# Estimation of Digitizing Error in Program InpGraph 

(S. Taova, G. Pikulina, S. Dunaeva, S. Abramovich, CP-F/010, 2014-04-07)

It is proposed to redefine digitizing error obtained in the program InpGraph.
In the previous version of InpGraph the digitizing error ( $\delta$ ) was defined by a value of standard deviation (D_std) of the digitized value and true value on tics.

$$
\delta_{x}=\sqrt{\frac{\sum_{i=1}^{n x}\left(\bar{x}_{i}-x_{i}\right)^{2}}{n-1}} ; \quad \delta_{y}=\sqrt{\frac{\sum_{i=1}^{n y}\left(\bar{y}_{i}-y_{i}\right)^{2}}{n-1}} \text {, }
$$

where $x_{i}$ - value for the $\mathrm{i}-\mathrm{th}$ tics of x axis $(1<=\mathrm{i}<=\mathrm{nx})$ received from the digitization tool;
$\bar{x}_{i}-$ corresponding true value printed on the tics of x axis;
$\delta_{x}-\quad$ value of standard deviation on x axis;
$\bar{y}_{i}$ - value for the i-th tics of y axis ( $1<=\mathrm{i}<=$ ny $)$ received from the digitization tool;
$y_{i}$ - corresponding true value printed on the tics of y axis;
$\delta_{y}$ - value of standard deviation on y axis;
$n$ - number of tics on each axis.

Another possible approach is to define the digitizing error based on the quantization effect.

At present (due to development of electronic technique) the quality of plots to be digitized is rather high. Displacements of the digitized values and true values on tics are very small and standard deviation tends to zero.

It is obvious that such error (D_std) does not always explain the real deviation of the digitized value and true value on tics.

We propose to take into account additionally an error of quantization (D_qnt).
Quantization means a replace of continuous set of values with a set of quantized values (continuous image is replaced with a set of discrete cells (pixels)).

Quantization error arises when we obtain a bitmap image from the original one (i.e. when we scan or convert a picture from PDF format).

When transferring from continuous image to the quantized one, some distortion of original picture can take place. This fact may lead to some problems (for example, there may be difficulties in searching the geometrical center of a figure (Fig. 1)).

Quantization error $\Delta$ is determined by resolution of a scanner or display (in a case of PDF format). Quantization error cannot be smaller than one pixel.


Fig. 1
Knowing the coordinates of tics on OX and OY axes in a system of display and in a system of physical process it is possible to estimate the value of each pixel (in absolute and relative units, Fig. 2).


Fig. 2
For the plot presented in Fig. 2 quantization error is estimated in the following way:
$\Delta=\frac{E_{2}-E_{1}}{N}=\frac{15-14}{17}=0.0588 \mathrm{MeV}$,
where
$E_{1}$ and $E_{2}$ - true values of the neighbor tics;
$N$ - number of pixels between these tics.
Data treatment at image digitizing has shown that in some cases the value of quantization error is high enough.

For example, digitization of a high accuracy image gives the following results:
X -axis digitizing error:
o $\delta$ - 0.068 KEV ;
○ $\Delta$ - 0.14 KEV ;
Y -axis digitizing error:
o $\delta-0.58 \mathrm{MB}$;
o $\Delta-1.3 \mathrm{MB}$.

At present we have no clear understanding about the relation between $\delta$ and $\Delta$. We are going to continue investigation of this question to be on the safe side. But it seems that we cannot ignore quantization error any more, as it value is high.

One of the possible solutions in this situation is to define digitizing error as a maximum value between $\delta$ and $\Delta$.

Digitizing error $=\operatorname{Max}(\delta, \Delta)$
We would like to propose such approach for definition of digitizing error in the program InpGraph.

## Comment by N. Otsuka on Memo CP-F/010 after discussion with M. Aikawa

If the probability to find the symbol is equally distributed in the pixel, the error due to quantization averaged over all digitized points is $\sim \Delta / 2$. Does it explain the difference between $\delta$ and $\Delta$ seen in the above example?

