared to another, we disagree. We believe that if data are directly obtained in polar form, they should be fit in that form. If the measurements yield the data directly at the impedance level, they should be fit at that level rather than at the admittance or any other level. The reason for fitting with the data in its original form (a capability of LOMPF whatever the immittance level and form) is that to do otherwise will generally introduce bias into the data. Almost any usual transformation of the data will do so. If this point of view is accepted, then data obtained in polar form should not be fitted with MWT since to do so requires a transformation to rectangular form.

Zoltowski has suggested that MWT is preferable to PWT for data in which the errors in A_i^R and A_i^I residuals are not independent. Of course if all correlations are known or can be well estimated, one should carry out the fitting using an error matrix with non-zero non-diagonal elements [40], but this is usually impractical. Some such correlation actually always occurs for several reasons. First each exact A_i^R and A_i^I value is related to the other through the holistic Kronig-Kramers integral transforms when the system is minimum phase, the usual situation. But it is highly unlikely that these relations induce correlations of any consequence between the errors in the A_i^R and A_i^I values when errors are actually present and are non-systematic. Although it is possible that the measuring apparatus itself can lead to error correlation big enough to be important, one hopes and expects that this will not be the case when the errors are essentially random and non-systematic. A final source of correlation in the actual CNLS residuals is that there are $2N$ of them (for complex data fitting) but only $N_{DF} \equiv (2N - P)$ degrees of freedom of the system, where here P is the number of free parameters in the fit. When $2N \gg P$, the usual situation, this source of correlation is negligible, however.

Another putative reason for preferring MWT to PWT is that MWT leads to "fit results... independent of the representation" [13]. For example, it yields values of S_i , the estimated standard deviation of the

weighted residuals (a measure of the overall goodness of fit) which are nearly independent of the immittance level used in the fit. PWT, on the other hand, usually leads to significant differences in S_i for fitting results at different levels. In addition, MWT yields parameter estimates which are also nearly level independent, unlike PWT. These near equalities for MWT certainly initially suggest that it doesn't matter which level is employed for the fitting. But we believe that for good data one should indeed see differences between fitting results obtained at different levels, such differences arising from the biases introduced by the transformations employed. If this is true, then MWT actually obscures and covers up important differences and should be eschewed unless one is sure that there are appreciable equipment-induced correlations between the real and imaginary errors. Even then, the first order of business should be to try to alter the equipment so as to reduce or eliminate such correlations (which will usually be associated with the introduction of systematic errors by the equipment) before final measurement is carried out. In the next section the results of some comparisons between MWT and PWT fitting are presented for synthetic data and show unequivocally the PWT is much superior to MWT for such data, data without large error correlations.

Not only does the LOMFP include UWT, PWT, and MWT, it provides several other weighting possibilities. One of the most flexible is VWT, for which $S_i^a = |A_i^a|^n$, where $a=R$ (or I) of I (or $''$), and $0 \leq n \leq \infty$. When $n=0$, one has UWT, and $n=1$ yields PWT. There are likely to be situations where an intermediate weighting, say $n=0.5$, will be superior to either of these choices, and the value of n is a simple input choice.

Another more general weighting choice, CWT, is defined by $S_i^a = 1 + B_w |A_i^a|^n$, where B_w is a parameter whose value is set in the input. When it is zero, one again obtains UWT; when it is sufficiently large, one approaches VWT. This form, with $n=1$, has been used [41,42] in connection with a negative entropy objective function, SSE, to be discussed in the next section. A final weighting choice available in LOMFP is residual iteration weighting, RWT. Define the i th direct residual as $r_i^a \equiv A_{e_i}^a - A_{t_i}^a$. Then RWT begins with a fit using VWT, saves the resulting r_i^a 's, and carries out successive runs using $S_i^a = r_i^a$, where r_i^a is

the i th residual obtained from the previous run. Such residual weighting tends to decrease the effects on the fit of residuals of larger than average magnitude, and vice versa. It often converges well in one or two iterations, after which successive r_i^a values do not change appreciably. The weighted sum of squares, S_w , then is closely equal to unity. Even when iteration leads to instability, it may still be stopped at a point where S_w is very nearly unity, a condition where the weighting uncertainties and the residual vectors are very nearly the same. Although iterative residual reweighting schemes somewhat like RWT have sometimes been used in the past for linear least squares situations, there seems to be no adequate statistical basis for it, especially for NLLS fitting. Nevertheless, we find that it can sometimes (often?) lead to a better fit (smaller biases in the parameter estimates – see example in section 7.3.) than say PWT. Its trial use is therefore encouraged.

7.2. Statistics calculations

Here we describe the statistical calculations and outputs available in LOMPF. Define the weighted residuals (including unity weighting) as $R_{wi}^a \equiv r_i^a/S_i^a$ and the relative residuals as $R_{ri}^a \equiv r_i^a/A_{ei}^a$. They are the same only in the PWT case provided signs are properly accounted for. Further, define N_x as N for real or imaginary fitting ($a=R$ or I), and as $2N$ for full complex data fitting. Then $N_{DF} \equiv N_x - P$. The LEV part of LOMPF provides the following outputs and statistics: estimates of all the free parameters; r_i and r_i'' ; R_{ri} and R_{ri}'' ; $AV(R_{wi})$; $AV(R_{ri})$; $SD(R_{wi})$; $SD(R_{ri})$; and SSE. Here $AV(R_{ri})$, for example, is given for complex fitting by

$$AV(R_{ri}) = N_x^{-1} \sum_{i=1}^N (R_{ri}' + R_{ri}''). \quad (9)$$

For real-only fitting, $R_{ri}'' \equiv 0$. Similarly, for example, the estimated standard deviation of R_{wi} is

$$SD(R_{wi}) = \left[N_{DF}^{-1} \sum_{i=1}^N \{ (R_{wi}')^2 + (R_{wi}'')^2 \} \right]^{1/2} \\ \equiv [S_w/N_{DF}]^{1/2} \equiv S_{fw} \equiv S_f. \quad (10)$$

Since the R_{wi} 's are the residuals actually minimized

in the fit, we expect that $AV(R_{wi})$ will be very small for a good fit. For such a fit, we certainly expect $|AV(R_{wi})|$ to be less than $3S_f$ since the expectation value of $AV(R_{wi})$ is zero. Let $SD(R_{ri}) \equiv S_{fr}$. Then again we expect $|AV(R_{ri})|$ to be no larger than $3S_{fr}$. Usually the averages are much smaller than the above values for a good fit. Only in the PWT case will the two estimated standard deviations defined above be identical.

As part of every LEV CNLS fit we actually calculate three S_{fr} values, that for the real part of the data separately, that for the imaginary part, and that for both together. When one of the first two S_{fr} values turns out to be appreciably smaller than the third (and, conversely, one bigger), it is an indication that a separate full fit for the part of the data with the smallest S_{fr} should be carried out for comparison with the full CNLS fit since some parameter estimates may then be better for the former fit than for the latter.

The negative entropy objective function SSE, which is not minimized directly in LOMFP, is [41,42]

$$SSE = 1 + [\ln(N_x)]^{-1} \sum_{i=1}^N q_i \ln(q_i), \quad (11)$$

where for CNLS fitting

$$q_i \equiv [(R_{wi}')^2 + (R_{wi}'')^2] / \sum_{i=1}^N [(R_{wi}')^2 + (R_{wi}'')^2]. \quad (12)$$

The quantity SSE, another measure of the goodness of fit in addition to the two estimated standard deviations defined above, falls in the range of zero to unity. It is unity if all but one of the q_i 's are zero and is zero if they are all equal to N_x^{-1} . It is thus a measure of the constancy of the magnitudes of the weighted residuals. It has been suggested [41,42] for ordinary nonlinear least squares fitting that the parameter B_w of CWT be selected to make SSE a minimum. We have not evaluated this approach sufficiently for CNLS fitting to be able to make a recommendation for this case.

It should be noted that SSE and S_{fr} are normalized quantities and do not vary greatly with the type of weighting and fitting immittance level, as does S_{fw} since it depends directly on the scale of the weighting. Thus, in determining the weighting which gives the best fit of the data, the first two measures should be compared for different weightings and immitt-

ance levels in preference to the last.

The OLSON subprogram of LOMFP provides the same residual listings as does LEV and also includes a S_{fw} output. In addition, OLSON yields the parameter variance-covariance matrix, the correlation matrix for the parameters, their estimates, and their estimated standard deviations and estimated relative standard deviations. Only if the parameter estimates and S_{fw} values are very nearly identical for the LEV and Olson outputs should the LEV statistics be judged appropriate and should it be assumed that a true LSS has been obtained; and even then it is a good idea to fit again with quite different starting parameter guesses.

7.3. Fitting results for different weightings

In order to illustrate the use of LOMFP with different weightings, we have generated "exact" reproducible immittance data from a relatively simple equivalent circuit and rounded them to various numbers, n , of decimal places. We used a circuit consisting of C_g , R_∞ , and a ZC DCE at the impedance level. In circuit (a) of fig. 1 the five elements present were C_g , $R_\infty=R_1$, and DE set to the ZC. The ZC parameters were R_D (equal to R_0-R_∞), τ , and ψ . Their exact values, used in generating "exact" 13-place data, are given in table 2. Twenty-seven angular frequencies, distributed equally on a log scale, were used with $10^{-4} \leq \omega \leq 10^9$. The resulting ratio of the maximum to minimum value of Z'' in this frequency span was about 250.

A three-dimensional perspective plot [2] at the impedance level of the data and its three projections is presented in fig. 2. The solid lines represent the exact data and the dashed ones (which are hardly visible over most of the range) show the result of a

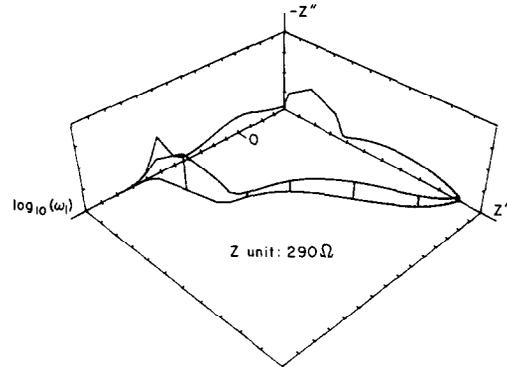


Fig. 2. Three-dimensional perspective plot of exact synthetic impedance data (solid lines) and CNLS PWT fit of the data rounded to two decimal places (dashed lines). The $\log_{10}(\omega)$ scale unit used here is unity. See text for identification of circuit and element values employed in generating the data.

PWT LOMFP fitting of rounded 2-place data to the equivalent circuit used to generate the data. The plotting routine connects points with straight lines and there were too few points in the high frequency ($\omega \sim (R_\infty C_g)^{-1}$) region to yield the true rounded appearance, that of a semicircle with its center on the real axis. Some overlap is apparent between the bulk R_∞ , C_g effects and the wide ZC displaced semicircle.

Also shown in table 2 are estimates of the relative standard deviations, S_{rj} , for the $j=1$ to 5 parameters for fits of the exact impedance data with UWT and with PWT. Here the designation ZR-ZR indicates that the original data was in rectangular impedance form and was fitted in the same form. Although the table shows that UWT can yield excellent estimates of the parameter values using the exact data, it is evident that the precision of the parameter estimates is

Table 2

Fitting parameters and their estimated relative standard deviation, S_{rj} , with U and P weighting; 13-place data and ZR-ZR fits.

Parameter no.	Identity	Exact value	S_r : UWT	S_r : PWT
1	R_D (M Ω)	2.0	2.5×10^{-7}	3.4×10^{-14}
2	τ (s)	1.0	1.6×10^{-6}	1.9×10^{-13}
3	ψ	0.30	4.4×10^{-7}	3.5×10^{-14}
4	R_∞ (M Ω)	1.0	2.3×10^{-7}	4.7×10^{-14}
5	C_g (pF)	1.0	5.1×10^{-7}	3.9×10^{-14}

six or seven orders of magnitude worse here for UWT as compared to PWT.

Table 3 shows some results of PWT fitting of the data rounded to various levels. Since the exact parameter values used to generate the exact data are known, one is able to calculate the relative error, E_{vj} , in the parameter estimates, Q_{vj} . If Q_{ej} denotes the exact parameter values, then $E_{vj} \equiv (Q_{ej} - Q_{vj})/Q_{ej}$. In the table we compare E_{vj} values with the estimated parameter relative standard deviations, S_{vj} , for several rounding levels. Although the results show expected statistical variability, they indicate that the S_{vj} 's generally provide reasonable and useful estimates of the magnitudes of the relative error actually present in the parameter estimates.

The last line in table 3 is for $n=2$ fitting with iterative RWT, and the results should thus be compared with those in the line above. Iterative weighting was carried out in LEV, and the resulting residual weights used in OLSON to obtain the values shown in the table. As expected, S_f has been driven very close to unity; further the relative error in four of the five parameter estimates is smaller than that with PWT, suggesting that RWT may indeed be useful. It should be noted, however, that rounding does not yield a close approximation to a Gaussian distribution of the actual random errors in the data, although it certainly does not introduce systematic errors. Perhaps in part because of this, converged RWT fitting leads here to much smaller (and unrealistic) S_{vj} estimates than does PWT.

Although our main comparisons will use PWT and MWT and will all be at the $n=2$ level in order to approximate typical experimental error levels, it is worthwhile to present a few $n=2$ UWT comparisons as well, as in table 4. Here in such data-fit designations as ZR-MR the M stands for the complex mod-

ulus immittance level, Y for the admittance level, and E for the complex dielectric constant level. The original exact data at the Z level was first rounded to two places and then converted to other levels. Thus the Z -level data should be representative of reasonably good experimental data taken at that level, and the other fits show what can be obtained from transforming the data and using only UWT. Here O and L designate the OLSON and LEV subprograms.

First, on comparing the ZR-ZR line of table 4 with the $n=2$ PWT results of table 3, one sees that PWT yields much decreased relative error of the parameter estimates and much smaller S_{vj} values. Second, the ZR-MR results for LEV and OLSON fitting seem comparable based on their S_f values but yield very different parameter estimates! The OLSON run has changed the LEV parameter estimates used in its input very much for the worse while still maintaining nearly the same S_f value. Here we know that the LEV results are actually much better, but if no additional information were available one would presumably choose the OLSON ones since they show a slightly smaller S_f . This example indicates what can happen when inappropriate weighting is used and a true LSS is not reached.

No convergent OLSON results were obtained for the ZR-YR and ZR-ER cases. The LEV output shows some large biases and, in fact, even the LEV results are not particularly trustworthy here. In any event, the major deleterious effects of transformation of the data immittance level before fitting are well illustrated here for a poor weighting-type choice.

Now it can be readily shown for both PWT and MWT (but not UWT) that fitting results for either of these weightings will be identical for the Z and the M immittance levels. The transformation from Z to M or vice versa makes no difference. Similarly, fit-

Table 3
Relative error and relative standard deviations for ZR-ZR PWT data, rounded to n places.

n	$10^n S_f$	$10^n E_{vj} 10^n S_{vj}$				
		1	2	3	4	5
4	1.55	0.242 0.353	-2.60 1.96	-0.195 0.418	-0.486 0.458	0.063 0.335
3	1.20	-0.225 0.272	2.74 1.51	0.516 0.322	0.218 0.353	-0.190 0.258
2	1.29	-0.420 0.294	-1.51 1.65	-0.576 0.351	-0.178 0.383	-0.004 0.280
2i	99.92	-0.414 0.006	-1.05 0.228	-0.666 0.037	-0.157 0.006	$-4.61 \times 10^{-4} 0.001$

Table 4

Comparison of fit results with U weight using data rounded to two places for all four immittance levels.

Input-fit PGM	S_f	$10^2 E_{ij} 10^2 S_{ij}$				
		1	2	3	4	5
ZR-ZR O	1.58×10^4	3.49 1.24	6.44 7.44	-5.74 2.12	-2.72 1.18	-0.335 2.40
ZR-MR L	2.286×10^9	2.76 -	-75.7 -	10.6 -	-0.358 -	-0.498 -
ZR-MR O	2.280×10^9	722 439	527 476	-27.7 233	-1.13 3.98	-0.505 0.103
ZR-YR L	5.17×10^{-7}	-39.3 -	-48.1 -	-16.2 -	-1.34 -	-0.120 -
ZR-ER L	2.26×10^{-6}	1.06 -	-98.1 -	-84.4 -	-10.5 -	5.2×10^7 -

ting results for the Y and ϵ levels will also be identical. Thus since all the rest of our results will be for either PWT or MWT it is sufficient to carry out fits only for the Z and Y levels.

Tables 5 and 6 present our main PWT and MWT results. Those weight letters with an asterisk superscript indicate fits made without immittance level transformation of the data, and thus, according to our earlier hypothesis, they should yield best fits. For example, in table 6 the ZP-ZP|P* line involved original exact 13-place ZR data which were first converted to polar form, then both components of all the ZP data points were rounded to two places, and the results were finally employed in ZP, PWT fitting. The data for other lines in the ZP group of this table were calculated, however, from the ZP-ZP 2-place data by transformation. Thus, the lines involving asterisks maintain to the maximum degree possible with 2-place rounded data the maximum information from the original exact data. We have tried to

present as many fitting combinations as possible in tables 5 and 6, but it will be noted that ZP-ZR|P, ZP-YR|P, YP-ZR|P, and YP-YR|P lines are missing from table 6. Not even a good LEV fit could be obtained for these particular cases, presumably because the transformations carried out to obtain them destroy and/or scramble too much needed information from the original 2-place data.

There is a lot of information in tables 5 and 6. Here we shall only summarize some main conclusions with the hope and expectation that most of them will apply to the fitting of real data which have negligible systematic errors. First, on the average, based on both E_{ij} and S_{ij} estimates, the asterisked lines yield better results than the others, showing, as expected, the danger of transforming data with appreciable errors before fitting, even when the errors are random. In increasing order of goodness of fit judged also by S_f , one has ZR-ZR|P, ZP-ZP|P, YR-YR|P, and YP-YP|P. One must conclude that YP-YP|P is bet-

Table 5

Comparison of fitting results for various levels, weighting, and forms of the data for 2-place data rounded from exact ZR level data.

Input-fit WT	$10^2 S_f$	$10^2 E_{ij} 10^2 S_{ij}$				
		1	2	3	4	5
ZR-ZR P*	1.29	-0.420 0.294	-1.51 1.65	-0.576 0.351	-0.178 0.383	-0.004 0.280
ZR-YR P	2.76	-0.195 1.35	-5.29 4.33	-0.903 0.736	-0.026 0.494	-0.021 0.945
ZR-ZP P	1.83	-0.494 0.603	-3.73 2.42	-0.777 0.502	-0.197 0.501	-0.129 0.654
ZR-YP P	1.82	-0.432 0.598	-3.71 2.41	-0.780 0.499	-0.095 0.501	-0.156 0.646
ZR-ZR M	1.10	2.30 1.64	2.10 11.9	-3.73 2.49	-2.04 0.916	-0.470 0.435
ZR-YR M	1.10	2.26 1.63	2.33 11.8	-3.76 2.48	-1.88 0.908	-0.481 0.431

Table 6

Comparison of fitting results for various levels, weighting, and forms of the data using 2-place data rounded from exact ZP, YR, and YP level data.

Input-fit WT	$10^2 S_f$	$10^2 E_{rj} 10^2 S_{rj}$				
		1	2	3	4	5
ZP-ZP P*	1.24	0.175 0.409	-1.05 1.62	0.102 0.337	-0.096 0.339	-0.324 0.441
ZP-YP P	1.24	0.232 0.409	-1.02 1.62	0.097 0.337	+0.010 0.338	-0.353 0.441
ZP-ZR M	1.13	2.78 1.71	9.57 12.4	-4.44 2.58	-1.54 0.940	-0.291 0.444
ZP-YR M	1.13	2.48 1.71	9.71 12.4	-4.36 2.58	-1.32 0.944	-0.304 0.444
YR-ZR P	1.11	-0.371 0.254	-1.56 1.40	0.501 0.299	-0.308 0.329	0.428 0.241
YR-YR P*	0.826	-0.456 0.405	-1.12 1.28	0.058 0.219	-0.012 0.146	-0.253 0.283
YR-ZP P	0.868	-0.494 0.286	-1.21 1.13	0.262 0.236	-0.196 0.237	-0.141 0.306
YR-YP P	0.867	-0.483 0.286	-1.21 1.13	0.262 0.235	-0.187 0.236	-0.161 0.309
YR-ZR M	0.410	-1.84 0.566	-10.2 4.02	2.43 0.901	0.484 0.307	0.368 0.161
YR-YR M	0.410	-1.82 0.568	-10.2 4.02	2.42 0.900	0.486 0.307	0.355 0.161
YP-ZP P	0.681	-0.205 0.225	-0.818 0.889	0.153 0.185	-0.219 0.186	0.016 0.242
YP-YP P*	0.683	-0.193 0.225	-0.812 0.891	0.154 0.185	-0.204 0.186	-3.4×10^{-4} 0.243
YP-ZR M	0.501	-1.48 0.704	-2.63 5.00	1.72 1.11	0.676 0.377	0.125 0.196
YP-YR M	0.502	-1.50 0.705	-2.69 5.02	1.78 1.11	0.706 0.376	0.108 0.197

ter than, for example, ZR-ZR|P because the former retains more of the original information in the exact data (at either the ZR or YR level) than the latter does when rounding to $n=2$ is carried out.

The tables also show that the transformed runs (those without asterisks) yield worse results than the untransformed ones which were derived directly from the exact data by rounding. This conclusion is not indeed obvious for the MWT results if one looks only at the S_f values. But it is clear that even though the ZR|M and YR|M S_f values are essentially the same and are lower than those from corresponding PWT runs, the MWT parameter estimates, and thus the relative bias values in the tables, are generally very much worse. Therefore, the S_f values are strongly misleading for MWT. This is one reason why the LEV program also calculates the S_{fr} estimates of the relative residuals. They show a different story indeed. Although $S_{fr}=S_f$ for PWT, the quantities are quite different for MWT. For example, the S_{fr} values for the ZR-ZR|M and ZR-YR|M runs are 2.87×10^{-2} and 3.66×10^{-2} , respectively, results which should be compared to the far better ZR-ZR|P value of 1.29×10^{-2} . Similarly, for example, the YR-ZR|M and YR-YR|M S_{fr} estimates are 2.07×10^{-2} and 1.95×10^{-2} , respectively, which should be compared to the YR-YR|P value of 8.26×10^{-3} . These results

suggest that MWT runs generally yield S_{fr} results between two and three times worse than those using PWT. Finally, the results of tables 5 and 6 suggest that when one takes statistical variability into account, S_{rj} values are generally good estimates of the corresponding relative error magnitude values to be expected. For PWT the actual $|E_{rj}|$ values nearly always fall within one or two times the S_{rj} standard deviation estimates.

Some of the foregoing conclusions are made even clearer by study of the relative residuals, R_{ri}^a , resulting from the various runs. Figs. 3, 4 and 5 present R'_{ri} and R''_{ri} results for $n=2$ ZR input data in two different forms. The (a) graphs are complex plane plots, and the (b) ones show the dependences of the relative residuals on $\log_{10}(\omega_i)$. Of course for PWT, weighted and relative residuals are the same. Scaling is such that the residual values shown are in per cent. One should examine these plots in conjunction with the correlation matrix of table 7. Correlations are presented for the three relative residual sets of figs. 3-5, for the log angular frequency vector, $[LW] \equiv [\log_{10}(\omega_i)]$, and for the exact $n=2$ input relative residuals, designated ZR|E, calculated by subtracting the $n=2$ rounded ZR data from the "exact" $n=13$ ZR data and normalizing with the exact data values themselves.

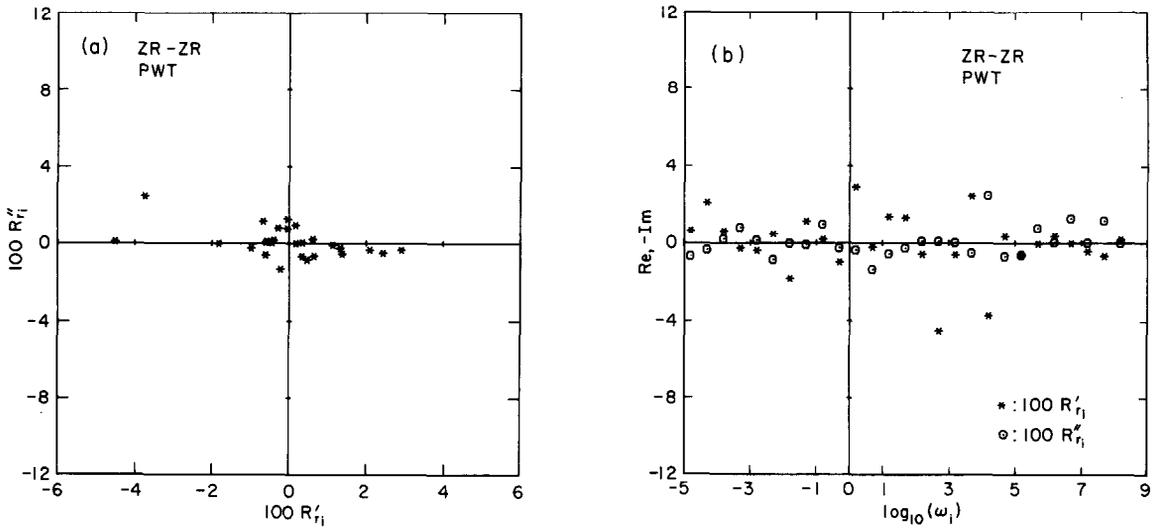


Fig. 3. Plots of relative residuals, R'_{ri} and R''_{ri} , from a ZR-ZR|P CNLS fit: (a) complex plane; (b) R'_{ri} and R''_{ri} versus $\log_{10}(\omega_i)$.

First, table 7 shows that there are no very large correlations with [LW], although the ZR|M results show the largest such correlations. We might expect that a good ZR-ZR|P fit should yield relative residuals in good agreement with the ZR|E ones since no transformations have been carried out. In fact, one finds correlation values of 0.991 for the real sets and 0.861 for the imaginary ones, quite large values. If one fits the ZR|P results versus the ZR|E ones, one finds a slope parameter of (1.009 ± 0.027) for the real sets, consistent with unity slope, but (0.67 ± 0.09) for the imaginary sets. Thus the ZR|P fit, while good,

is unable to reproduce the imaginary input relative residuals very closely.

Fig. 4 shows that the ZR-YR|P relative residuals, derived from a CNLS fit carried out after transformation of the immittance level, leads to very appreciably larger residuals than does the ZR-ZR|P fit. Further, fig. 4a shows a strong correlation between R'_{ri} and R''_{ri} for the transformed fit. As table 7 indicates, the actual correlation is 0.885, greatly increased from the -0.414 of ZR|E and the -0.453 of ZR|P. These results are striking evidence of increased correlation associated with a Z-to-Y transformation.

Table 7

Pearson correlations between [LW] and various $n=2$ relative residual sets, all derived from ZR input. The set marked ZR|E comprises the exact input relative residuals. Here ZY|P is an abbreviation for ZR-YR|P, etc.

	[LW]	ZR E		ZR P		ZY P		ZR M	
		[R'_i]	[R''_i]	[R'_i]	[R''_i]	[R'_i]	[R''_i]	[R'_i]	[R''_i]
[LW]	1								
ZR E	[R'_i]	-0.097	1						
	[R''_i]	0.501	-0.414	1					
ZR P	[R'_i]	-0.211	0.991	-0.464	1				
	[R''_i]	0.276	-0.438	0.861	-0.453	1			
ZY P	[R'_i]	-0.064	-0.840	0.129	-0.826	0.116	1		
	[R''_i]	0.007	-0.965	0.381	-0.946	0.485	0.885	1	
ZR M	[R'_i]	-0.476	0.749	-0.499	0.786	-0.501	-0.439	-0.680	1
	[R''_i]	0.566	-0.143	0.181	-0.196	0.402	0.030	0.197	-0.284

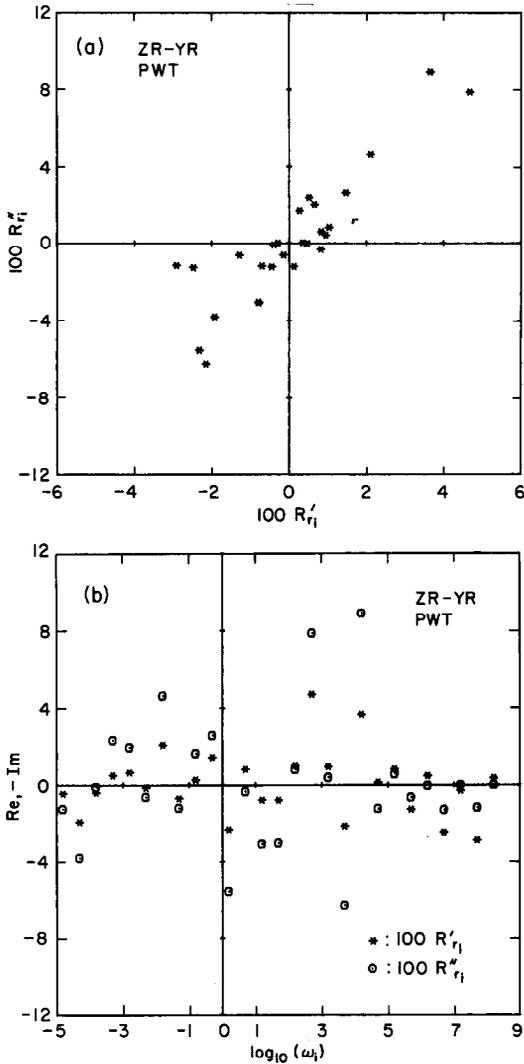


Fig. 4. Plots of relative residuals, R'_{ri} and R''_{ri} , from a ZR-YR|P CNLS fit: (a) complex plane; (b) R'_{ri} and R''_{ri} versus $\log_{10}(\omega_i)$.

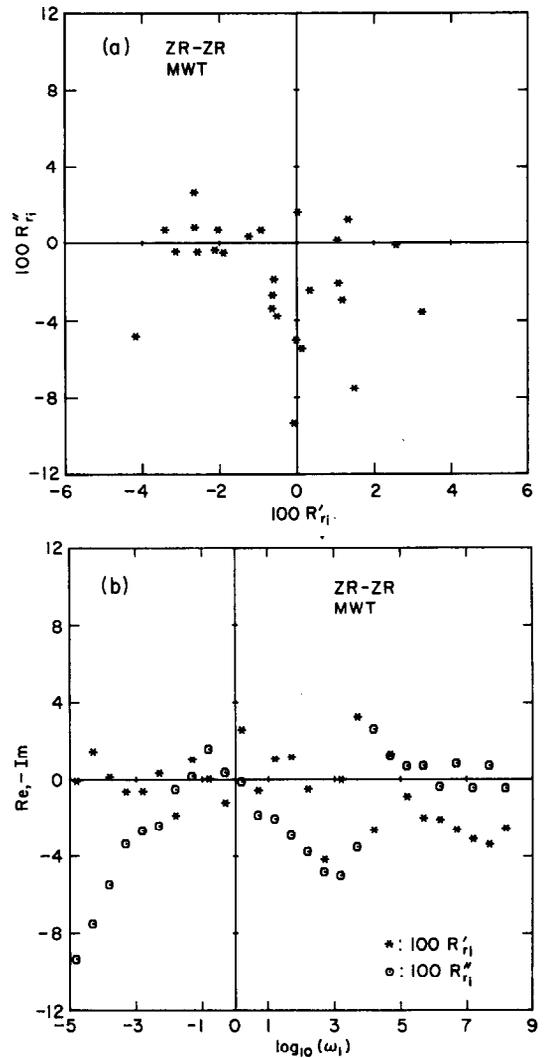


Fig. 5. Plots of relative residuals, R'_{ri} and R''_{ri} , from a ZR-ZR|M CNLS fit: (a) complex plane; (b) R'_{ri} and R''_{ri} versus $\log_{10}(\omega_i)$.

Fig. 5 shows the normalized residual results for the ZR|M fitting. Note that this fitting, as compared for example to that of ZR|P, has much larger relative residuals and shows in fig. 5b a much more systematic behavior of R''_{ri} with $\log_{10}(\omega_i)$ than do any of the other fits. Further it has reduced the correlations with the input ZR|E relative residuals from the values 0.991 and 0.861 of the ZR|P fit down to 0.749 and 0.181! When autocorrelation calculations with lags of 1, 2, 3... are carried out for all the relative

residual sets considered here, only the MWT [R''_{ri}] set yields significant results: for a lag of 1 the auto-correlation found is (0.692 ± 0.192) , a very significant result. Thus MWT increases serial correlation greatly. When the MWT [R''_{ri}] is cross-correlated with [LW], one finds values more than twice their estimated standard deviations for lags of 0, -1, -2, and -3.

It is interesting to note that the original input relative residuals involve a cross-correlation of -0.414

between $[R'_r]$ and $[R''_r]$; the corresponding value of -0.453 for $ZR-ZR|P$ is close; but the value for $ZR-ZR|M$ is -0.284 . None of these values is large enough to indicate much significant correlation between the real and imaginary components, but again the MWT results yield less agreement with the input residuals. Even though the PWT correlation is appreciably larger in magnitude than the MWT one, such larger correlation does not keep PWT from yielding far better results than does MWT, contrary to Zoltowski's conclusions [39].

The foregoing results make it quite clear how such level transformations as Z to Y (or vice versa) can lead to undesirable results: worse overall fits and larger error in parameter estimates. In addition, it should be evident that at least in fits like those discussed here and very probably generally, MWT yields much degraded results compared to PWT. This conclusion is usually entirely obscured, as we show in tables 5 and 6, if one compares only $S_f \equiv S_{f_w}$ values, but it is made obvious by comparison of S_{f_r} values, those given by eq. (10) with R_{w_i} replaced by R_{r_i} . In conclusion, we therefore suggest that the LOMFP be used for immittance data analysis with zero or minimum data transformation and with PWT, VWT, or RWT in preference to MWT.

In future we plan to add further weighting choices to LOMFP and to provide, in addition to least squares, the capability of minimizing the sum of the absolute values of the residuals. Finally, anyone who has been using a version of LOMFP dated earlier than December 1986 may request a free update to the present more extensive, general, and flexible version, as described herein, by writing one of us (J.R.M.).

Acknowledgement

One of us (J.R.M.) wishes to thank all of his past associates who have contributed to LOMFP. We also thank the Army Research Office for their support.

References

[1] R.W. Hamming, Numerical methods for scientists and engineers (McGraw-Hill, New York, 1962).

- [2] J.R. Macdonald, J. Schoonman and A.P. Lehen, Solid State Ionics 5 (1981) 137.
- [3] R.L. Hurt and J.R. Macdonald, Solid State Ionics 20 (1986) 111.
- [4] J.R. Macdonald and G.B. Cook, J. Electroanal. Chem. 168 (1984) 335; 193 (1985) 57.
- [5] J.R. Macdonald, J. Appl. Phys. 58 (1985) 1955, 1971.
- [6] I.D. Raistrick, Solid State Ionics 18/19 (1986) 40.
- [7] J.R. Macdonald, ed., in: Impedance spectroscopy, emphasizing solid materials and systems (Wiley-Interscience, New York, 1987).
- [8] J.J. Fontanella, M.C. Wintersgill, M.K. Smith and J. Semancik, J. Phys. 60 (1986) 2665.
- [9] J.R. Macdonald and J.A. Garber, J. Electrochem. Soc. 124 (1977) 1022.
- [10] J.R. Macdonald, J. Schoonman and A.P. Lehen, J. Electroanal. Chem. 131 (1982) 77.
- [11] Y.-T. Tsai and D.H. Whitmore, Solid State Ionics 7 (1982) 129.
- [12] B.A. Boukamp, Solid State Ionics 18/19 (1986) 136.
- [13] B.A. Boukamp, Solid State Ionics 20 (1986) 31.
- [14] D.R. Franceschetti and J.R. Macdonald, J. Electroanal. Chem. 82 (1977) 271.
- [15] J.R. Macdonald, J. Electroanal. Chem. 66 (1975) 143.
- [16] J.R. Macdonald and D.R. Franceschetti, J. Chem. Phys. 68 (1978) 1614.
- [17] J.R. Macdonald and R.L. Hurt, J. Electroanal. Chem. 200 (1986) 69.
- [18] J.R. Macdonald, J. Electroanal. Chem., to be published.
- [19] J.R. Macdonald, IEEE Trans. on Electrical insulation EI-16 (1981) 65.
- [20] D.R. Franceschetti and J.R. Macdonald, J. Electroanal. Chem. 101 (1979) 307.
- [21] J.R. Macdonald, J. Chem. Phys. 40 (1964) 1792.
- [22] H. Vogel, Physik. Z. 22 (1921) 645.
- [23] H. Fricke, Philos. Mag. 14 (1932) 310.
- [24] K.S. Cole and R.H. Cole, J. Chem. Phys. 9 (1941) 341.
- [25] J.R. Macdonald, Solid State Ionics 13 (1984) 147.
- [26] J.R. Macdonald, Solid State Ionics 15 (1985) 159.
- [27] D. Ravaine and J.-L. Souquet, C. R. Acad. Sci. (Paris) 277C (1973) 489.
- [28] S. Havriliak and S. Negami, J Polym. Sci. C14 (1966) 99.
- [29] J.R. Macdonald and R.H. Hurt, J. Chem. Phys. 84 (1986) 496.
- [30] R. Kohlrausch, Ann. Phys. (Leipzig) 12 (1847) 393.
- [31] G. Williams and D.C. Watts, Trans. Faraday Soc. 66 (1970) 80.
- [32] A.K. Johscher, The universal dielectric response: a review of data and their new interpretation (Chelsea Dielectrics Group, Chelsea College, University of London, 1978) pp. 27, 29.
- [33] J.R. Macdonald, J. Appl. Phys. 61 (1987) 700.
- [34] D.R. Franceschetti and J.R. Macdonald, J. Electroanal. Chem. 101 (1979) 316.
- [35] D.R. Franceschetti and J.R. Macdonald, J. Electrochem. Soc. 129 (1982) 1754.
- [36] S.W. Liu, Phys. Rev. Letters 55 (1985) 529.

- [37] T. Pajkossy and L. Nyikos, *J. Electrochem. Soc.* 133 (1986) 2061.
- [38] K. Funke, *Solid State Ionics* 18/19 (1986) 183.
- [39] P. Zoltowski, *J. Electroanal. Chem.* 178 (1984) 11.
- [40] A. Celmins, in: *Mathematical programming with data perturbations II*, ed. A.V. Fiacco (Marcel Dekker, New York, 1983) pp. 131–152.
- [41] K. Nielsen, *Acta Crystallogr.* A33 (1977) 1009.
- [42] A. Celmins, *Comp. Chem.* 8 (1984) 81.