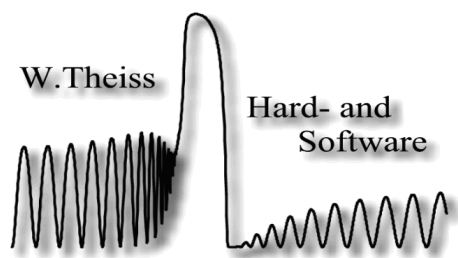**W.Theiss Hard- and Software for Optical Spectroscopy**
Dr.-Bernhard-Klein-Str. 110, D-52078 Aachen
Phone: (49) 241 5661390   Fax: (49) 241 9529100
E-mail: theiss@mtheiss.com   Web: www.mtheiss.com

Digit

A utility for digitizing scientific graphs

by Wolfgang Theiss

Printed: 9/17/2012, 1:50 PM in Aachen, Germany

Table of Contents

Digit

Part

I

1   Overview

1.1   About Digit

by Wolfgang Theiss

January 2012

W. Theiss – Hard- and Software for Optical Spectroscopy
Dr.-Bernhard-Klein-Str. 110, D-52078 Aachen, Germany
Phone: + 49 241 5661390   Fax: + 49 241 9529100
e-mail: theiss@mtheiss.com
web: www.mtheiss.com

Purpose

Digit is an easy to use program to get digitized data from scanned pictures containing scientific graphics.

Requirements

You must have a scanned bitmap in the Windows bitmap format (*.bmp) which can be created by most scanner software packages. If you have any other format like GIF, TIF, JPEG or anything else you must consult one of the numerous file conversion utilitities (Graphic Workshop, IPhotoDeluxe, Paint Shop Pro, ...) that convert various bitmap formats.
You should also have a reasonable screen resolution, i.e. at least 800 by 600 pixels. The more, the better.

1.2   Program overview

There are two ways of getting data:

1. Manual point selection

After the definition of three scaling points you simply click along the curve that you want to digitize. Every mouse click is recorded and the coordinates of the mouse cursor position are added to the list of data points.
This method is easy to learn and recommended for getting a few points in a quick manner. On the other hand, it is not very accurate and rather tedious if you want to collect many points.
Click here to learn more about this method.

### 2. Semi-automatic acquisition

This method is more accurate than the first one, and certainly much more elegant. On the other

hand, you have to learn a little more to use it.
After the definition of three scaling points and the background color you only have to move the cursor roughly along the curve to be digitized. All pixels whose color differs significantly from the background color are identified and recorded. Based on this set of points an equidistant set of output points is constructed by suitable averaging.
Click here to learn the handling of this method.

The main window of *Digit* is the following:
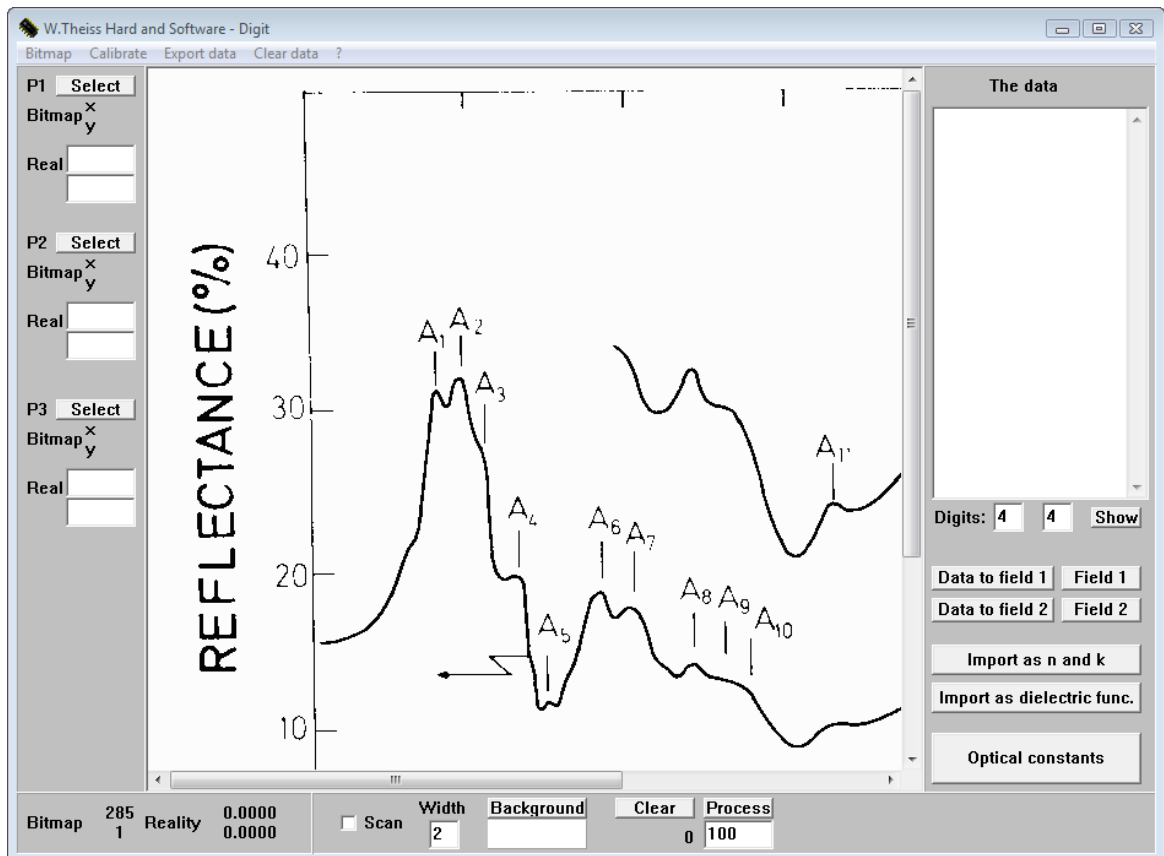
The center part of the window holds the image to be digitized. If the bitmap is larger than the space available in the window you can move the image using scrollbars.
The panel on the left side of the window holds the controls to set three definition points (P1, P2, P3). For these three points you have to enter the real word coordinates and you have to mark them on the bitmap by a mouse click. The **Calibrate** menu command activates the computation of the relation between pixel coordinates and real world coordinates.
After this calibration both the bitmap and the real world coordinates of the cursor position are displayed in the lower left corner of the window.
The lower right panel holds all items used for the semi-automatic scan. These will be explained later.
Finally, the right side named 'The data' is used to display the result, namely the real world coordinates of the digitized data. You get access to data fields which have methods to export the new data.
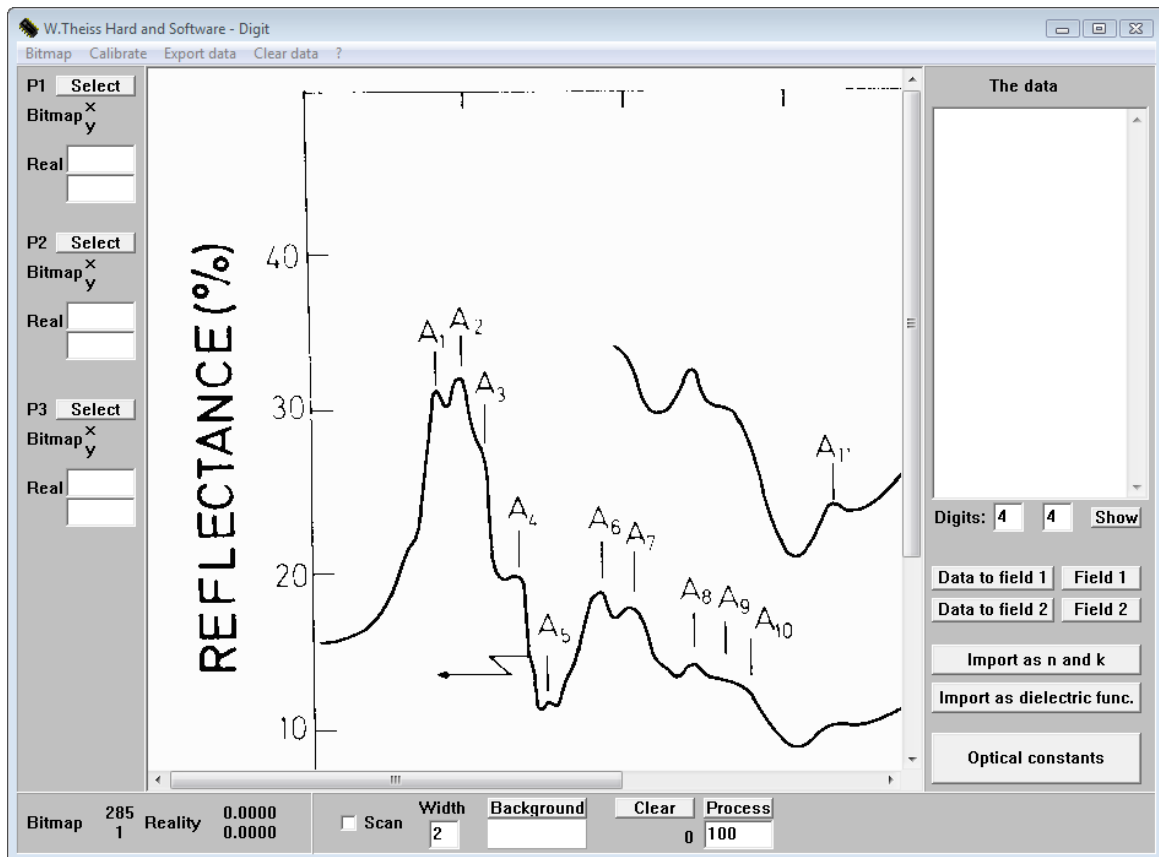
Digit

Part

II

2   Manual point selection

2.1   Step 1

This part of the manual shows you how to digitize points manually.

As first step, you should use the **Bitmap** command to load the image to be digitized (the bitmap DEMO.BMP can be found where digit.exe is located). You can load the image from a file or the clipboard:

Note that you might see just a small fraction of the whole picture. What you see is a one-to-one copy of the original bitmap, no zoom manipulations are used. Hence what you see on the screen depends on the scan resolution and on your screen resolution. Nevertheless, since you have the complete picture on paper anyway for your orientation, you can live with only a small fraction visible. Of course, you must use the scrollbars to move around.

2.2   Step 2

Depending on the quantities you want to digitize you should set the numbers of digits to be used (lower right corner, labeled as **Digits**:) for the obtained x- and the y-values (left number and right number, respectively). The default values are 4 digits for both x and y.

2.3   Step 3

Now you have to specify three points on the bitmap whose 'real life' coordinates are known. Of course, in most cases, you can make use of the axis labels to choose suitable points. Since these three points (labeled as P1, P2 and P3 on the left side of the windows) are used to calculate a conversion matrix from bitmap to 'real life' values there are two rules for their choice:

- **Do not choose all three points on one line**
- **Use large separations between the three points**

On the example bitmap which looks like this
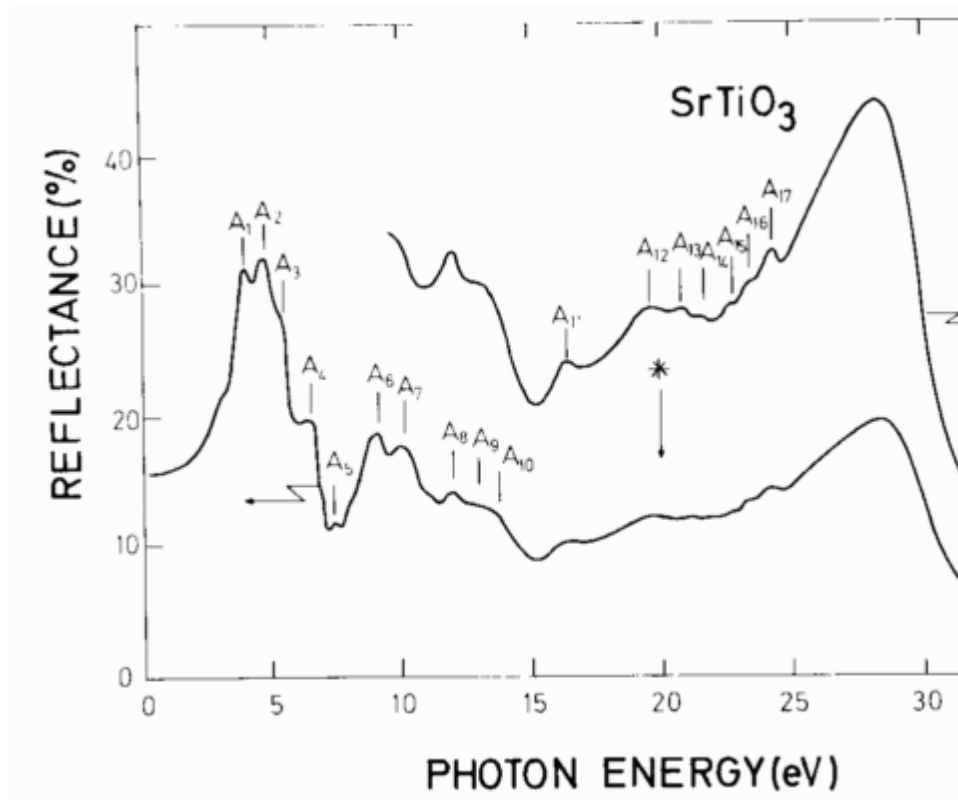
I would suggest the following points:

       P1: 0 eV, 40% reflectance
       P2: 0 eV, 0% reflectance
       P3: 30 eV, 0% reflectance

Here are the actions to do:
- Enter in the 'real' section of P1 the values 0 and 0.4 (i.e. 40% reflectance)
- Press the **Select** button and click on the point on the bitmap with these coordinates
- Check if the program has entered the correct bitmap coordinates in the bitmap section of P1 (I have got 129 and 152 as bitmap coordinates).

Do the same with P2 and P3:
- For P2 set the real values to (0,0) and obtain about (137,673) as bitmap coordinates
- For P3 set the real values to (30,0) and obtain about (920,668) as bitmap coordinates

2.4   Step 4

Select the menu item **Calibrate** to calculate the conversion matrix between bitmap and real coordinates. From now on the mouse coordinates on the bottom panel of the window should give the correct reality coordinates which should be checked by positioning the mouse cursor to points on the bitmap with known real values.

2.5   Step 5

Now the real work starts: whenever you click somewhere on the bitmap a new pair of x-a nd y-values is created in the data list on the right of the window. In addition a small circle is drawn on the bitmap to remind you where you have collected points already.
Click a few points and watch the list of data points grow.
What is left to do is to move along the curve to be digitized and collect as many points as desired. Before you start this real work you can free the list from previous points by **Clear list** and start with an empty list.
While collecting points you should know that the list of data points can be edited - in particular, wrong points (where you failed to click correctly) can be deleted. Be aware, nevertheless, that the list should not contain blank lines because this will cause errors when you process the data later.

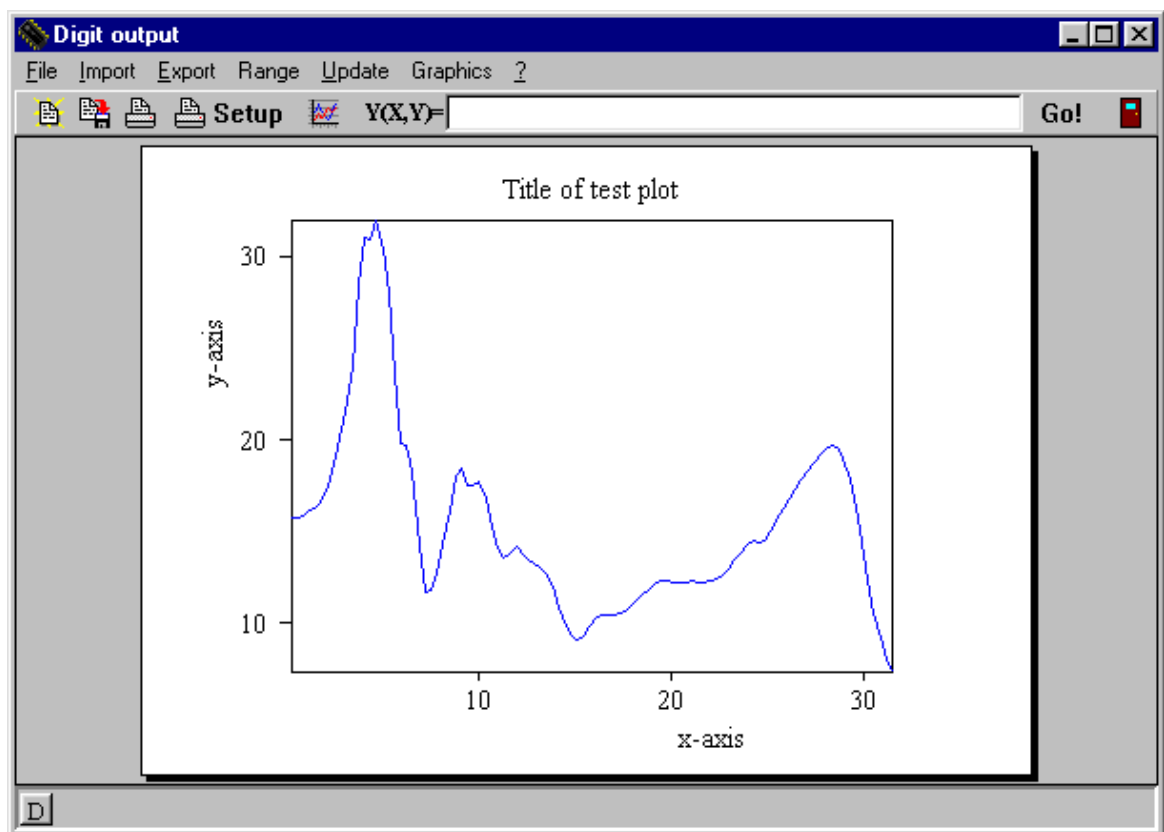In between the window may look like this:
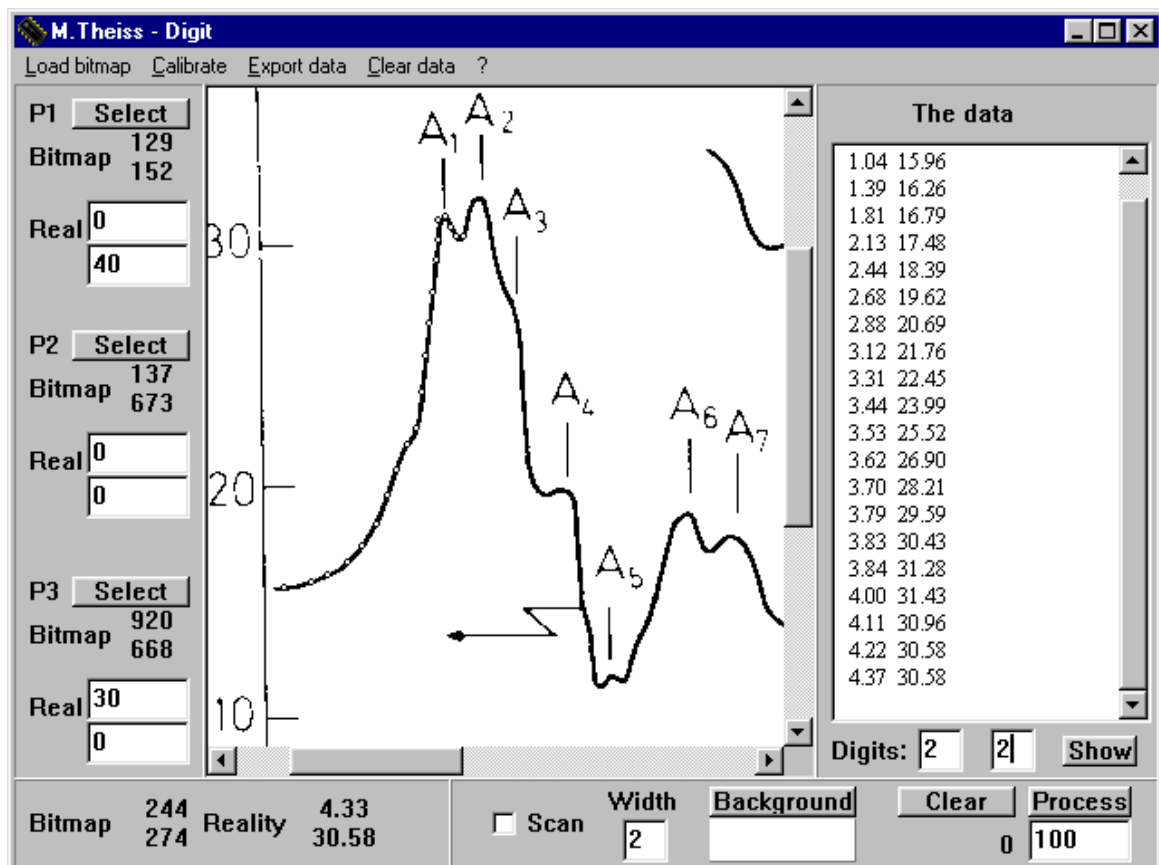
2.6   Step 6

Export the data using the **Export data** command. Files created this way are simple text files with columns of the obtained x- and y-values. They should have the extension *.xy. These files can be re-imported by most applications from *W.Theiss Hard- and Software* and others like Microsoft Excel. However, you can also use the Drag&Drop method described below to copy the data to other applications of *W.Theiss Hard- and Software*.
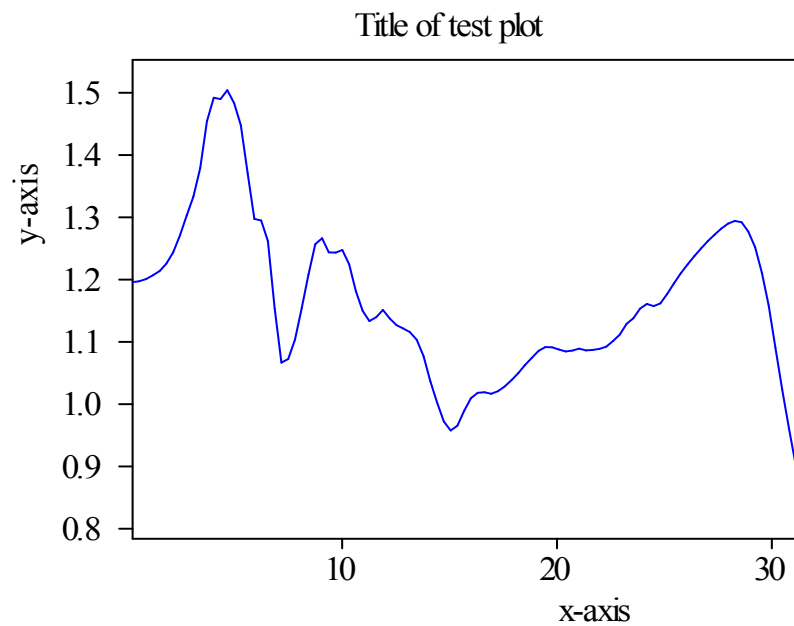
Note, however, that all *W.Theiss Hard- and Software* objects that require data sets with equidistant x-values (such as all fitting objects) interpolate the non-equidistant input data to an equal number of equidistant data points - hence the number of your input data points should be large enough to ensure a reasonable sampling even on equidistant 'x-grids'.

To examine your results you can also get a graphical view of the digitized data. Press the **Show** button and a second window opens showing the data:

Besides looking at the data, you can use this window also for data export in several formats and also for data manipulation.

In the line following **Y(X,Y)=** you can type in a formula (refer to the present y-values using the symbol 'Y' in the formula). Pressing the **Go!** button is followed by a re-computation of the data based on the formula. For example, the formula string 'log(y)' leads to the calculation of the logarithm (base 10) of the y-data:

Title of test plot

The formula editor as well as the handling of the graphics in this window is well-known to experienced W.Theiss Hard- and Software customers. The formula editor is described in the help of the program Data Factory that is shipped with Digit to registered users, the graphics handling is described in the helpfile graphics.hlp (also shipped with Digit).

Using the 'Export' submenu you can export the data to a file or to the clipboard. Alternatively you can write the data to the built-in workbook (using the Workbook submenu). Once the data are in the workbook you can copy&paste data blocks to other applications or write the whole workbook to an Excel file.

That's all for this simple, manual method to collect data points!

Click here if you think you are ready to learn about the more automatic and elegant way to digitize curves.

Digit

Part III
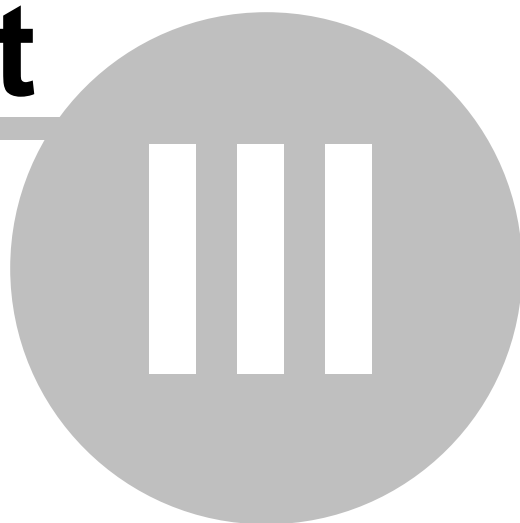
3   Semi-automatic scan

3.1   Step 1

Before the more semi-automatic data acquisition method can be used you have to perform **Steps 1 to 4 of the manual method** (Loading the image, setting the three definition points and doing the calibration).

Since the automatic method is based on the discrimination of the colors of data points and background you have to define the background color. Push the **Background** button in the Scan section of the main window and then click at a typical background pixel in the bitmap. While you move the cursor the rectangle below the **Background** button displays the current pixel color. After you have picked a background pixel the rectangle keeps its color to show you the selected background.

3.2   Step 2

Now check the checkbox labeled **'Scan'** to indicate that you want to use the semi-automatic scanning procedure.
Move the cursor close to the wanted curve and press the left mouse button down. Now, with the mouse kept down, move along the curve and see that some of the black points disappear. At the same time, the number of converted pixels is displayed below the **Clear** button.
Move the cursor several time across the curve. The main window may look like this after a while:
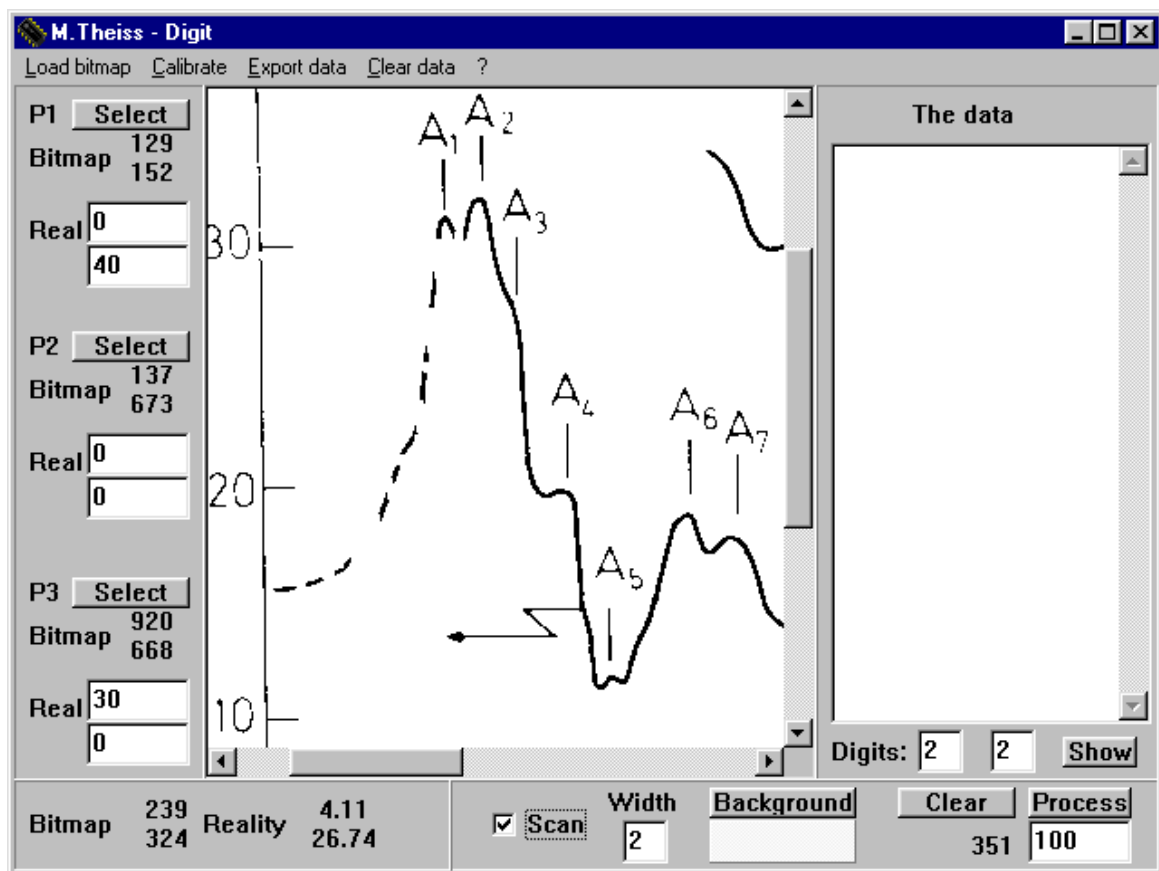
Note that there are no points in the data area yet.
What happened up to now is the following: While you move the cursor in the picture (with left mouse button pressed down) Digit tests all pixels in a rectangle around the present cursor position. The

size of the rectangle is given by the number you specify below **Width**. The entry of 2 means that a rectangle from -2 to +2 pixels centered at the cursor position (in both the x- and y-direction) is investigated, i.e. 25 pixels are tested. If the color of a pixel differs significantly from the background color, it is taken as a data point. Its coordinates are recorded internally and its color is set to the background color. This avoids that a pixel is counted twice. In addition, you can easily see which parts of the curve have already been digitized.

The difference of two colors is measured in the three-dimensional rgb-space. If the difference vector of the red, green and blue values has a length above 100 the two pixels are considered to be different.

3.3   Step 3

To digitize the whole curve you have to convert all data pixels to background pixels. Of course, this is quickly done by setting a large **Width** value, but you have to avoid to pick wrong points such as black labels or arrows. During the scan you can change the Width and continue with small values in regions close to parts of the picture which are no wanted data points.

After a while I managed to scan all pixels of the bottom curve of the picture so that the curve has disappeared completely (all pixels converted to background):
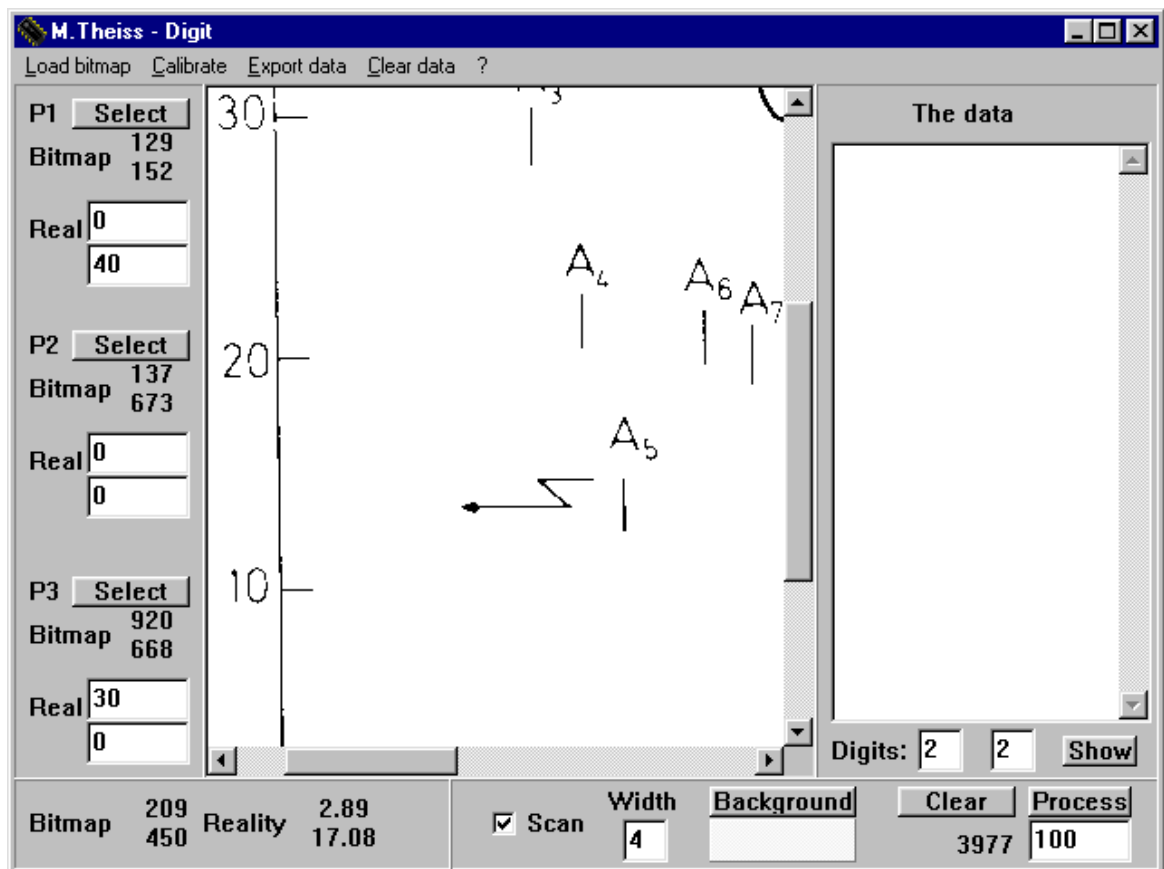

If you have done something wrong (e.g. deleted too many points) you have to reload the bitmap and click on the **Clear** button to delete the list of digitized pixels. If you load exactly the same bitmap you do not have to do the calibration again.

3.4   Step 4

It would not be very useful to export all the 3977 data points. With the **Process** button you can activate an averaging and interpolating procedure which constructs as many output points as you have specified below the **Process** button (in this case 100).

The output points cover the x-range of the digitized points. They are equi-distant.
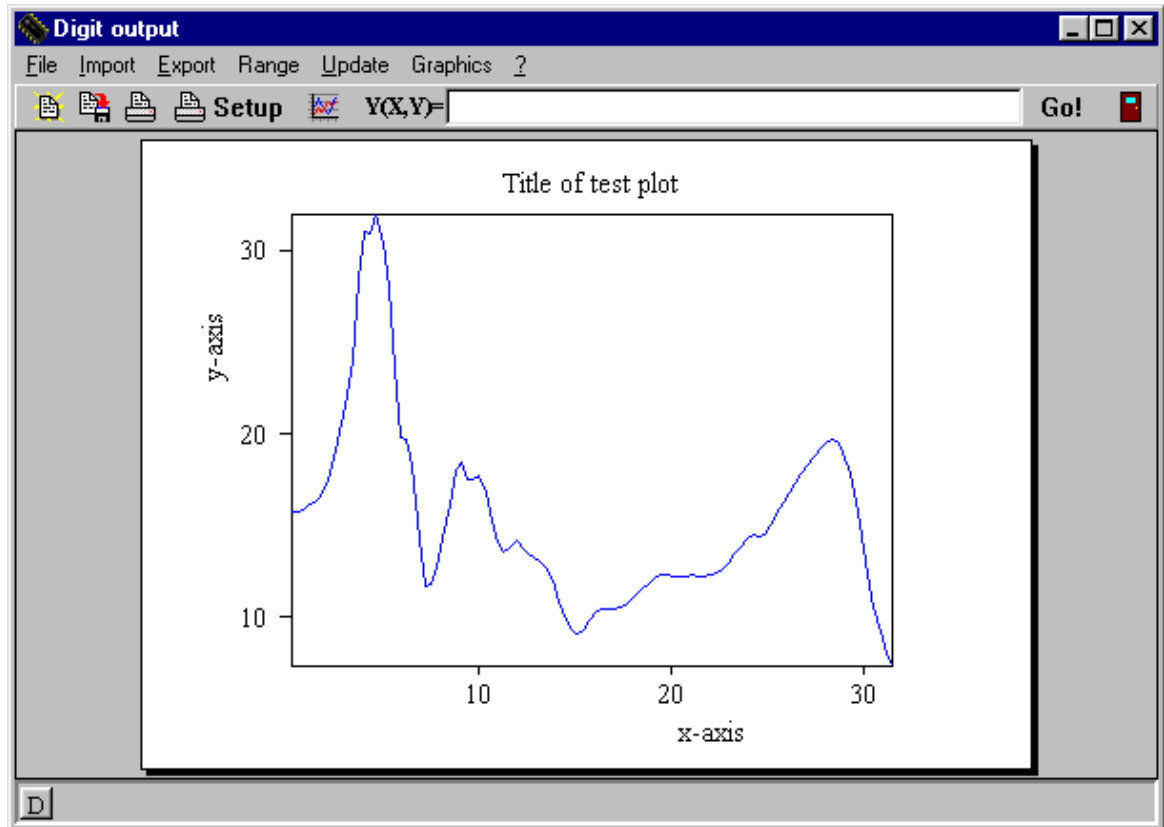
With the present settings one gets, after pushing the **Process** button, the following results:
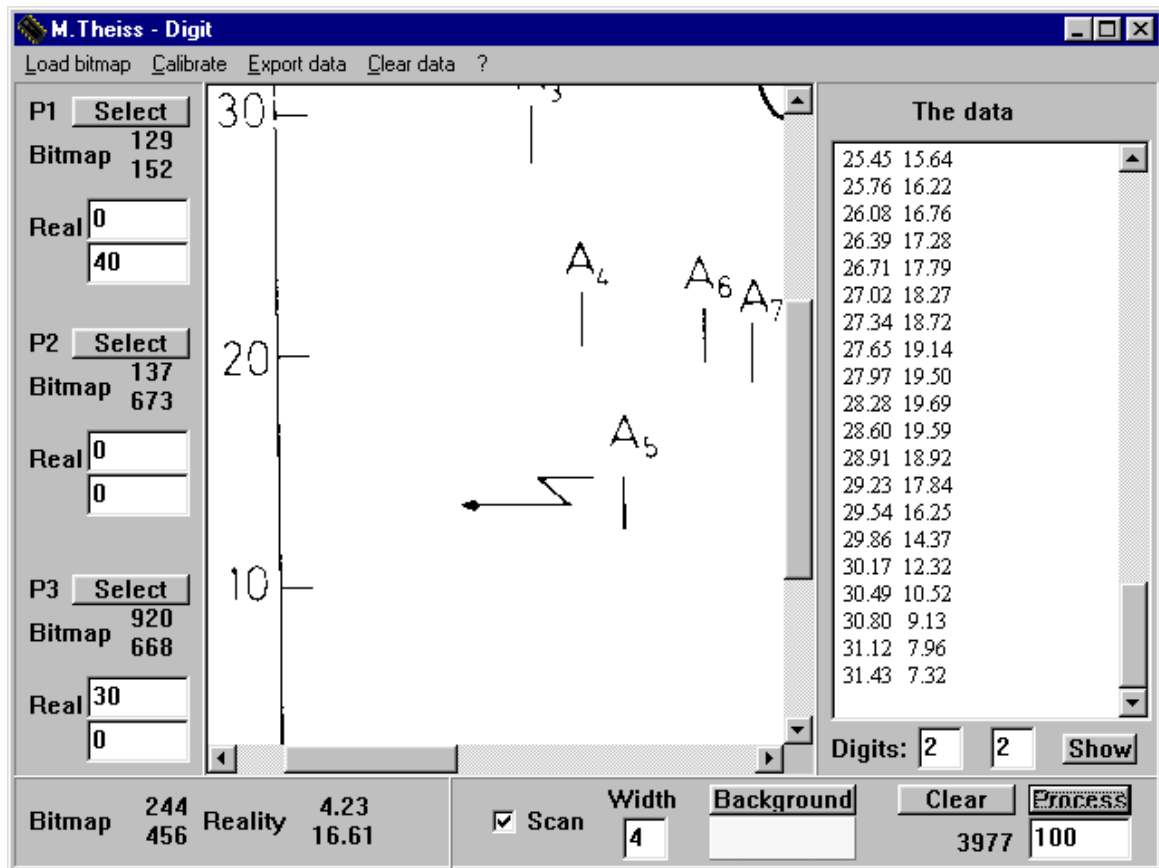
With the menu command **Export data** you can write these values to a text file of your choice. Or go on with step 5.
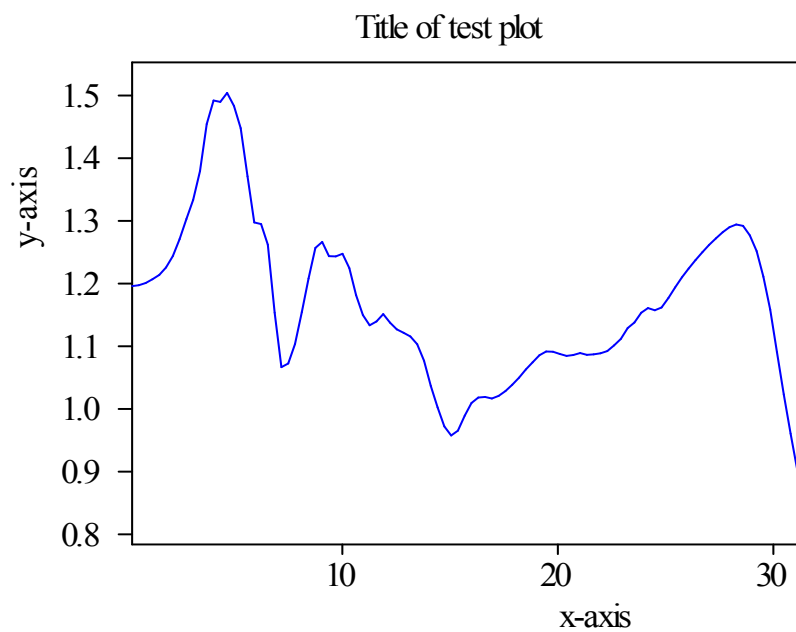
3.5   Step 5

To examine your results you can also get a graphical view of the digitized data. Press the **Show** button and a second window opens showing the data:

Besides looking at the data, you can use this window also for data export in several formats and also for data manipulation.

In the line following **Y(X,Y)=** you can type in a formula (refer to the present y-values using the symbol 'Y' in the formula). Pressing the **Go!** button is followed by a re-computation of the data based on the formula. For example, the formula string 'log(y)' leads to the calculation of the logarithm (base 10) of the y-data:

Title of test plot

The formula editor as well as the handling of the graphics in this window is well-known to experienced W.Theiss Hard- and Software customers. The formula editor is described in the help of the program Data Factory that is shipped with Digit to registered users, the graphics handling is described in the helpfile graphics.hlp (also shipped with Digit).

Using the 'Export' submenu you can export the data to a file or to the clipboard. Alternatively you can write the data to the built-in workbook (using the Workbook submenu). Once the data are in the workbook you can copy&paste data blocks to other applications or write the whole workbook to an Excel file.

Digit

Part IV

4   Digitizing optical constants

Getting optical constants from literature data means to get 2 sets of data - either n and k (which are the real and imaginary part of the complex refractive index) or real and imaginary part of the dielectric function. In both cases the following actions are to be done.

Real part
First digitize the real part as described above. When you are ready and the data look satisfying, click the button **Data to field 1**. This copies the data to a separate data field called 'Field 1' where we will pick up the data later on. Verify that the data have correctly arrived by clicking on the button **Field 1** which opens a separate window. Pressing the key 'a' performs automatic scaling of the data so that we can see what we have.

Imaginary part
Do the procedure that has been described for the real part for the imaginary part as well. However, now use the commands for 'Field 2' instead - after the digitizing press the button **Data to field 2** and to inspect the data press **Field 2**.

Joining the data fields
Finally we have to join the digitized data sets for real and imaginary part.
If you have digitized data of a graph showing the complex refractive index, click the button **Import as n and k**. You will be asked to enter the wanted spectral range and the wanted number of data points, and finally a joined data set is generated and displayed in a separate window. This optical constant window can represent the data either as dielectric function or refractive index. Use the submenu **Property** to select which representation you want to see. You can save the data to a data file using the **Export** submenu, or to the workbook using the **Workbook** submenu.
In case of a digitized dielectric function click on **Import as dielectric func.** in order to join the 2 data fields.

Using the digitized values in SCOUT or CODE
If you want to do computations in our thin film analysis and design programs SCOUT and CODE, you have to generate a new object of type 'Imported dielectric function' in the list of materials. The new object then has to import the new digitized data, either from a generated data file or a workbook data block entered by copy&paste.