Linearity in silicon light-sensors: A case study

Alan Roberts, June 2014 version 2

Silicon light-sensors are used everywhere, in cameras (stills, astronomy, movies, mobile phones etc), in bar-code readers, in spectro-radiometry, production-line automation and in many other imaging applications. Their linearity is often taken for granted, but that doesn't appear to be the case. The non-linearity is not a problem in most usages, where a non-linear response can be a benefit, but in spectroradiometry, where actual light levels must be measured precisely, it causes great problems. There are many possible causes of non-linearity, but investigations by Downing et al[1] showed that only one cause is predominant, the quantum-efficiency of silicon itself. They found that the allied circuitry (analogue amplifiers, power supplies, ADCs etc) are not at fault, so the only possible remaining cause *must* be the quantum-efficiency, the conversion efficiency of photons-to-electrons. In practical terms, this can be explained as an increasing difficulty in getting electrons to fit into the charge well as it progressively fills with charge, electrons are being repelled rather than accumulated. Bezawada et al[2] did similar experiments and concluded that the non-linearity is due to small changes in the capacitance of the charge well as charge increases. However, it seems impossible to separate these possible causes, but the result is the same anyway.

Downing et al and Bezawada et al did their experiments on CCD array sensors used in optical telescopy, but the same science must hold for all uses of silicon sensors. This report gives details of an experiment to investigate and correct the non-linearity in a low-cost spectro-radiometer, the ASEQ LR1.

1. Basics

In most spectro-radiometric software, a standard procedure is followed: expose it to the light source, adjust the exposure duration to achieve a sensible signal level in 'Scope mode', then cap the entry port to measure the $Dark$ current, then make the measurement. This can be expressed mathematically as:

$$Result_\lambda = \frac{Scope_\lambda - Dark_\lambda}{Calibration_\lambda} Reference_\lambda$$

… where $Calibration$ is a measurement of a known $Reference$ source and λ is the wavelength, usually over at least 380 to 760nm, the range of visible wavelengths. Note that the Calibration data itself comes from a measurement:

$$Calibration_\lambda = ScopeCal_\lambda - DarkCal_\lambda$$

[1] 'CCD Riddle: a) signal vs time: linear; b) signal vs variance: non-linear': Mark Downing, Dietrich Baade, Peter Sinclaire, Sebastian Deiries, Fabrice Christen. Society of Photo-Optical Instrument Engineers 2006

[2] 'Conversion gain non-linearity and its correction in hybridised near infrared detectors': Nagaraja Bezawada, Derek Ives, David Atkinson. UK Astrononomy Centre, Royal Observatory, Edinburgh.

However, since it must now be assumed that the opto-electronic transfer function of the spectro-radiometer is non-linear, the equation must now be expressed as:

$$Result_\lambda = \frac{f(Scope_\lambda) - f(Dark_\lambda)}{f(Calibration_\lambda)} Reference_\lambda$$

… where $f(v)$ is a correction for the non-linear transfer function, as yet unknown. The trick is to know this function, and then to estimate a correction for it.

2. Measurements

A wide-range (189.86nm to 1283.88nm) ASEQ LR1 was set up to make measurements of a small tungsten-filament lamp, at close range. Measurements were made with a variety of exposure intervals (up to 300ms) to obtain a series of *Scope* files. At each exposure setting, a separate *Dark* file was saved. For both types of file, several exposures were averaged together in the LR1 software to reduce the noise levels. Further noise reduction has been won by passing the data through a low-pass filter with a bandwidth of 10nm.

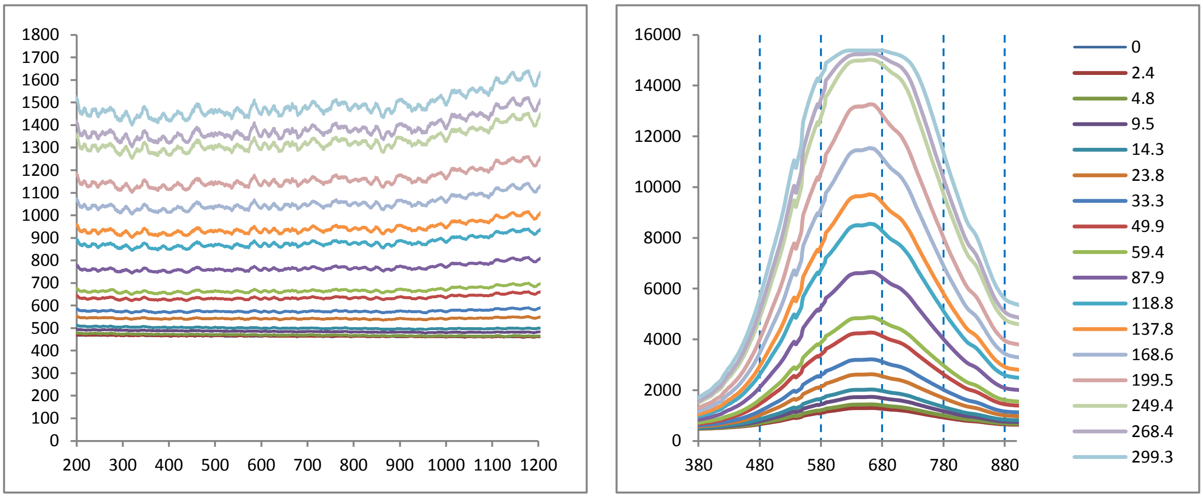

Figure 1. Basic measurements - a) *Dark* signal **b) *Scope* signal**

Figure 1 shows the results, plotted to different scales. Clearly, the *Dark* signal increases significantly with exposure duration (although it retains its basic shape), and the *Scope* signal is clipped at high exposure. But we need to know more than just this.

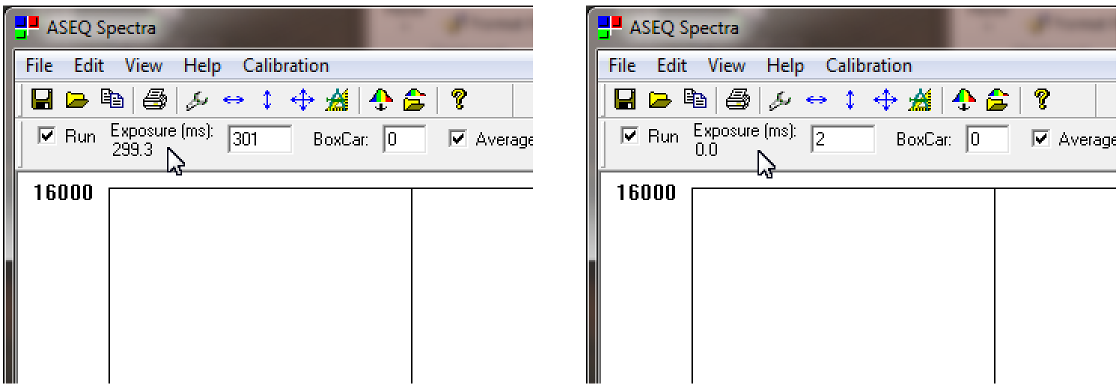

Figure 2. ASEQ LR1 Exposure control - a) 301ms **b) 2ms**

One slightly worrying piece of information comes from this analysis; the indication of exposure time cannot be correct. Figure 2a shows this. In the **Exposure (ms)** edit box, the requested time is 301ms, but the software reports that it is only 299.3. But more significantly, Figure 2b shows that when the requested time is less than 2.4ms, the reported value is 0.0. Clearly this cannot be true since when the exposure is nominally zero, a spectral power distribution can still be captured (the bottom curve in Figure 1b). However, it is quite noticeable that the signal levels change precisely as the indicated exposure changes, in coarse steps.

The reported exposure periods are integer multiples of 2.375ms (rounded to one decimal place), which implies a clock frequency of about 421.053Hz

3. Simple analysis

The simplest possible analysis is to assume that the opto-electronic function is linear for small signal levels, and to find the signal level at which non-linearity starts to become a problem. Then, if all measurements are restricted to below this level as a maximum in Scope mode, then the results should be reliable. This is the normal process advised for spectroradiometry in the TLCI software.

Figure 3 shows the results, *Scope-Dark* plotted as a ratio to the data for 23.8ms nominal exposure. As expected, the line for 23.8ms exposure is flat with a value of exactly unity. The 'frilliness' below

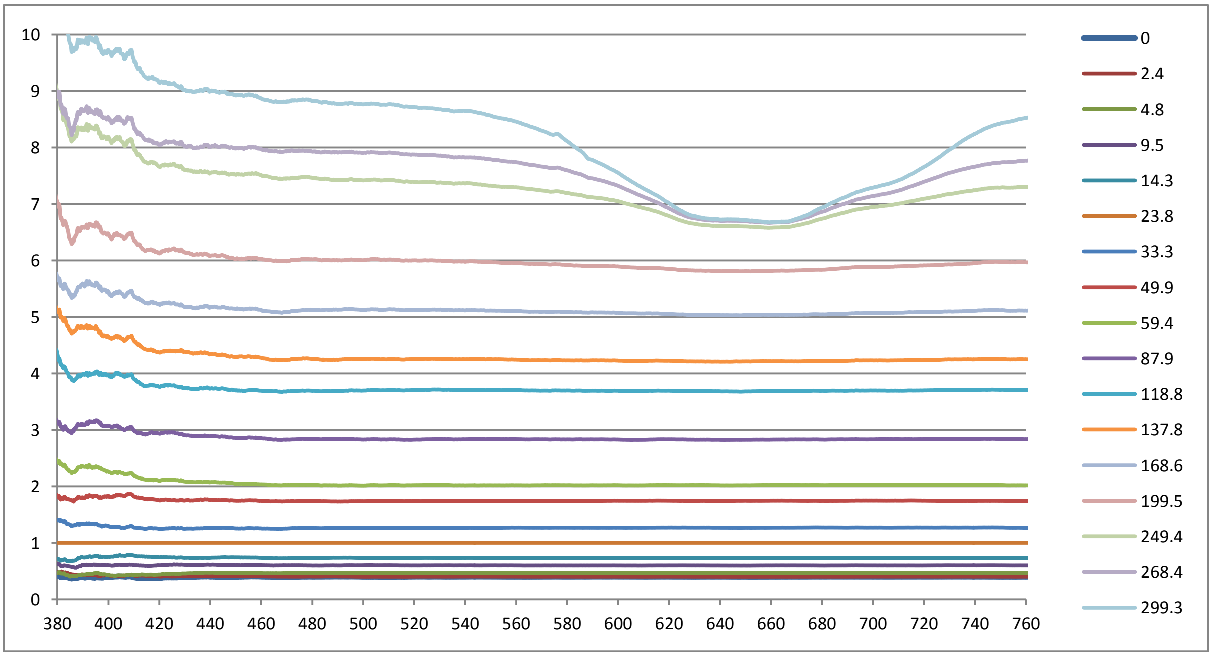

Figure 3 Ratios of *Scope-Dark* signals to that at 23.8ms nominal exposure

450nm comes from taking ratios of small numbers, both of which contain noise, and can be ignored for the time being.

Clearly, the curves lose flatness above 137.8ms exposure, which, referring back to Figure 1 b) tells that the exposure level must be kept such that the *Scope* mode signal level does not go significantly above about 8,500, a little over half the available range of 16,384 (clearly, the software uses a 14-bit ADC). In video and photographic terms, this means that normal exposure should not go above 50%

of the available dynamic range, leaving one photographic stop of overexposure in which the linearity is not important.

However, for spectro-radiometric use, that is not really good enough. It is clear that the non-linearity increases with signal level, but it would be better to characterise it than to try to avoid it. That way, it should be possible to apply a correction for it, and then much of the rest of the available signal range should be useable reliably.

4. More rigorous analysis

It is evident in Figure 1b and Figure 3 that the signal level is reasonably constant between 635 and 660nm, so values in this range can be averaged to provide a better picture of the non-linearity and *Dark* signal. Figure 4 shows how the *Dark* signal level changes with nominal exposure duration.

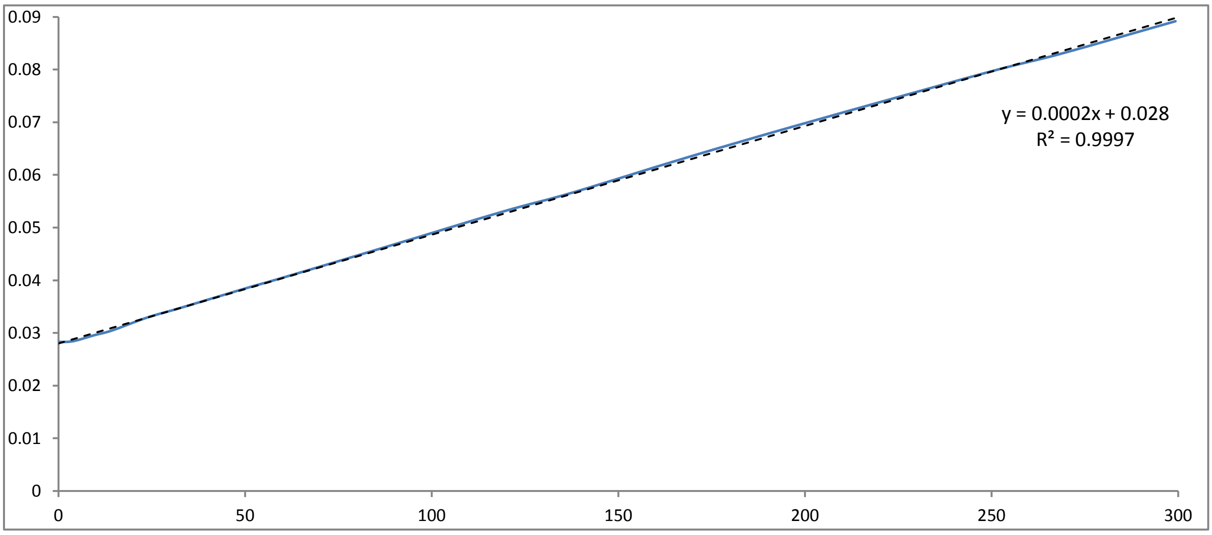

Figure 4 *Dark* signal versus nominal exposure, using averages of signal levels in the range 635 to 660nm

Unsurprisingly, the relationship is linear, the longer the exposure, the greater the accumulated charge due to leakage in the diode structure. There will also be relationship with temperature, but that can be ignored for the purposes of this investigation. The intercept with the vertical axis (which is scaled to the maximum signal level, 16,384, so the Dark signal ranges from 2.8% to 9% in this plot) is probably just an arbitrary offset. The calculated trend-line has a low slope, but there is no need for more than a linear calculation to produce an extremely good fit.

Figure 5 shows how the *Scope* signal changes with nominal exposure duration. We should expect to see a largely-linear function, curving only in the upper 50%, and the difference between the *Scope* plot and that for *Scope-Dark* is only a matter of slope and offset (since the *Dark* signal is linear with nominal exposure). The plot does not include the values for 299.3ms exposure, because it is clear from Figures 1b and 3 that the peak values are either clipped or very nearly clipped.

The trend-line equations need to be 4^{th}-power polynomials to achieve a good fit, and the constant offsets are considerable, but the coefficients of the higher orders are quite small. The offsets are due to the offset in the Dark signal, which is probably arbitrary, and so it seems reasonable to estimate and remove it from the important line, that of *Scope-Dark*. There is no point in doing this for the *Scope* signal alone, since the *Dark* signal must always be subtracted before it can be of any use. A few minutes trying values in Excel established that adding 12.1232448025 to all the nominal

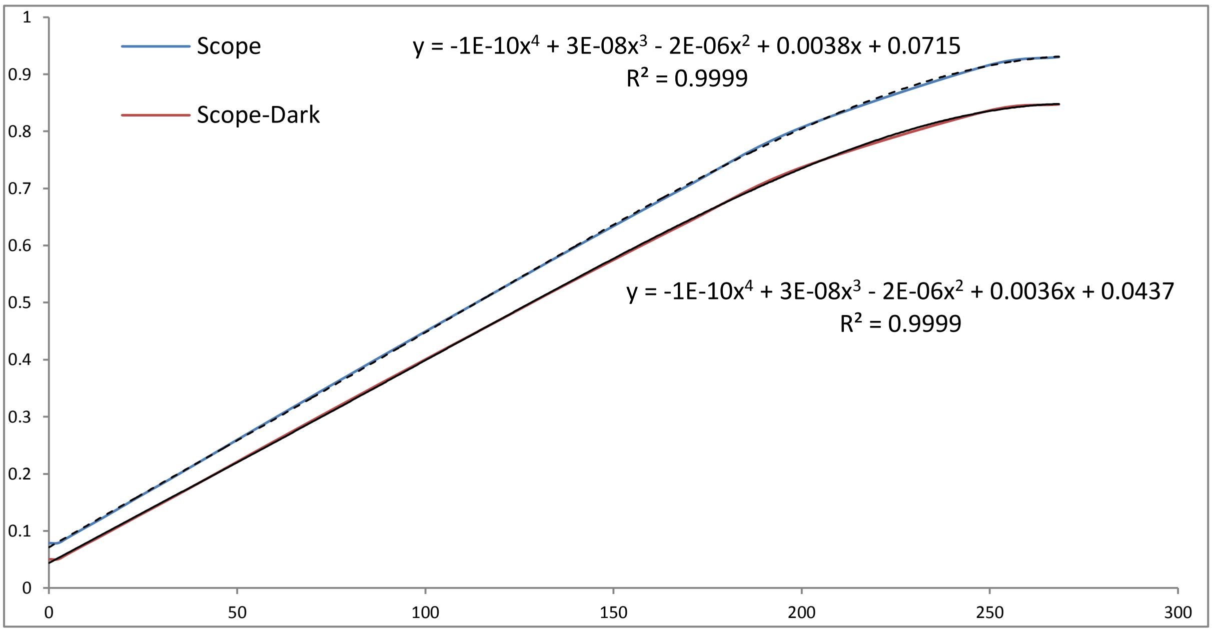

Figure 5 Scope and Scope-Dark vs nominal exposure duration

exposure values reduces the trend-line offset to miniscule proportions. Figure 6 shows the result. The curve is now linear with nominal exposure up to about 65%, and this curve can be used to model the performance of the LR1.

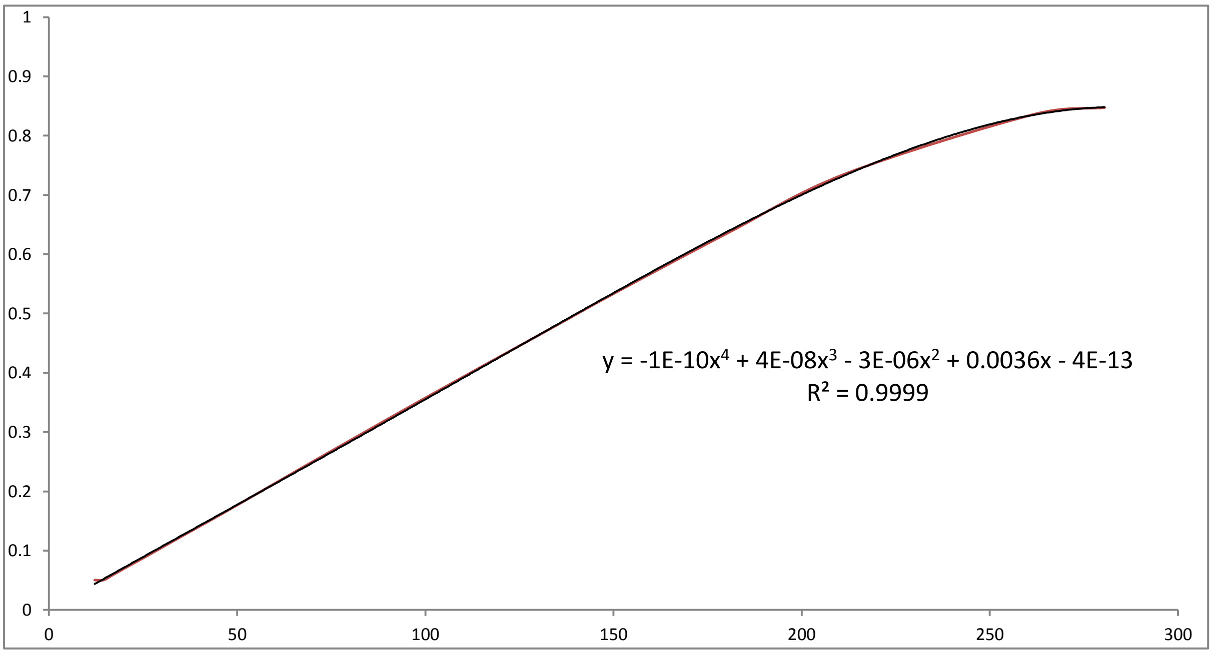

Figure 6 Scope-Dark vs adjusted exposure duration

This trend-line equation can be interpreted as the relationship between the *Scope-Dark* signal and light level, even though an arbitrary constant has been added in order to make it linear through the origin. Therefore, by swapping the axes, the trend-line will be the correction factor needed to linearise the *Scope-Dark* data, and this is what is needed to linearise the LR1.

Figure 7 show this curve plotted with the axes swapped. Since all the coefficients are larger, the offset is more sensitive and so a better estimation can be got, in this case adding 9.2154747391535442 to all the nominal exposure values generates a linear slope below 50% which passes through the origin. The constant value in this equation (-3E-10, or -0.000,000,000,3) can be safely ignored.

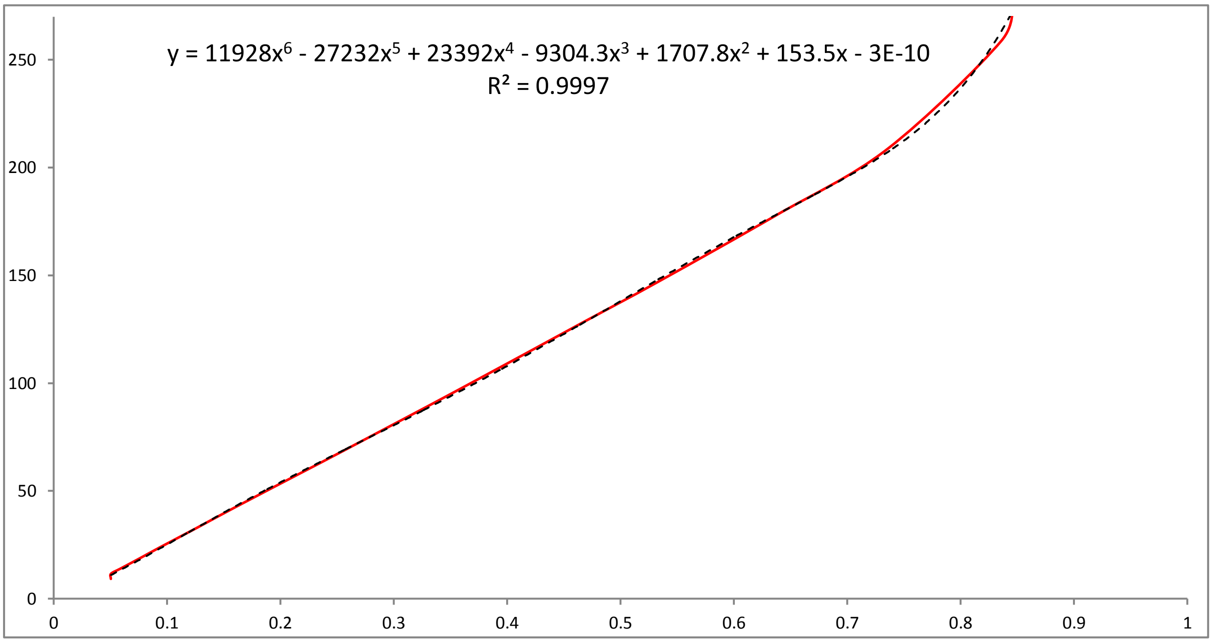

Figure 7 Adjusted exposure vs Scope-Dark

Thus, the data from the LR1 can be made linear using the equation:

$$Light_{\lambda} = 11928x^6 - 27232x^5 + 23392x^4 - 9304.3x^3 + 1707.8x^2 + 153.5x$$

… where x is the scaled value difference $(Scope\text{-}Dark)/16384$ at each wavelength λ.

However, the user must always beware that this applies only when the peak value of the Scope signal is less than 15,400, since that is the peak value of the data set at 268.4ms nominal exposure, and the next exposure measured at 299.3ms produced clipped data. This limiting value is 94% of the available range, 0~16384.

5. Application of the correction expression

Although the expression fits the data well, the data source is suspicious because of the way it was derived. Ideally, the various exposure captures which provide the data should all be made with the same temp[oral exposure, so that the *Dark* signal is constant. Figure 4 shows that the *Dark* signal can be modelled as:

$$Dark = Dark_T + Dark_{DC} + k\ Exposure$$

… where, in this case, $Dark_{DC}$ is the residual value of about 0.028 and k is about 0.0002. $Dark_T$ is the contribution due to temperature T, which we do not know.

Thus, since the various exposure levels were captured with different temporal exposures, the resulting curve must be polluted with at least the linear part of this $Dark$ expression. Therefore the curve will not accurately fit any single measurement (since the $Dark$ offset will, inevitably, be different), it can only be an approximation. However, the fit is still reasonably good provided the signal level is not too high. Visually, this can be explained by looking at Figure 7, where the curve changes directly very abruptly at the top end; this is due to signal clipping which must be avoided at all costs.

Therefore, the measurements must be restricted such that clipping does not occur, even on noise peaks and peaks caused by individual gain differences between pixels in the sensor array. In practice, this means deliberately avoiding signal levels which exceed about 85% after the subtraction of the $Dark$ signal, or 94% before the subtraction.

6. Conclusion

Silicon sensors are non-linear. The source of the non-linearity appears to be a property of silicon itself and not of the electronics. The non-linearity can be measured, and a correction derived, provided precautions are taken to avoid the possibility of signal clipping.