PERMITTIVITY PROFILE RECONSTRUCTIONS USING TRANSIENT ELECTROMAGNETIC REFLECTION DATA

P. Fuks
Dept. of Electromagnetic Theory
Royal Institute of Technology
S-100 44 Stockholm, Sweden

G. Kristensson
Dept. of Electromagnetic Theory
Lund Institute of Technology
P.O. Box 118
S-221 00 Lund, Sweden

G. Larson
Dept. of Electromagnetic Theory
Royal Institute of Technology
S-100 44 Stockholm, Sweden

1. Introduction
2. Theory
   2.1 Mathematical Model
   2.2 Scattering Representation and Wave Splitting
   2.3 Invariant Imbedding
   2.4 The Green Functions
3. General Considerations
4. Deconvolution
5. Experimental Set-up
6. Analysis of Errors
   6.1 Signal Quality
   6.2 Cable Loss
   6.3 Internal Reflection
6.4 Oscilloscopes
6.5 Input Impedance Error
6.6 Baseline Error
6.7 SNR of an A/D Converter
6.8 Amplitude Linearity
6.9 Jitter and Noise
6.10 Sweep Errors
6.11 Bandwidth
6.12 Residual Ringing

7. Results

References

1. INTRODUCTION

In this paper an experimental implementation of two inversion algorithms for inhomogeneous dielectric media is presented. The first algorithm is based upon an imbedding procedure which was suggested originally by Corones, Davison and Krueger [2]. The second algorithm utilizes the Green functions of the problem and this approach was first introduced by Krueger and Ochs [7]. These algorithms and the underlying theory are presented in Section 2.

The permittivity profile is assumed to vary only with depth and the variation is assumed to be continuous. Otherwise, it is arbitrary. Thus, no assumptions are made about a piecewise constant permittivity profile. The assumption on a continuous permittivity profile can be relaxed, and extensions of the theory so that finite jump discontinuities in the permittivity profile can be treated are possible, see [5–6]. These extensions are, however, not presented in this paper. Furthermore, it is assumed that the medium is dispersion free and lossless. It is, however, possible to cope with lossy and dispersive profiles, but this requires an extension of the theory presented in Section 2. The theory of lossy and dispersive media are presented in [3–6] and [1], respectively.

To meet the assumptions made above and to have an easily controlled environment for the experiment a coaxial component set-up is adopted. The one-dimensional variation in the permittivity is then easy to realize and the dispersion free wave propagation in the wave guide is guaranteed by the TEM wave mode.

One of the prominent advantages of the theory and algorithms used in this paper is that they only require a finite time trace of reflection data as input data for the inversion of the permittivity profile. More
specifically, one round trip of reflection data is needed, where one round trip is defined as the time it takes for the pulse to propagate through the medium and back again. The design of the experiment can therefore be made so that all unwanted reflections from the experimental set-up arrive after one round trip. In this way these reflections do not affect the measurements and it is possible to avoid some unwanted interference with the experimental set-up. This is not possible with an experimental set-up that measures the fixed frequency response of the dielectric sample.

Some general considerations are found in Section 3. In Section 4 the deconvolution problem encountered in this paper is described and analyzed. The experimental set-up is described in Section 5 and a detailed presentation of the error analysis is found in Section 6. Finally, in Section 7 the results of this paper are presented.

2. THEORY

In this section the mathematical model is presented, some notations are introduced and the precise statement of the inverse problem is given.

2.1 Mathematical Model

An inhomogeneous slab occupies the region $0 \leq z \leq L$. The slab is assumed to be lossless and is modeled by a relative permittivity $\epsilon(z)$ that varies with depth $z$. A homogeneous lossless medium is situated on each side of the slab. These homogeneous media have a constant permittivity $\epsilon_1$ and $\epsilon_2$, respectively, see Figure 1. The permeability is that of vacuum everywhere in space.
To simplify the theoretical analysis in this section it is assumed that the permittivity profile is continuous at the ends, $z = 0$ and $z = L$, respectively, and, furthermore, continuously differentiable inside the slab, $0 < z < L$. A typical plot of the permittivity profile is shown in Figure 2. The assumption of continuity at the left and right ends of the slab is no loss of generality, see e.g., [5], and for the purpose of this paper, the present assumptions are quite sufficient. More explicitly, reflection and transmission data for a permittivity profile with finite jump discontinuities at the edges can always be transformed to reflection and transmission data for a permittivity profile where these jumps have been removed, i.e., the permittivity profile is continuous everywhere. These transformations consist of solving Volterra integral equations of the second kind. These equations are well-posed and numerically efficient methods to solve these integral equations are easy to find. If the jump discontinuities are inside the slab similar methods apply.

The profile is excited by an electromagnetic wave impinging normally on the slab from the left, i.e., the field does not depend on $x$ or $y$. Assuming that all fields only depend spatially on the depth $z$, the Maxwell equations in the absence of free charges and currents im-

**Figure 2.** The permittivity profile $\epsilon(z)$ as a function of depth $z$. 

ply (the fields $E$ and $D$ are assumed to have components along the $x$-axis and $B$ and $H$ along the $y$-axis)

$$\begin{align*}
\frac{\partial}{\partial z} E(z,t) &= -\frac{\partial}{\partial t} B(z,t) \\
\frac{\partial}{\partial z} H(z,t) &= -\frac{\partial}{\partial t} D(z,t).
\end{align*}$$

(2.1)

Here $\partial_z$ and $\partial_t$ denote (partial) differentiation with respect to depth $z$ and time $t$, respectively. The constitutive relations

$$\begin{align*}
D(z,t) &= \varepsilon_0 \varepsilon(z) E(z,t) \\
H(z,t) &= B(z,t)/\mu_0,
\end{align*}$$

and (2.1) then imply that the electric field $E(z,t)$ inside the slab satisfies the wave equation

$$\frac{\partial^2}{\partial z^2} E(z,t) - c^{-2}(z) \frac{\partial^2}{\partial t^2} E(z,t) = 0,$$

(2.2)

where the local phase velocity $c(z) = \{\varepsilon_0 \varepsilon(z) \mu_0\}^{-\frac{1}{2}}$.

### 2.2 Scattering Representation and Wave Splitting

To the left of the slab, $z < 0$, the electric field is a sum of two parts, one right going incident wave, $E^{i}(t)$, and one left going reflected wave, $E^{r}(t)$. Similarly, to the right of the slab, $z > L$, the electric field consists of one right going transmitted wave, $E^{t}(t)$. The total field $E(z,t)$ outside the slab is therefore

$$E(z,t) = \begin{cases} 
E^{i}(t - z/c(0)) + E^{r}(t + z/c(0)), & z < 0 \\
E^{t}(t - \tau - (z - L)/c(L)), & z > L
\end{cases}$$

where $c(0) = \{\varepsilon_0 \varepsilon_1 \mu_0\}^{-\frac{1}{2}}$ and $c(L) = \{\varepsilon_0 \varepsilon_2 \mu_0\}^{-\frac{1}{2}}$ are the phase velocities on the left and right hand side of the slab, respectively (remember the continuity in the permittivity profile at the ends). The quantity $\tau = \int_{0}^{L} \sqrt{\varepsilon_0 \varepsilon(z) \mu_0} dz$ is the time it takes for the wave front to go through the slab from $z = 0$ to $z = L$.

The incident and the scattered fields, respectively, are related by scattering operators. These relations are integral operators represented by

$$E^{r}(t) = \int_{-\infty}^{t} R^{+}(t - t')E^{i}(t')dt'$$

(2.3)

$$E^{t}(t) = \sqrt{\frac{c_2}{c_1}} \left\{ E^{i}(t) + \int_{-\infty}^{t} T^{+}(t - t')E^{i}(t')dt' \right\},$$
where the kernels $R^+(t)$ and $T^+(t)$ are the reflection and the transmission kernels of the slab, respectively, for an incident wave from the left. These kernels are independent of how the slab is excited, i.e., totally determined by the $\varepsilon(z)$ profile. Notice that if $E^i(t) = \delta(t)$ (where $\delta$ is the Dirac delta function) then it follows that $E^i(t) = R^+(t)$ and $E^i(t) = \sqrt{\varepsilon_2 / \varepsilon_1} \{\delta(t) + T^+(t-t')\}$. Hence, the scattering kernels $R^+$ and $T^+$ are the impulse responses of the medium.

One of the keystones in the theory of this paper is the wave splitting transformation. This is a transformation of dependent variables from the pair $\{E, \partial_z E\}$ to another pair $\{E^+ , E^-\}$ defined by [2, 3]

$$E^\pm(z,t) = \frac{1}{2} \left\{ E(z,t) \mp c(z) \int_{-\infty}^{t} \partial_z E(zt') dt' \right\}.$$ 

This wave splitting transformation can be written in a matrix shorthand notation as

$$\begin{pmatrix} E^+ \\ E^- \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & -c \partial_z^{-1} \\ 1 & c \partial_z^{-1} \end{pmatrix} \begin{pmatrix} E \\ \partial_z E \end{pmatrix} = T \begin{pmatrix} E^+ \\ E^- \end{pmatrix}.$$ 

(2.4)

The operator $T$ has a formal inverse

$$T^{-1} = \begin{pmatrix} 1 & 1 \\ -c \partial_z^{-1} & c \partial_z^{-1} \end{pmatrix},$$

that will be used below.

In a region where the phase velocity $c$ is constant this wave splitting transformation has the effect of projecting out the left and the right going parts of the field. More explicitly, in a region where the phase velocity $c$ is constant the general solution to (2.2) is

$$E(z,t) = f(t - z/c) + g(t + z/c),$$

where $f$ and $g$ are arbitrary functions. It is then easy to calculate the fields $E^+(z,t)$ and $E^-(z,t)$ defined in (2.4). They are

$$\begin{cases} E^+(z,t) = f(t - z/c) \\ E^-(z,t) = g(t + z/c). \end{cases}$$

In a region where $c$ is not constant the transformation defined in (2.4) is still well-defined. In this case the fields $E^+(z,t)$ and $E^-(z,t)$ are
defined as the left and the right going parts of the field, respectively, even though no interpretation such as in the constant phase velocity case above can be made. Notice that

\[ E(z, t) = E^+(z, t) + E^-(z, t), \quad (2.5) \]

for all profiles.

The fields \( E^+(z, t) \) and \( E^-(z, t) \) satisfy the following partial differential equation

\[
\partial_z \begin{pmatrix} E^+ \\ E^- \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} E^+ \\ E^- \end{pmatrix}, \quad (2.6)
\]

where

\[
\begin{align*}
\alpha &= -c^{-1}\partial_t + \frac{c'}{2c} \\
\beta &= -\frac{c'}{2c} \\
\gamma &= -\frac{c'}{2c} \\
\delta &= c^{-1}\partial_t + \frac{c'}{2c}
\end{align*}
\]

This can most easily be seen by combining (2.2) and (2.4). This equation is equivalent to the wave equation, (2.2), and gives the dynamics of the fields \( E^+(z, t) \) and \( E^-(z, t) \).

2.3 Invariant Imbedding

Consider now a subsection \([z, L]\) of the region \([0, L]\), see Figure 3. Mathematically, the original problem, \([0, L]\), is imbedded in a family of problems where the left edge of the slab, \( z \), is the parameter that is varied.

The fields \( E^+(z, t) \) and \( E^-(z, t) \), defined at the position \( z \), are related to each other in a similar way as the incident field \( E^i(t) \) and the reflected field \( E^r(t) \) in (2.3) of the original physical problem are related to each other. This relation is represented as an integral operator as

\[ E^-(z, t) = \int_{-\infty}^{t} R^+(z, t-t')E^+(z, t') dt'. \quad (2.7) \]

The kernel \( R^+(z, t) \) can be interpreted as the reflection kernel for the subsection \([z, L]\), where the medium to the left of \( z \) is of constant
permittivity $\epsilon(z)$. The field $E^+(z,t)$ serves as an incident field, while $E^-(z,t)$ is a reflected field, for this subsection problem. For the special value $z = 0$, the physical reflection kernel $R^+(t)$ in (2.3) is identical to $R^+(0,t)$. Hence, the reflection kernel $R^+(t)$ for the physical region $[0, L]$ is imbedded in a family of subsection problems $[z, L]$ with reflection kernels $R^+(z,t)$.

The dynamics, (2.6), and the relation between the fields $E^+(z,t)$ and $E^-(z,t)$, given by (2.7), imply that the reflection kernel $R^+(z,t)$ satisfies a non-linear differential equation. Lengthy, but straightforward, calculations show that

$$\partial_z R^+(z,t) - \frac{2}{c(z)} \partial_t R^+(z,t) = \frac{c'(z)}{2c(z)} \int_0^t R^+(z,t-t') R^+(z,t') \, dt' \quad (2.8)$$

$$R^+(z,0) = \frac{1}{4} c'(z) \quad (2.9)$$

$$R^+(L, t) = 0. \quad (2.10)$$

It is intuitively clear that the dependence of the reflection kernel $R^+(z,t)$ on $z$ is related to the local properties of the slab at $z$. This is expressed mathematically in (2.9). Equation (2.10) implies that the reflection kernel is zero at $z = L$, i.e., no scatterer present. Notice that (2.8) is non-linear due to the convolution integral on the right hand side of the equation, and that (2.8) has a directional derivative in the $(z, -\frac{2}{c(z)})$ direction. This latter property solves the inverse problem, which now can be stated more explicitly.

The inverse problem solved in this paper is the reconstruction of the permittivity profile $\epsilon(z)$ from reflection data. Specifically, given
reflection data for one round trip, i.e., $R^+(t)$, for $0 \leq t \leq 2\tau$, find the permittivity profile $\epsilon(z)$, $0 \leq z \leq L$, see Table 1. The numerical algorithm based upon the embedding equation (2.8) is now presented.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Known</th>
<th>Sought</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>$\epsilon(z)$, $0 \leq z \leq L$</td>
<td>$R^+(t)$, $0 \leq t \leq 2\tau$</td>
</tr>
<tr>
<td>Inverse</td>
<td>$R^+(t)$, $0 \leq t \leq 2\tau$</td>
<td>$\epsilon(z)$, $0 \leq z \leq L$</td>
</tr>
</tbody>
</table>

Table 1.

In order to make the numerical computations more easy a normalized travel time coordinate transformation is made. The transformation is

$$
\begin{align*}
x &= x(z) = \int_0^z \frac{dz'}{\tau c(z')} \\
s &= t/\tau \\
R(x, s) &= \tau R^+(z, t),
\end{align*}
$$

where $\tau = \int_0^L \sqrt{\epsilon(z)} \mu_0 \, dz$ is the time it takes for the wave front to go through the slab. This transformation maps the slab $z \in [0, L]$ to the interval $x \in [0, 1]$, and the time $s$ is normalized so that $s = 1$ is the time it takes for the wave front to go through the slab. The new scaled reflection kernel $R(x, s)$ satisfies

$$
\partial_x R(x, s) - 2\partial_s R(x, s) = -\frac{1}{2} A(x) \int_0^s R(x, s-s') R(x, s') \, ds' \quad (2.12)
$$

$$
R(x, 0) = -\frac{1}{4} A(x) \quad (2.13)
$$

$$
R(1, s) = 0, \quad (2.14)
$$

where

$$
A(x) = -\tau c'(z(x)) = -\frac{d}{dx} \ln c(z(x)), \quad 0 < x < 1. \quad (2.15)
$$

The inverse transformations of (2.11) and (2.15) are

$$
z(x) = c(0) \tau \int_0^x \left\{ \exp \left\{ -\int_0^{x'} A(x'') \, dx'' \right\} \, dx' \right\}, \quad 0 < x < 1 \quad (2.16)
$$

$$
\epsilon(z(x)) = \epsilon_1 \exp \left\{ 2 \int_0^x A(x') \, dx' \right\}, \quad 0 < x < 1. \quad (2.17)
$$
Thus, the knowledge of \( A(x), 0 \leq x \leq 1 \) and the constants \( \tau \) and \( \epsilon_1 \) determine \( z, L \) and \( \epsilon(z), 0 \leq z \leq L \). How \( A(x), 0 \leq x \leq 1 \) can be reconstructed numerically from (2.12) and (2.13) is now presented.

The numerical implementation of the imbedding equation (2.12) is most easily done by writing (2.12) as

\[
\partial_x R(x, s - 2x) = -\frac{1}{2} A(x) \int_0^{s-2x} R(x, s - s') R(x, s') \, ds',
\]

and integrate from \( x - h \) to \( x \) and let then time be \( s + 2x \). The result is

\[
R(x, s) - R(x - h, s + 2h) = -\frac{1}{2} \int_{x-h}^x A(x') (R \ast R)(x', s + 2(x - x')) \, dx',
\]

where star \( \ast \) denotes time convolution, i.e.,

\[
(R \ast R)(x, s) = \int_0^s R(x, s - s') R(x, s') \, ds'.
\]

Introduce a uniform grid of points \((x_i, s_j)\) in \((x, s)\) space, where \( x_i = ih, s_j = 2jh, i = 0, 1, 2, \ldots, N, j = 0, 1, 2, \ldots, N - i, h = 1/N \) and \( N \) is an integer. Denote

\[
R_{i,j} = R(x_i, s_j), \\
A_i = A(x_i).
\]

With the use of the trapezoidal rule in (2.18) the following algorithm is easily found

\[
R_{i,j} = \left\{ R_{i-1,j+1} - \frac{h^2}{2} \left\{ A_i \sum_{k=1}^{j-1} R_{i,j-k} R_{i,k} \\
+ A_{i-1} \sum_{k=1}^{j+1} R_{i-1,j+1-k} R_{i-1,k} \right\} \right\} \left\{ 1 - \frac{h^2}{8} A_i^{-2} \right\}^{-1},
\]

where \( i = 1, 2, \ldots, N, j = 0, 1, 2, \ldots, N - i \), and the values of \( A_i \) are determined from (2.13) and by using (2.18) with \( s = 0 \)

\[
A_i = -4R_{i-1,1} \left\{ 1 + \frac{h^2}{8} A_{i-1}^{-2} \right\}.
\]
The error made in (2.19) and (2.20) is of order $O(h^3)$.

The input data of the numerical algorithm is

$$
R^+(2j\tau h), \quad i = 0, 1, 2, \ldots, N
$$

(2.21)

The reconstruction of the permittivity profile goes in two steps. First, reconstruct $A(x)$ from the reflection data $R^+$ data (the constant $\tau$ is also needed determined by the time scaling). Then from $A(x)$ and the constants $\tau$ and $\epsilon_1$ determine $z(x)$ and $\epsilon(z(x))$. This is done from (2.16) and (2.17). The length of the slab is also determined by $L = z(1)$.

The initialization of the numerical algorithm is made by assigning

$$R_{0,j} = R(0, 2jh) = \tau R^+(0, 2j\tau h) = \tau R^+(2j\tau h), \quad j = 0, 1, 2, \ldots, N.$$  

The algorithm then proceeds as follows, see also Figure 4.

1. Use (2.20) to calculate $A(x_i)$ from data at grid line $i - 1$. The two arrows in Figure 4 indicate this operation.
2. Use (2.19) to calculate $R_{i,j}$, $j = 1, 2, \ldots, N - i$ from old and new data at grid lines $i - 1$ and $i$, respectively. Notice that the right hand side of (2.19) is now known from the previous step.
3. Repeat the previous steps to move one grid line deeper into the medium, until the right edge of the slab, $i = N$, is reached.

### 2.4 The Green Functions

It is also possible to arrive at a numerically faster algorithm by using another relation between the fields $E^+(z,t)$ and $E^-(z,t)$, defined at the position $z$. The representation used to derive the imbedding equation in Section 2.3, given by (2.7), gives the reflection kernel for the subsection problem $[z,L]$. It is also possible to relate the fields $E^+(z,t)$ and $E^-(z,t)$ to the external excitation $E^+(0,t)$. These relations are integral operators represented by

$$E^+(z,t) = \sqrt{\frac{c(z)}{c_1}} \left\{ E^+(0,t - \tau x(z)) 
+ \int_{-\infty}^{t-\tau x(z)} G_1(z,t - t')E^+(0,t') dt' \right\}$$  

(2.22)
Figure 4. The numerical algorithm.

\begin{align}
E^{-}(z,t) &= \sqrt{\frac{c(z)}{c_1}} \int_{-\infty}^{t-x(z)} G_2(z,t-t')E^+(0,t')dt', \quad (2.23)
\end{align}

where \(x(z)\) is defined in (2.11). The two kernels \(G_1(z,t)\) and \(G_2(z,t)\) are called the Green functions. The kernels \(G_1\) and \(G_2\) are ordinary functions; the singular part is represented by the first term on the right hand side of (2.22). The sum of (2.22) and (2.23) gives the total internal field in the slab, see (2.5) and Figure 5. This is not true for the sum of the fields \(E^+(z,t)\) and \(E^-(z,t)\) in Section 2.3 since the left end point of the slab is varied and thus the physical set-up changed.

The representations in (2.22) and (2.23) lead to a very efficient way to calculate the internal field, see [7]. However, calculations of the internal fields are not the main topic of this paper and this matter is, therefore, not pursued here, since the focus is on solving the inverse problem.
The following boundary conditions on the Green functions are easy to see by letting $z = 0$ in (2.22) and (2.23), and comparing with (2.7).

\begin{align*}
G_1(0, t) &= 0 \\
G_2(0, t) &= R^+(0, t) = R^+(t).
\end{align*}

Following the same line of analysis as in Section 2.3, the dynamics, (2.6), and this new relations between the fields $E^+(z, t)$ and $E^-(z, t)$, given by (2.22) and (2.23), imply that the Green functions $G_1(z, t)$ and $G_2(z, t)$ satisfy a system of first order linear differential equations. Lengthy, but straightforward, calculations show

\begin{align*}
\partial_z G_1(z, t) + \frac{1}{c(z)} \partial_t G_1(z, t) &= -\frac{c'(z)}{2c(z)} G_2(z, t), \quad (2.24) \\
\partial_z G_2(z, t) - \frac{1}{c(z)} \partial_t G_2(z, t) &= -\frac{c'(z)}{2c(z)} G_1(z, t), \quad (2.25)
\end{align*}

with the boundary conditions

\begin{align*}
G_1(z, \tau x(z)) &= \frac{1}{8} \int_0^z \frac{c'(z')^2}{c(z')} dz' \quad (2.26) \\
G_2(z, \tau x(z)) &= \frac{1}{4} c'(z). \quad (2.27)
\end{align*}

The numerical computations, just as in Section 2.3, become easier if a normalized travel time coordinate transformation is made.

\[ x = x(z) = \int_0^z \frac{dz'}{\tau c(z')} \]
\[ s = t/\tau \]
\[ G^+(x, s) = \tau G_1(z, t) \]
\[ G^-(x, s) = \tau G_2(z, t). \]

The system of first order linear differential equations then becomes

\[
\partial_x G^+(x, s) + \partial_s G^+(x, s) = \frac{1}{2} A(x) G^-(x, s) \tag{2.28}
\]
\[
\partial_x G^-(x, s) - \partial_s G^-(x, s) = \frac{1}{2} A(x) G^+(x, s), \tag{2.29}
\]

with the boundary conditions

\[
G^+(x, x) = -\frac{1}{8} \int_0^x A^2(x') \, dx' \tag{2.30}
\]
\[
G^-(x, x) = -\frac{1}{4} A(x), \tag{2.31}
\]

and where \( A(x) \) is defined as in (2.15).

For the numerical computations create a grid of points \((x_i, s_j)\) in \((x, s)\) space similar to the one in Section 2.3. The discretized points are defined by \( x_i = ih, \ s_{i+2j} = (i + 2j)h, \ i = 0, 1, 2, \ldots, N, \ j = 0, 1, 2, \ldots, N - i \), where \( h = 1/N \) and \( N \) is an integer. Denote

\[
G^+_i = G^+(x_i, s_{i+2j})
\]
\[
G^-_i = G^-(x_i, s_{i+2j})
\]
\[
A_i = A(x_i).
\]

With the use of the trapezoidal rule in (2.28) and (2.29) the following algorithm is easily found

\[
G^+_{i,j} = \begin{cases} G^+_{i-1,j} + \frac{h}{4} \left\{ A_i G^-_{i-1,j+1} + A_{i-1} G^-_{i-1,j} \right\} \\ + \frac{h^2}{16} A_i A_{i-1} G^+_{i-1,j+1} \left\{ 1 - \frac{h^2}{16 A_i^2} \right\}^{-1} \end{cases} \tag{2.32}
\]
\[
G^-_{i,j} = G^-_{i-1,j+1} + \frac{h}{4} \left\{ A_i G^+_{i,j} + A_{i-1} G^+_{i-1,j+1} \right\}, \tag{2.33}
\]
where \( i = 1, 2, \ldots, N \), \( j = 0, 1, 2, \ldots, N - i \). The discretized boundary values of \( G^+ \) and \( G^- \) in (2.30) and (2.31) become

\[
G_{i,0}^+ = G_{i-1,0}^+ - \frac{h}{16} \left\{ A_i^2 + A_{i-1}^2 \right\}
\]
(2.34)

\[
G_{i,0}^- = -\frac{1}{4} A_i
\]
(2.35)

where \( i = 1, 2, \ldots, N \). The error made in (2.32), (2.33) and (2.34) is of order \( O(h^3) \). The values of \( A_i \) are determined by solving

\[
A_i \left\{ 1 + hG_{i-1,0}^+ - \frac{h^2}{16} A_{i-1}^2 - \frac{h^2}{16} A_i^2 \right\} + 4 \left\{ G_{i-1,1}^- + \frac{h}{4} A_{i-1} G_{i-1,1}^+ \right\} = 0,
\]
(2.36)

for \( A_i \). This equation is obtained by letting \( j = 0 \) in (2.33) and then using (2.34) and (2.35).

The input data in the numerical algorithm are the same as in Section 2.3, given by (2.21), and the reconstruction proceeds in two steps. First a reconstruction of \( A(x) \) and then a transformation to \( z(x) \) and \( \epsilon(z(x)) \), where the latter step is identical to the one in Section 2.3. To find \( A(x) \), initialize \( G^+ \) and \( G^- \)

\[
G_{0,j}^+ = 0
\]
\[
G_{0,j}^- = R(0, 2jh) = \tau R^+(0, 2j\tau h) = \tau R^+(2j\tau h),
\]

where \( j = 0, 1, 2, \ldots, N \). The algorithm then proceeds similar to the one in Section 2.3.

1. Solve (2.36) for \( A_i \) from data at grid line \( i - 1 \).
2. Use (2.32) and (2.33) to calculate \( G_{i,j}^+ \) and \( G_{i,j}^- \), \( j = 1, 2, \ldots, N - i \) from old and new data at grid lines \( i - 1 \) and \( i \), respectively. Notice that the right hand side of (2.32) is now known from the previous step.
3. Repeat the previous steps to move one grid line deeper into the medium, until the right edge of the slab, \( i = N \), is reached.

This numerical algorithm based upon the Green functions approach is considerably faster than the one presented in Section 2.3, which was based upon an invariant imbedding procedure. This is simply due to the absence of time convolution integrals in the Green functions approach. However, the invariant imbedding approach provides a
stronger mathematical tool for the theoretical analysis and it is therefore motivated to present both methods here. It is also possible to compare the reconstructions made with the two algorithms. No significant difference in the accuracy of the reconstructions is, however, found for the class of profiles that is considered in this paper.

3. GENERAL CONSIDERATIONS

The reconstruction algorithms presented in Section 2 require that the impulse response $R^+(t)$ of the medium is known for one round trip. However, this is an idealized experimental situation where the incident field is a Dirac delta function. The real incident pulse has always a finite width. The algorithm can be used if the reflected waveform is deconvolved with the incident waveform, i.e., the reflection kernel or impulse response $R^+(t)$ can be extracted from (2.3)

$$E^r(t) = \int_{-\infty}^{t} R^+(t-t')E^i(t') \, dt'.$$

This deconvolution is also very efficient as a tool for removing certain errors which will be discussed in Section 6 below, but as any tool it has limitations, which restrict its use. It is sometimes stated that the deconvolution or some other suitable data processing can significantly improve the experimental results. This is generally not true because the measured data contain noise. Thus, outside a limited frequency band the quality of the data is too low to be processed. This is easily demonstrated by the following experiment:

1. Measure the same signal twice and calculate the Discrete Fourier Transform (DFT) of the two measurements, see Figure 6a for an example of a measurement of the incident pulse and Figure 6b for its DFT.

2. Divide the Fourier transform of the first measurement with the Fourier transform of the second. The result is showed in Figure 7.

The expected result is that this quotient is identical to one and this is indeed the case at low frequencies. However, when the frequency approaches the upper frequency limit of the system, the quotient of two small noisy quantities results in violent oscillations. This effect can only be reduced by bandlimiting the measured signal in the time
Figure 6a. An example of a recording of the incident waveform. Time scale in ps.

Figure 6b. The DFT of Figure 6a. Frequency in GHz.

or frequency domain or both. This filtering, if carefully done, can somewhat extend the useful bandwidth. Additional information can be extracted from the “out of band” part of the data if additional properties of the signal are known by other means, i.e., a priori information. The more a priori knowledge about the unknown sample, the more specific algorithms can be used and the better results are expected. A good example to this is found in Section 6.6.

The main practical problem in deconvolution is to obtain a relevant incident field or waveform. The pulse must be measured with the same system as used for the actual measurement, i.e., for the reflected field or
waveform. This can be obtained by introducing a short in the sample cell at the position of the sample. However, this requires frequently disassembling and assembling of the cell which wears and tears the precision connectors.

4. DECONVOLUTION

The kernel $R^+(t)$, which is required by the reconstruction algorithm, is computed by a deconvolution in the frequency domain. There is also the possibility to perform the deconvolution in the time domain. A recent contribution to the time domain deconvolution utilizing the Singular Value Decomposition (SVD) is found in [10]. The time domain approach has not been pursued in the present paper.

The signal reflected from the sample is sampled in the time domain and transformed to the frequency domain through a DFT. Assume that $t_k = T \frac{k}{N}, k = 0, 1, 2, \ldots, N - 1$, is a partition of the time interval $[0, T]$ into $N$ equidistant points. Then the DFT of $f_k = f(t_k)$ is defined as

$$F_p = \sum_{k=0}^{N-1} f_k e^{-i\omega_p t_k} = \sum_{k=0}^{N-1} f_k W_{pk}, \quad p = 0, 1, 2, \ldots, N - 1,$$

(4.1)
where $\omega_p = \frac{2\pi p}{T}$, $p = 0, 1, 2, \ldots, N - 1$ and where

$$W = \exp \left\{ \frac{-i2\pi}{N} \right\}.$$ 

This can be computed in an efficient way using any of the FFT algorithms that are available. The inverse transform is

$$f_k = \frac{1}{N} \sum_{p=0}^{N-1} F_p e^{i\omega_p t_k}, \quad k = 0, 1, 2, \ldots, N - 1.$$ 

The limitation of this approach is that the sample points $z_k = \exp \left\{ \frac{i2\pi k}{N} \right\}$ are equally spaced around the unit circle in the complex $z$-plane, and $\Delta t \Delta f = \frac{T}{N} \frac{1}{T} = \frac{1}{N}$. Thus, the resolution in the time domain and the number of sampling points determine the resolution in frequency domain. This is often undesirable. In this paper the limited resolution in frequency domain was unacceptable. This problem can be solved by using “zero padding”, which means increasing the $N$ with a number of dummy zeros, but the computing time will then increase drastically. A more efficient way to remedy this problem, and to be free from the limitations inherent in the ordinary DFT, is to adopt another algorithm to generate the sampled amplitudes in the frequency domain - the Chirp z-Transform (CZT), see Ref. [9].

The CZT transform has the advantage that it is possible to choose the number of time-samples different from the number of frequency-samples and it is possible to calculate the frequency-samples only in the frequency interval of interest. The reference signal is also transformed to the frequency domain using the CZT transform.

The CZT algorithm is based on the identity $pk = \frac{1}{2} \left( p^2 + k^2 - (k - p)(k - p) \right)$. The DFT defined in (4.1)) is then

$$F_p = W^{\frac{pk^2}{\pi^2}} \sum_{k=0}^{N-1} g_k W^{-\frac{(k-p)^2}{2}}, \quad p = 0, 1, 2, \ldots, M - 1,$$

where $g_k = f_k W^{\frac{k^2}{2}}, k = 0, 1, 2, \ldots, N - 1$. This can be viewed as a three-step process consisting of two multiplications and one convolution, which can be performed using FFT. The most important feature of this approach compared to the standard FFT is that the number
of time samples need not be equal to the number of frequency samples, i.e., the spacing and the starting point of the argument of $F_p$ are arbitrary, see Ref. [9].

Before the CZT transformation, some windowing is sometimes necessary. Depending on the shape of the signals the window function is chosen so that a good compromise between frequency resolution and amplitude accuracy is obtained. No weighting function is ideal. The bandwidth, ripple, sidelobe level, sidelobe fall-off rate, shape factor can be optimized but not all with the same window. The Kaiser-Bessel weighting function, defined as

$$w(t) = 1 - 1.24 \cos \frac{2\pi t}{T} + 0.244 \cos \frac{4\pi t}{T} - 0.00305 \cos \frac{6\pi t}{T},$$  \hspace{1cm} (4.2)

is superior to the other filters with respect to selectivity. This makes it excellent for separation of closely spaced components with widely different levels.

A formal inverse DFT of the quotient of the DFT of the reflected and the incident waveforms, respectively, is not possible because of noise at the higher frequencies (above 12 GHz). The deconvolution problem is an ill-posed problem and filtering is necessary to yield a stable and physically consistent result. A regularization filter is used here to stabilize these ill-posedness effects. The reflection kernel $\hat{R}^+(\omega)$ in the frequency domain then has the form

$$\hat{R}^+(\omega) = \frac{Y(\omega)X(\omega)^*}{|X(\omega)|^2 + \lambda C(\omega)},$$

where $Y(\omega)$ and $X(\omega)$ are the Fourier transformed reflected and incident waveforms, respectively, and where star ($*$) denotes the complex conjugate. The function $C(\omega)$ is chosen as $C(\omega) = \omega^4$, see also Ref. [8] for more details about this choice. To determine the optimum value of the parameter $\lambda$ the procedure described above is used. Two recordings of the same signal are measured and deconvolved using one of them as a reference. The constant $\lambda$ is then adjusted so that a smooth transition from one to zero at the high frequency end of the spectrum is obtained. A Kaiser-Bessel window, see (4.2), is used on the kernel before it is transformed into the time domain. The transformation to the time domain is made by using an inverse CZT transform. It has the advantage that the number of time points, the beginning of
the time trace, and the length of the time trace for the kernel can be varied arbitrarily.

The deconvolution procedure is illustrated below in Section 7.

5. EXPERIMENTAL SET-UP

The measurement system can be described as a pulsed radar. A short pulse is generated and sent towards the sample to be investigated. The reflected waveform is recorded and using the algorithms described in Section 2.3 or 2.4 the permittivity profile of the sample is reconstructed.

The set-up is similar to a Time Domain Reflectometer (TDR). However, a short duration pulse is used as excitation instead of a step. A traditional TDR uses a through-line sampler. This gives the best possible signal-to-noise ratio (SNR) and the largest bandwidth. In our application a terminated sampler and a power divider is used because the amplitude errors due to multiple reflections are reduced by attenuations in the divider. The loss of resolution and dynamic range due to losses in the divider are insignificant. Because the algorithms used are developed for a one-dimensional case, the system is built using coaxial components. As long as the underlying PDE for the system is the one-dimensional wave equation, (2.2), it is a possible realization of the theory. Other TEM transmission lines, such as striplines, have the advantage of being open structures which makes it easier to insert and remove samples in the sample cell. Flat samples are, however, more difficult to machine than cylindrical ones. A section of a coaxial air line was used as a sample cell. The longest air line available is only 200mm long which requires short samples and, consequently, pulses with very short rise time have to be used in order to obtain adequate relative resolution.

The two main components of the system are the pulse generator (Picosecond Pulse Labs Model 3050A) and the sampling oscilloscope (Tektronix TDR/Sampler 7S12 with a S-6 head installed in a 7854 frame). The specifications of these components are summarized in Table 2 and 3, respectively, and an outline of the experimental set-up is given in Figure 8. The signal from the generator is connected to the sample cell and to the oscilloscope through a power divider optimized for broadband pulses. The generator is also a source of trigger pulses. The reflection from the sample is continuously monitored and can be sampled, averaged, and stored under the control of a desktop computer. Data are then transferred to a mainframe computer for processing.
Immediately after recording of the reflection data from an unknown sample a short is placed in its place and a new recording is made. This waveform is an inverted copy of the incident waveform (including all imperfections of the measurement system) and is used as a reference for deconvolution.

**Figure 8.** The experimental set-up.
Specifications

<table>
<thead>
<tr>
<th>Specifications</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pulse width (FWHM, 50%)</td>
<td>60ps</td>
</tr>
<tr>
<td>Pulse width (10%)</td>
<td>115ps</td>
</tr>
<tr>
<td>Rise &amp; Falltime (10%-90%)</td>
<td>45ps</td>
</tr>
<tr>
<td>Amplitude</td>
<td>4V</td>
</tr>
<tr>
<td>Residual ringing</td>
<td>±3%</td>
</tr>
<tr>
<td>Precursor</td>
<td>±1%</td>
</tr>
<tr>
<td>VSWR</td>
<td>3:1</td>
</tr>
<tr>
<td>Jitter</td>
<td>±2.5ps</td>
</tr>
<tr>
<td>Repetition rate</td>
<td>1MHz</td>
</tr>
<tr>
<td>Trigger risetime</td>
<td>0.5ns</td>
</tr>
</tbody>
</table>

Table 2. Specifications of the pulse generator PSPL Model 3050A.

Specifications

<table>
<thead>
<tr>
<th>Specifications</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risetime</td>
<td>30ps</td>
</tr>
<tr>
<td>Displayed noise</td>
<td>5mV</td>
</tr>
<tr>
<td>Jitter</td>
<td>10ps</td>
</tr>
<tr>
<td>Maximum input voltage</td>
<td>1V p-p</td>
</tr>
<tr>
<td>Input resistance</td>
<td>10kOhm</td>
</tr>
<tr>
<td>Deflection factor accuracy</td>
<td>3%</td>
</tr>
<tr>
<td>Time/Div accuracy</td>
<td>3%</td>
</tr>
</tbody>
</table>

Table 3. Specifications of the Tektronix TDR/Sampler 7S12 with S-6 head installed in a 7854 frame.

The following dielectric materials have been used (data supplied by the manufacturer)
1. PTFE \( \varepsilon_r = 2.0 \) at 1 MHZ
2. POM \( \varepsilon_r = 3.7 \) at 1 MHZ
3. PA \( \varepsilon_r = 3.5 - 3.7 \) at 1 MHZ
6. ANALYSIS OF ERRORS

In this section a detailed analysis of the errors of the experimental set-up is presented.

6.1 Signal Quality

The most important parameter of the signal source is the rise time of the pulse, because it determines the resolution of the system. The generator is one of the fastest that is available commercially. Moreover, it has two important characteristics: relatively large output voltage, which gives a good SNR of the system, and very little residual ringing. This last property is essential, because low level perturbations on the baseline, which can only partially be removed from the recorded data, will seriously degrade the performance of the system. The generator also gives clean and stable trigger pulses which reduces the system jitter.

6.2 Cable Loss

As the pulse travels down the line it is attenuated by dielectric losses and finite losses in the conductors. The losses are frequency dependent and high frequencies are generally more attenuated. This degrades the risetime of the pulse and limits the distance resolution. To avoid this problem the cables are kept as short as possible. However, certain lengths are necessary to avoid multiple reflections and to create a clean portion of the baseline.

6.3 Internal Reflection

The discontinuities present in the pulse source, connectors and other parts of the system will degrade the performance. Consider a transmission line with two discontinuities on it. If the first one has a reflection coefficient of 0.1, the error in determining the reflection coefficient of the other will be 6%. The connectors (APC7 and SMA) contribute very little to this error when they are new. However, after a few months of use, even if handled with care, reflections became noticeable. This effect can be reduced using deconvolution.
6.4 Oscilloscopes

The fastest available oscilloscopes have a rise time of the order of 300 ps. For analysis of faster signals the sampling technique must be used. The obvious drawback of this approach is that only repetitive waveforms can be captured. The main problem is that the design of the sampling bridge makes it difficult to determine the baseline level.

6.5 Input Impedance Error

The input of the oscilloscope has an impedance of 50 Ohms. For Tek S-6 terminated by a HP load, the VSWR is \(< 1.05\) which corresponds to a variation in resistance of 47.6 to 52.5 Ohms. When connected to a 50 Ohms source, the measured voltage can differ by 2.5% from the correct value. The attenuators, which generally have lower performance, contribute to this error. However, if, as in our measurements, all data are measured using the same system layout and the same sensitivity settings, this effect is not noticeable. The discontinuities cause reflections back into the source. If the source is not matched, these reflections will be re-reflected and will distort the measured waveform. This can be avoided by using a suitable time window.

6.6 Baseline Error

Due to the design of the sampling head, it is difficult to determine the exact position of the baseline. The reconstruction algorithm assumes that this position is at zero level. If this is not the case, the error is integrated and an erroneous slope in the reconstructed permittivity profile is introduced. Remember, that without any \(a\ priori\) knowledge, the reconstruction algorithms in Section 2.3 or 2.4, will always generate a permittivity profile that corresponds to the given reflection kernel including all errors. It is possible to eliminate this problem by an offset of the recorded data and a repetition of the calculation until the relative permittivity outside the sample is 1. This procedure requires that the approximate length of the sample be known, which is the case in our tests, but not in general. This corresponds to the \(a\ priori\) knowledge introduced into the inverse scattering problem. This baseline error, if not corrected, can easily create a 20% error in the reconstructed permittivity profile. If more information is available (i.e., that the sample has constant or piecewise constant permittivity), additional degrees of correction can be introduced. Figures 9 a and b
clearly illustrate this effect.

**Figure 9a.** Reconstruction with 2% baseline offset. Length scale in m.

**Figure 9b.** Reconstruction with corrected baseline error. Length scale in m.
6.7 SNR of an A/D Converter

A 10 bits converter in the TEK oscilloscope gives a SNR = 62 dB, which is more than enough compared to the SNR of the signal.

6.8 Amplitude Linearity

All the measurements are made at the same sensitivity setting. Consequently, the absolute value of the amplitude error has no importance. The vertical linearity is investigated by measuring a number of identical waveforms which are offset by different amounts. This offset is then compensated by adding a constant to the stored waveform. No significant differences are detected when the processed waveforms are compared, as long as the input is kept under 250 mV.

6.9 Jitter and noise

Jitter is the random time uncertainty of a waveform point relative to a reference point. In a TDR application jitter is particularly critical because the information is extracted from time relations. The amount of jitter depends on the slope of the trig signal and the amount of noise in it. During the sampling of a signal with a slope different from zero, values randomly distributed around the signal value are recorded. This makes jitter difficult to distinguish from amplitude noise. However, the amount of noise due to jitter is largest at the large slope portions of the signal as opposed to the amplitude noise.

Jitter always limits the bandwidth of the system because the distribution of the noise around the signal is not symmetric. The mean value at points where the signal has negative second derivative is below the signal and where the second derivative is positive the mean value is above the signal. After averaging the original waveform will be reconstructed only for segments with constant slope. All other parts of the signal will be smoothed which results in loss of resolution. The distortion caused by this effect cannot be determined exactly because the probability distribution of the error depends on the waveform and is therefore unknown. It can be estimated, however, by measuring the noise at the most vertical part of the signal. Both kinds of noise can be removed from the signal by measuring it several times and then averaging. Generally, the larger the number of recordings averaged, the better the signal quality becomes. However, the process takes time and if the number of recordings is growing one must take into account
the long-time drift of the oscilloscope. For the set-up reported in this paper, 100 times is about optimum.

6.10 Sweep Errors

Sweep errors consist of the absolute error and the linearity error. Both can be calibrated by using a stable sinewave and using the zero crossing points as time markers. The frequency of the wave must be high in order to obtain a large number of calibration points. Unfortunately, due to the design of the TEK oscilloscopes trigger circuit, the highest possible frequency is about 900 MHz which is not sufficient. This problem is solved by using a chain of frequency multipliers connected to a very stable 500 MHz source which is also used as a trigger. Alternatively, a high Q-value resonator is used, triggered by short pulses with repetition frequency of 1 MHz.

6.11 Bandwidth

Many kind of errors can be normalized or filtered out. The price for these improvements in signal quality is generally a loss of effective bandwidth. Thus, one should try to start with as large bandwidth as possible even when the required resolution does not motivate this. In this set-up all the components were specified for operation to at least 18 GHz. The resulting bandwidth of the system is 12 GHz. This corresponds to a resolution of about 7 mm for permittivity values in the range 2–4 where the measurements are made. A 20 mm long sample with constant permittivity showed a clean flat portion of permittivity in the middle of the profile.

6.12 Residual Ringing

The ringing, which mainly originates from the generator, can be divided into two parts. Some ringing is created in the step generating parts, and cannot be affected, and some is created in the following, pulse shaping components, and can be delayed with respect to the main pulse, see Figure 10. This makes it possible to create a "quiet zone" where the first ringing has died out and the second has not arrived yet. The sample should be placed in this zone. This zone limits the maximum length of the sample to about 90mm.
7. RESULTS

The general scheme for processing the recorded data is illustrated in Figures 11 a–11 j. These figures show the step by step procedure to obtain the inversion of the permittivity profile. The recorded incident and reflected waveforms, Figures 11 a and b, are first transformed into the frequency plane, Figures 11 c and d, using the transformation described in Section 4. Figures 11 e and f show the deconvolved reflection kernel in the frequency domain with different normalizations. The time domain behavior of the kernel from Figure 11 e is showed in Figures 11 g. Finally, in Figures 11 i and j the reconstructions corresponding to Figures 11 g and h, respectively, are showed illustrating the trade off between the resolution and amplitude accuracy when different windows are used.
Figure 11a. The incident waveform. Time scale in ps.

Figure 11b. The reflected waveform. Time scale in ps.
Figure 11c. The DFT of the incident waveform. Frequency in GHz.

Figure 11d. The DFT of the reflected waveform. Frequency in GHz.
Figure 11e. Deconvolved reflected waveform with optimum normalization. Frequency in GHz.

Figure 11f. Same as Figure 11e, but with insufficient normalization. Frequency in GHz.
Figure 11g. Deconvolved reflected waveform with optimum windowing. Time scale in ns.

Figure 11h. As Figure 11g, but with rectangular windowing. Time scale in ns.
Figure 11i. Permittivity profile reconstruction from waveform in Figure 11g. Length scale in m.

Figure 11j. Permittivity profile reconstruction from waveform in Figure 11h. Length scale in m.
Two additional sets of figures illustrate the reconstruction of permittivity profiles which are inhomogeneous, see Figures 12–13 and 15–16. The original samples are depicted in Figures 14 and 17. In both the cases the samples are turned round and new measurements with these reflected profiles are made for comparison. The two reconstructions should be mirror images of each other. From these comparisons it is possible to conclude that the accuracy decreases slightly when the length of the sample increases. This is due to the losses. If the losses are a priori known, then these losses can be compensated for. No such compensation is made here.

Figure 12. Reconstruction of a composite sample. Length scale in m.

Figure 13. Reconstruction of a composite sample (sample turned round). Length scale in m.
Figure 14. The original sample of Figure 13.

Figure 15. Reconstruction of a composite sample. Length scale in m.
The resolution of the measurement system is also clearly shown. A comparison of the length of the samples (measures with a sliding calliper) with the length obtained from the reconstructions of the permittivity profiles shows an error of 2%. This is indeed a very sensitive test because the length of the medium is calculated from the reconstructed permittivity.

The last set of reconstructions, see Figures 15–17, also illustrates the resolution of the system in a different way. (The reconstruction depicted in Figure 15 has already been given in Figure 11 i, but it is repeated here to facilitate the comparison.) Between the two dielectric samples (of different lengths, but the same permittivity) there is an air-gap. The length of this air-gap can be varied. The smallest possible air-gap that still gives a fair reconstruction of the vacuum value of the permittivity is a measure of the resolution of the system. This is illustrated in Figures 15–16.

![Figure 16. Reconstruction of a composite sample (sample turned round). Length scale in m.](image)
Figure 17. The original sample of Figure 16.

REFERENCES


