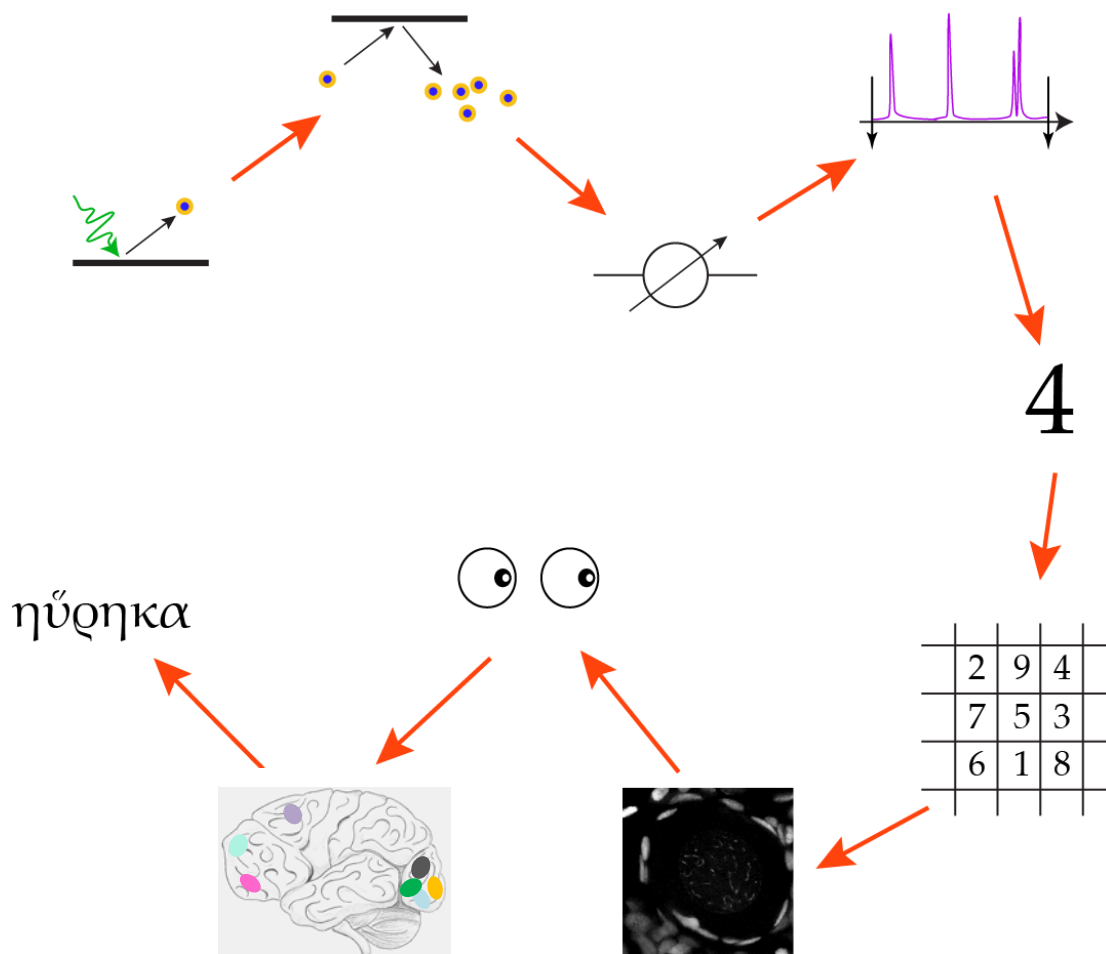


SENSORS AND MEASURING TECHNIQUES IN CONFOCAL MICROSCOPY



*Richtiges Auffassen einer Sache und
Missverstehen der gleichen Sache schließen
einander nicht vollständig aus.*

Franz Kafka: Der Prozess

Introduction

This paper outlines the most important sensors used in confocal microscopy. By confocal microscopy, we mean 'True Confocal Scanning', i.e. the technique that illuminates and measures one single point only. The aim is not to impart in-depth specialist knowledge, but to give the user a small but clear overview of the differences between the various technologies and to advise on which sensor may be most suitable for which applications.

First, we shall give a brief description of the photomultiplier tube (PMT). Having established how light is converted into an electric signal, we shall go on to deal with different methods of primarily picking up this electric signal with the sensor. Basically, these methods apply equally to all sensors, although the suitability of the conversion techniques varies from sensor to sensor.

Finally, we describe the two recent sensor techniques deploying semiconductor technology (at least in part): the avalanche photo diode, APD and the hybrid detector, HyD.

1. The Photomultiplier Tube

This classic technology utilizes the outer photoelectric effect, which we shall first describe in brief. It will then be easy to understand why different materials are used for different purposes. Gain is produced by generating secondary electrons which are then amplified again in several stages. The amplification takes place at the so-called dynodes. Finally, the many electrons that have been created have to be discharged into the outer world. This job is done by the anode.

1.1 Turning Photons into Electrons (Cathode)

The outer photoelectric effect describes the observation that free electrons are generated at (initially metal) surfaces when light shines on them. The first photoelectric effect to be observed was the 'Bequerel effect', where an electric potential can be seen at electrodes in conductive liquids when these electrodes are irradiated with light of different intensity (Bequerel 1839). This effect, similar to the effect seen with semiconductors, is usually called the 'photovoltaic effect' nowadays and is the reason for the booming solar panel business on house roofs. Free electrons are not generated in this technology, which is why it is called an 'inner photoelectric effect'. This inner photoelectric effect is important for the avalanche photo diodes described in Chapter 3.

The outer photoelectric effect was observed and described by Heinrich Hertz (1887), who determined a promotive impact of UV light on the generation of his spark gaps ('Tele-Funken'). Wilhelm Hallwachs (1888) continued the studies on light-induced electricity, which explains why the outer photoelectric effect was called Hallwachs effect for a while. The measured phenomena were understood and interpreted by Albert Einstein (1905). By reviving the corpuscular theory of light (Newton, 1704) he was able to explain all the observations that could not be derived from wave theory.

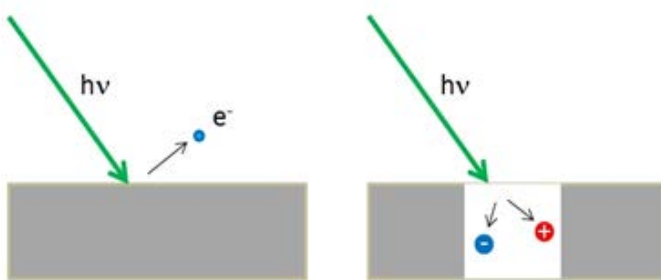


Fig 01: Outer (left) and inner (right) photoelectric effect. In the outer photoelectric effect, an electron is released from the material (such as an alkali metal) after absorption of a light quantum and can then be electrically accelerated (e.g. in a vacuum). The inner photoelectric effect delivers a charge pair (electron/hole) after a photon has been absorbed. These charges can then be accelerated in the semiconductor material, too.

Measuring the number of electrons released upon irradiation of one electrode (i.e. the current that can be detected via the second electrode), we find that this number is dependent on the intensity of the light. This is compatible with wave theory: the greater the field strength, the more energy is available for releasing the electrons. Examining the amount of kinetic energy of these electrons (the 'starting speed'), however, we notice that this speed does not depend on the intensity, but on the wavelength only. Moreover, there is a maximum wavelength beyond which no more electrons are released. This maximum wavelength is a property of the cathode material. This contradicts the idea in wave theory that if we only add enough energy – just a bit longer for weak intensities – we can excite the receiver to release electrons.

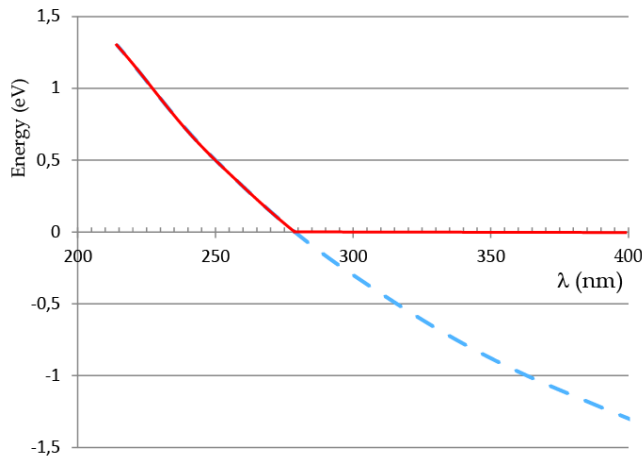


Fig 02: Kinetic energy of the photoelectrons (red) and reduced energy of the photons (blue) as a function of the wavelength. The data apply for a zinc photocathode. If the photon energy is smaller than the work function, no photoelectrons will be released (energy remains zero). If the photon energy is higher than the work function, the photoelectron receives this surplus energy as kinetic energy, and the electron can be accelerated further in a vacuum.

This behavior is explained by understanding photons as particles of light: after Max Planck had already introduced the “quantum of action” in 1900 to be able to describe the correct relationship between temperature and radiation energy, Einstein transferred this quantization to light energy itself. Light is thus divided up into energy packets of the size $E=h*c/\lambda$, where h is Planck’s constant. If a light particle of a certain energy E strikes the metal surface, this energy can be absorbed and can release an electron from the structure of the cathode material. For this to happen, however, the photon must have at least as much energy as that binding the electron to the material. This is the above-mentioned minimum energy described by the work function W_A . Light particles with less energy are not capable of releasing any electrons at all, no matter how many of them hit the surface per time unit (intensity). As c and h are physical constants, the photon energy depends only on the wavelength λ , being inversely proportional, in fact. Above a certain wavelength λ_{min} the energy of the photons is $E=c/\lambda_{min}$, i.e. smaller than the work function, and the measuring instrument does not indicate any current.

Photons with a shorter wavelength can release electrons and we find that their kinetic energy is linearly related to the wavelength: $E_{kin} = h*c/\lambda - W_A$, but independent of the radiation intensity (photon density). All that initially happens at higher intensities is that more free electrons are generated.

This relationship is shown for zinc in Fig. 02. As can be seen from the graph on the right, the light has to be at least ca. 280 nm ‘short’ for a photoelectric effect to be detectable. Zinc cathodes are therefore unsuitable for visible light. The work function of the material used must be as low as possible here. As can be seen from the periodic system, alkali metals are particularly suitable, as their outer electron is only very weakly bound to the atomic core. Therefore, the effective components of photocathodes are normally alkali metals or their mixtures (e.g. Cs, Rb, K, Na). In order to get good results for visible light too, semiconductor crystals are now also often used in the photocathode material (e.g. GaAsP). Many details can be found in the Hamamatsu Photomultiplier Manual (2007). This also applies to all the following sections, particularly for the properties of photomultipliers.

A key parameter for the choice of sensor is the quantum efficiency of the photocathode, i.e. the relationship between arriving photons and generated photoelectrons. Alkali photocathodes perform well between 300 nm and 600 nm, achieving efficiencies of almost 30% in the blue range of the spectrum. Semiconductor variants achieve up to 50% and can be used between 400 nm and 700 nm (GaAsP) or 900 nm (GaAs). Semiconductor photocathodes are therefore being increasingly used for life science applications requiring longer wavelengths, as optical scattering is a major problem with thick specimens that decreases with the fourth power of the wavelength.

For the huge diversity of applications there is therefore a great number of varieties differing in their spectral sensitivity, their quantum efficiency and their temporal resolution.

1.2 Multiplication (Dynodes)

In the photocathode, an electron is released by every arriving and effective photon. Of course, this photoelectron is just as difficult to measure as a single photon. Unlike photons, however, electrons are extremely easy to manipulate: the energy of an electron can be increased by a DC voltage. If the free electron passes through a voltage U before being discharged, it absorbs the energy $e \cdot U$ that is added to the electron as kinetic energy in a vacuum tube. For a potential difference of 1000 volts, for instance, the electron receives 1000 eV (electron volts) of kinetic energy. A photon of 620 nm has about 2 eV. As can be seen, by accelerating the photoelectron through high voltage it is possible to assign far higher energy to the 'one photon' signal than the photon itself contributes: this is a signal gain.

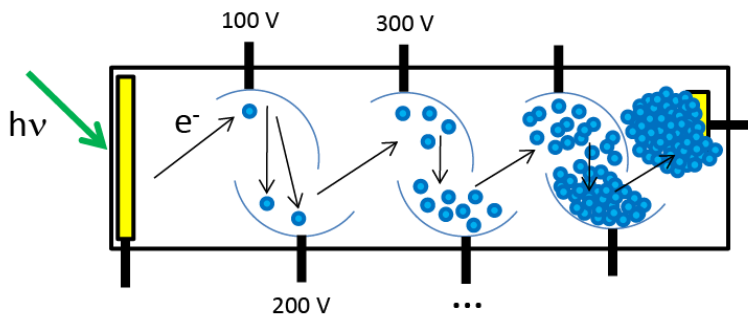


Fig 03: Design and function of a photomultiplier tube (PMT). The photon (green arrow) is absorbed at the photocathode (yellow), triggering the release of a photoelectron there (blue). At the dynodes of the cascaded amplifier (here: 6 dynodes) the kinetic energy is converted into several free electrons (multiplication). At the end of the tube, the resulting charge is measured at an anode (yellow).

In a classic PMT, this gain is distributed among several stages. Behind the photocathode there is an electrode carrying a positive voltage of typically between 50 and 100 volts. The released photoelectron is attracted by this positive potential and accelerates in this direction until it finally hits the electrode. The kinetic energy is here converted into the release of several more electrons ('secondary electrons') – very similar to the process at the photocathode after absorption of a photon. Typical values lie between 2-4 secondary electrons.

To create a measurable charge, this process is now repeated several times in a cascade, so that for 3 secondary electrons per electrode (called dynode here), for example, a total of 3^k electrons are released, where k represents the number of dynodes. Typically, PMTs have about 10 dynodes, which amounts to $3^{10} = 60,000$ electrons – a measurable charge of 10 fC.

The overall gain of the photomultiplier tube can be controlled via the total high voltage, which is normally distributed equally among the dynodes by a voltage divider. The higher the voltage, the greater the gain. The number of released electrons at each dynode is a mean value that can be continuously set through the voltage. This is why rather strange values such as '2.7 released electrons' are encountered here.

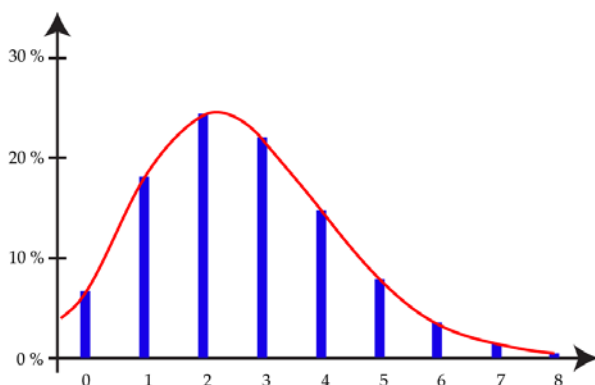


Fig 04: Probability distribution for the release of secondary electrons at the dynodes of a photomultiplier for an average gain of 2.7x. Most often, 2 or 3 electrons are released, although the number can be 8 or even 0. As such events occur consecutively at the cascaded dynodes, the pulses at the end of the chain are not distributed like the integers of the x axis. On the contrary, almost any intermediate values are registered (red curve) that generate a quasi-continuity. However, this curve is not better – it is just the result of extremely high noise.

Fig 4 represents the frequency distribution for the example of 2.7 electrons, describing the number of electrons that are actually released by an electron arriving at a dynode. Most events yield one, two, three or four electrons, and occasionally even more. At the 'late' dynodes, these events are averaged every time an electron shower from the preceding dynode arrives, and a gain of approx. 2.7x is obtained for every pulse. However, an average value cannot be taken at the first dynode, as only one single electron arrives there. The distribution in Fig. 4 is therefore an approximation of the distribution of the pulse heights at the end of the PMT, showing that this value fluctuates by a factor of 4 to 5. Thus the amount of gain varies greatly from photon to photon. This is a significant contribution to noise in integrative measurement methods (see below). These different pulse heights also lead to discrimination problems when counting photons.

It is also quite possible that electrons do not hit the target (the next electrode). Depending on the stage at which this happens, the total charge becomes more inaccurate and the efficiency is diminished. Instead of the photocathode's 'quantum efficiency', the term 'detection efficiency' is therefore often used, which considers losses along the whole chain.

1.3 The Signal (Anode)

After being amplified by the high voltage stages, the electrons finally reach the anode, where the charge can be measured. Because of the different paths the electrons can take in the tube, they do not all arrive at the same time. Here too, the events are distributed, and the temporal distribution is reflected in the width of the ideal electric pulse. In fact, the pulse shape is greatly influenced by the measuring setup and the measured pulse width is a convolution of the distribution of the arriving charges with the attenuation properties of the downstream electronics.

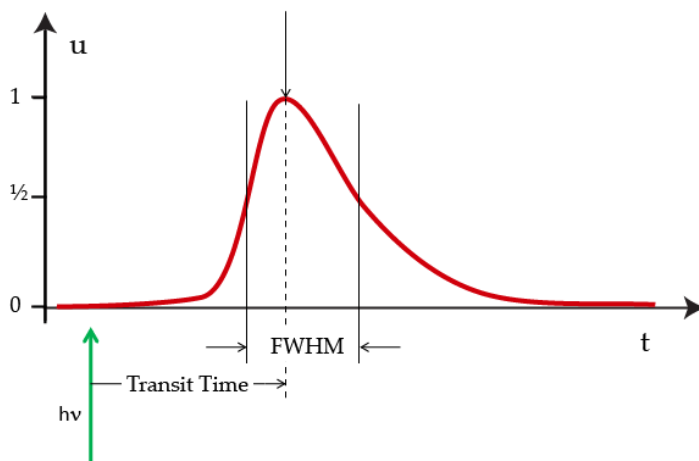


Fig 05: Shape of the charge pulse at the anode of a PMT. After absorbing a photon (green arrow) the charge increases rapidly after a certain time, before decreasing at a slower rate. The time between the arrival of the photon and the peak value of the pulse is called the transit time (TT). The pulse width is given as full width half maximum, FWHM).

To characterize the pulse shape, the full width half maximum (FWHM) is applied. For PMTs, this is roughly between 5 and 25 ns. Of course, the pulse is effective at the anode for longer than the full width half maximum, and for the time between 10% rise and 90% decay we can assume roughly twice the FWHM. However, the shape is not symmetrical; the rise time is two to three times faster than the decay time. To put it colloquially, this can be compared with the drops of water a dog shakes off its coat after having a bathe: first the number of drops increases quickly, and then it gradually decreases – because there are fewer and fewer left.

The time between the arrival of the photon at the cathode and the peak value of the output pulse is the transit time (TT), which is between about 15 and 70 ns, depending on the tube design. It is the time the signal is traveling in the tube. This time is randomly distributed around a mean value, too, and the transit time spread (TTS) is 1 – 10 ns. This is an important parameter for fluorescence lifetime applications, as it limits measurement accuracy.

Even if no photons hit the cathode at all, events will still be measured at the anode at a certain frequency. Most of them are triggered by thermal electrons emitted by the cathode. Thermal electrons from the dynodes can lead to minor events as well. These events are classed as 'dark noise' and diminish contrast, which is a particular problem for weak signals as they can no longer be differentiated from the background. In a PMT of course, the noise of the photocathode and the early dynodes is amplified by the following stages. This is not the case in single-stage systems. A two-stage hybrid detector (see Chapter 4) therefore generates much less multiplicative noise.

2. Measurement Methods and Histograms

The charge cloud arriving at the anode produces a voltage pulse. More precisely, the applied positive potential of the high voltage is slightly attenuated in pulse form by the arriving negative charge. The charge is now measurable, but still very small. Careful design is therefore necessary to ensure that a significant measurement result is obtained. Basically, there are three different measurement concepts, which are outlined below. They are valid for all sensors, including the avalanche photodiodes and hybrid detectors described in Chapters 3 and 4. However, the methods are not equally well suited to all sensors.

We would also like to touch upon the subject of evaluation here, as users are not so much interested in how the sensor works as in the significance of these signals for their measurements and research.

2.1 The Charge Amplifier

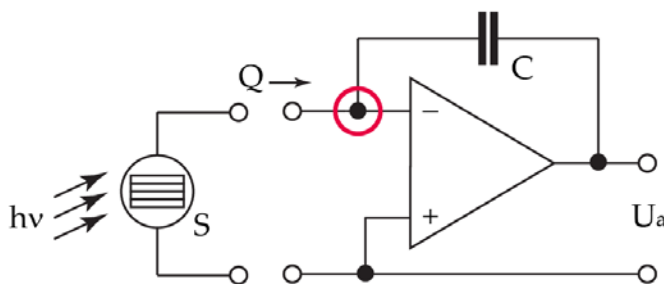


Fig 6: Upon absorption of a quantum of light $h\nu$ the sensor S generates a charge Q . The charge is summed up by an operational amplifier. The OP ensures that the voltage at its inverting input (-) is always equal to the voltage at the non-inverting input (+). So when a charge arrives from the anode of the PMT, it is immediately compensated (red circle) by charging the capacitor C . The capacitor then has a charge corresponding to the charge that has arrived during the complete measurement time and the output voltage U_a is then proportional to C and can be measured. At the end of one pixel the system is reset to zero.

The conventional technique for measuring the PMT signal uses a so-called charge amplifier. In a scanning microscope, the illumination beam is constantly moved over the sample and every line has to be divided into the desired number of pixels. A typical recording for classic structural images takes about a second for 1000 lines per image and 1000 pixels per line. The recording time for one pixel thus takes roughly 1 microsecond. In this microsecond, all charges coming from the PMT are summed up in a capacitor as the graph in Fig. 7 shows. A fraction of the total time is lost for resetting the charge amplifier, which has to start at zero again for the next pixel. The result is a voltage that is proportional to the total charge received from the PMT. This voltage is then converted into "gray values" by an analog-digital converter (ADC). Normally, 8-bit resolution is used, i.e. the gray values are integers between 0 and 255. By adjusting the high voltage at the PMT, it is possible to make sure the measured signal "fits" into these 8 bits, i.e. the entire dynamic range is exploited without the signal being cut off due to overload.

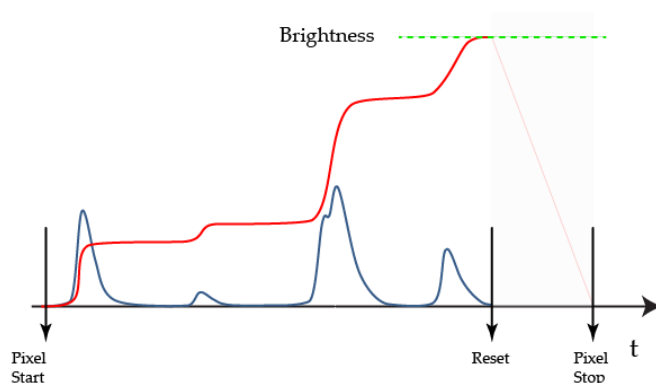


Fig 07: Charge pulses (blue) and integrated signal (red) over the time of a pixel, using a charge amplifier for measurement. The brightness reading is the final integration value. The capacitor is then discharged in a reset phase and measurement can start again for the next pixel.

Obviously, these gray values do not indicate absolute brightness values, but just enable comparison of relative brightnesses in different image areas – or of the brightness of images taken at identical settings of the measurement technique. As described above, the signal height for each arriving photon varies considerably. During the time taken to measure the brightness in one image element, several and sometimes many photons arrive during the measuring time. The possible values are all the combinations of the charges for a single photon, which are quasi-continuously different, anyway. Therefore, the photon events for a cumulative measurement are completely blurred and there is a continuum of signal heights. The digitization process then turns this into 256 separate brightnesses, the gray values.

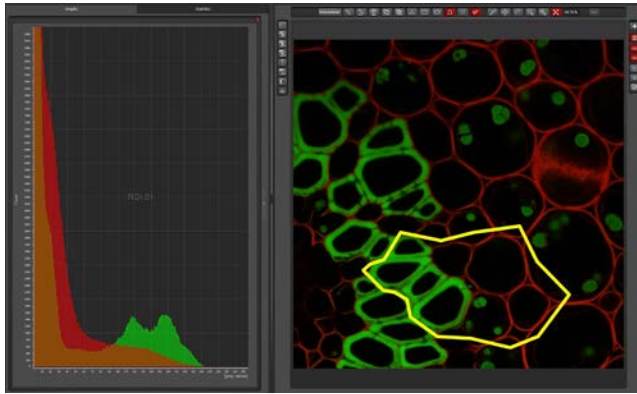


Fig 08: Two channel image of a plant rhizome section. The region of interest is indicated by the yellow line. Most of the area is black background, which appears as very high peaks in the histogram in the left range at low gray values. To make the histogram of the actual signal visible, the count numbers are restricted to show non-background results. The signal of the red channel is the shoulder in the red histogram (between 80 and 200 gray values). The green channels shows brighter and darker areas, represented by the two peaks at about 120 and 160 gray values.

Of course, the experimenter is not usually interested in the brightness of discrete image elements, but in the intensity of a structure, such as a whole cell or tissue component. He or she therefore selects a region of interest (ROI) and calls up its brightness. Depending on the image resolution, which is mostly around 1000 x 1000 pixels, i.e. 1 megapixel, many hundreds or even thousands of image elements are combined in a single number. This mean value can then be compared with values from other structures, or be plotted as a temporal change in an experiment with living material, etc. It is also possible, for example, to analyze the brightness distribution in this area more precisely. To do this, a histogram is generated, counting the number of image elements that have a certain brightness. As there are 256 different brightnesses, the numbers from 0 to 255 are to be found on the x axis of an 8-bit histogram. Because of the digitization there are no intermediate values here. So strictly speaking there should not be a continuous line in a histogram. Also, the frequencies plotted on the y axis are only integers, which is logical as they represent the numbers of the image elements. A histogram of this type is shown in Fig. 8. All sorts of details can be gleaned from this histogram; we are interested in three properties here.

If the brightness distribution is such that the curve is symmetrical, the mean brightness value is to be found at the histogram maximum. However, we frequently find a certain lopsidedness, indicating that the mean value is slightly to the left or right of the maximum.

For extreme brightness fluctuations, the curve becomes very 'wide'. If there was no variance at all, there would only be one gray value, and the 'curve' would then be infinitely narrow. The width is therefore an indication of the variance. But it is not only the distribution of the number of photons arriving in an image element for each measurement time that influences this variance, but also the noise properties of the measuring system, here in particular the PMT, and also the electronics. Only when these noise portions are small enough is it possible to estimate the number of contributing photons derived from the ratio of the mean value to the width.

As the zero point is fixed by the setting of the electronics, it is not possible to deduce from a gray value histogram the proportion of PMT dark current in the background brightness.

For a usable image, a signal-to-noise ratio (SNR) of about 3 is needed. This is the case when 10 photons are registered in an image element. For a 1 megapixel image with a line frequency of 1 KHz, this is 10 million photons per second. As a rule, however, we try to manage with less light, as high illumination intensities destroy the fluorochromes, and the resulting fractions are often toxic chemicals for the living material. Therefore, the illumination is reduced until a good compromise is reached between image quality and sample stability. So frequently only an average of 1...3 photons and less is registered per pixel. Signals that are extremely weak but can still be evaluated have a mean value of significantly less than one photon per pixel.

Why do we get far more than 200 different gray values if we only have 10 photons? Wouldn't 15 different brightness levels be enough? This would indeed be the case if each photon generated the same signal in the image. But, as we established above, the pulse heights vary tremendously (the intensity is noisy). If in a pixel the pulses from several photons are summed up, the possible number of brightness values is again multiplied. This is the reason for the large number of brightness levels. And the more levels we use (lower gray resolution with about 12 or 16 bits) the higher we rate the noise generated by the pulse variation. However, the result does not yield better information.

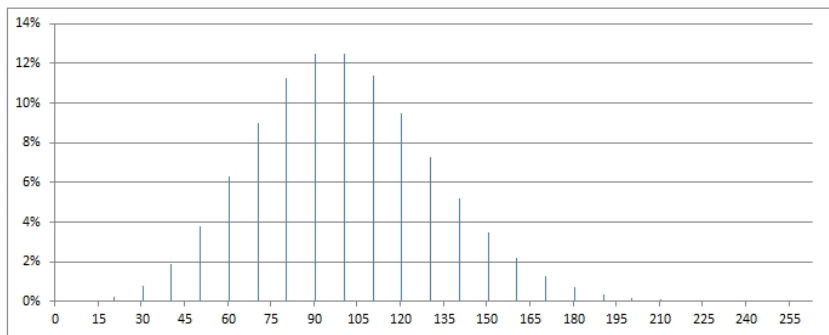
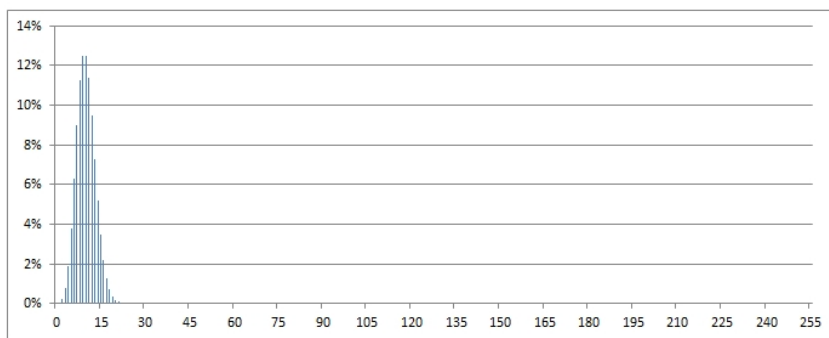


Fig 09 top: If only a few gray levels are used in an image, the histogram only shows values in a narrow range. If these gray levels are then multiplied so that the 8-bit dynamic range is fully utilized (e.g. to make the monitor image easier to view) the number of gray levels is retained and intermediate levels remain empty, leading to "gaps" in the histogram.

The histograms also supply further useful information. If, instead of exploiting the entire dynamic range of 0 to 255 during the recording, only the bottom tenth from 0 to 25 is used, the monitor image is extremely dark, as the monitor is only using the gray values from 0 to 25. We can only make the image brighter by 'widening' the histogram, a common technique automatically offered by all digital cameras today. This is done by multiplying the brightness in each pixel by the same number Z for all pixels. In the case above, we would multiply all values by 10 to cover the range from 0 to 255. However, as there were originally only 25 x values, there are only 25 x values in the image result. So the histogram has gaps; in our example 9 values are missing between all the plotted gray values. The graphs therefore look rather 'untidy', but are quite correct.

Although the charge amplifier was used for many years as a device for measuring PMT signals on confocal microscopes, it has now been mostly replaced by other techniques (see below). The charge amplifier delivers signals that are summed up over a whole pixel, so the result depends on the length of the pixel time. However, this pixel time changes if, for example, the scan format is changed, i.e. if only 100 instead of 1000 pixels per line are registered at the same scanning speed. The signal is then 10 times greater and necessitates adjustment of the high voltage on the PMT. Changing the scanning speed has the same effect. In addition, a small part of the measuring time is lost in each pixel because the resetting of the charge amplifier takes a certain amount of time. At high scanning speeds and high resolution the pixels are extremely short, and the reset loss becomes noticeable. This can be remedied by direct digitization.

2.2 Direct Digitization

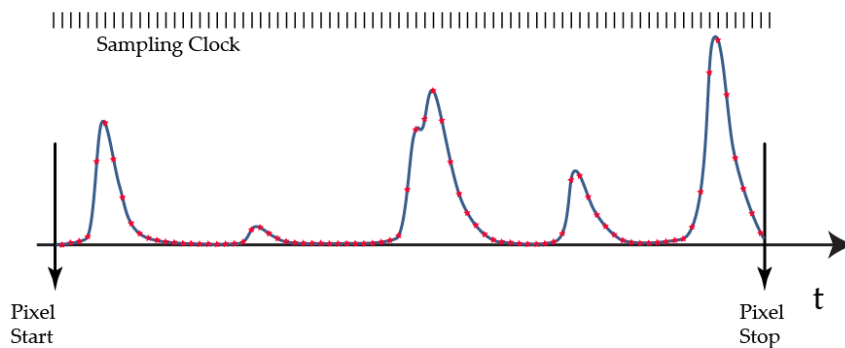


Fig 10: Same photon events as in Fig 7). As familiar from hifi technology, the signals at the output of the PMT are directly digitized at a high frequency (red asterisks). The digitization frequency (sampling clock) is set high enough to be faster than the electronic bandwidth of the measurement system ('oversampling'). This ensures that the signal is completely captured.

Modern digitization circuits are fast enough to convert the PMT signal directly into digital units. This is done by discharging the PMT charge via a suitably dimensioned resistor and converting the resulting voltage into gray values at a high clock rate. The data of a single image element is immediately averaged while it is being recorded, and the image is consequently given gray values that genuinely represent the intensity of the sample, however long it takes to record an image element. The high voltage at the PMT does not need readjusting and the brightnesses in particular can be compared directly. There is no time lost for resetting, either.

The histograms look exactly the same for direct digitization as when a charge amplifier is used and basically contain the same information. After all, the averaging method basically adds up the measurement data as well, but scales them immediately on the correct time axis. The scan zoom does not change the image brightness here either, except for the real changes in brightness caused, for example, by different degrees of fluorochrome bleaching – but that is valuable information, not a measuring artifact. Bleaching can be immediately quantified here, too.

A further advantage of this technique is that there is no dead time spent emptying a memory. This results in a better signal-to-noise ratio for extremely short pixel times.

2.3 Photon Counting

We now have splendid clear and simple histograms to describe intensities. Interestingly, however, the world is far from being so smooth and continuous. As Max Planck (1900) involuntarily discovered, light does not exist in a continuously intense form. This is why with our detectors we measure single events, the arrival of photons, rather than continuous intensities. The apparent continuum is the result of metrological vagueness: the histograms of analog measurements with a PMT show the merging of blurred measurements. As we already mentioned, a photon at the PMT generates a pulse of varying width that first rises in intensity and then falls. By integrating or averaging a large number of pulses of different widths and heights, an apparently continuous brightness histogram is obtained in which any intensity values can occur. What's more, the number of these intensity values depends on the arbitrary 'gray depth' setting for data recording. In reality, however, this is by no means the case. On closer consideration, the brightness in a pixel cannot be represented by a rational number. It can only be an integer, in fact it can only be the exact number of photons that arrived at the detector during the measuring time in this pixel.¹ So rather than integrating or averaging the charge (i.e. the area under the pulse curve) of each pulse, it would be desirable to count the charge pulses arriving at the anode without evaluating the pulse size. This would solve a major part of the noise problem.

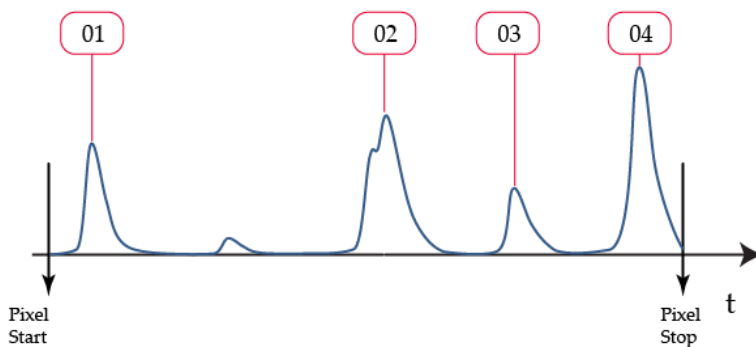


Fig 11: (Same photon events as in Figs. 7 and 10). Counting the charge pulses generated by single photons is called 'Photon Counting'. While this technique generates far lower noise than charge amplification, it can only be applied to pulses of similar height that do not merge with each other. Also, small pulses might be not recognized.

Such counters are a standard in electronics nowadays. The height or a slope of the pulse is used to activate a trigger that increments the counter by 1 each time. At the end of the pixel, the counter is reset to 0 (this takes next to no time) and the counter begins again. However, the limitations of this method are instantly apparent in Fig. 11: the pulses naturally have to be separately discernible. If a second pulse occurs during the first one, only one pulse is detected and the subsequently measured brightness is too low. Whether the pulses are separated or not depends on the interval between the pulses and on the pulse width. Wide pulses allow only dim brightnesses for the photon counting method, as otherwise the pulses follow each other too quickly with the risk of overlay. Narrow pulses can yield good counting results even at higher intensities, i.e. at shorter pulse intervals.

¹ In fact, the light intensity also depends on the color of the photons, and this will also slightly influence the pulse heights. However, this energy difference accounts for only about 2 eV for the entire range. Compared with about 80V at the first dynode, this is negligible. For the hybrid detectors (HyD) discussed below, this variation is even about a hundred times less significant.

The same applies for the initial velocity of the photoelectrons, which, as we mentioned, can leave the cathode in any direction. Some of them fly in the desired direction, but others have to be accelerated by a change of direction that reduces the kinetic energy when they hit the dynode. Here too, the variations are only in the range of ± 2 eV.

A typical PMT delivers pulses of a separable width of about 20 ns. If photons always occurred at the same time intervals, counting rates of up to 50 million per second (50 Mcps) could be realized. As they arrive stochastically, however, overlapping is much more probable, and the highest counting rate is reduced. By the highest counting rate, we mean that the relationship between the number of actually triggered photo electrons and the measured number of pulses is no longer linear above this counting rate. Admittedly, this is an arbitrary definition depending on the error tolerance. If a deviation of only 1% is allowed, the linearity ends at 0.5 Mcps, for 10% at 5 Mcps. An accuracy of one per cent is unattainable in everyday practice anyway, whereas 10 per cent is above the tolerance threshold. The threshold can be set at 6%, for example, to obtain a useful measurement up to counting rates of just over 10 Mcps.

If we assume for a moment that there are no brightness fluctuations (e.g. by illuminating a fixed point, without scanning movement) the photons arrive at the detector on average uniformly, but stochastically. In such a Poisson process, the time intervals between photons can be described by an exponential distribution. We need not go into detail here, suffice it to say that such a process can be described in exact mathematical terms. To a certain extent we can use it to compute the error resulting from pulse overlay that causes the deviation from linearity. If the error is known, however, we can use the measured number of pulses to count back to the actual number of arriving pulses. This works up to about five times the pulse rate at which the error became significant without correction. Therefore, this correction is called linearization. It is a well-known and well described method for extending the measurement range (Hamamatsu, 2007).

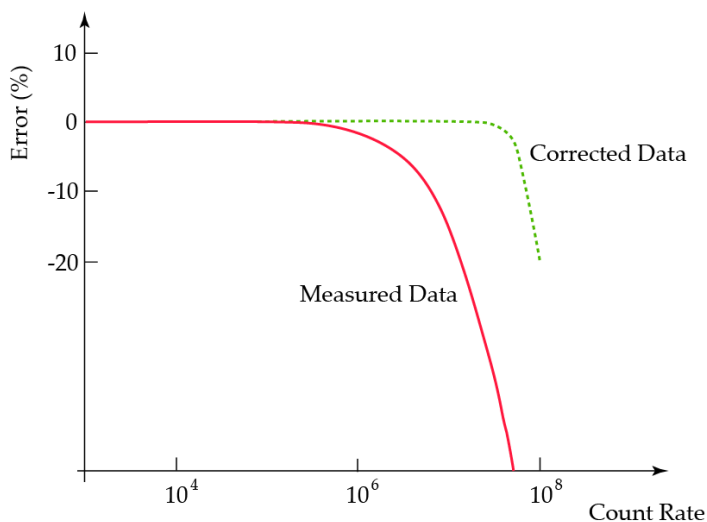


Fig 12: Linearization of the counting rates by correcting the statistic probabilities. Note the logarithmic x axis. (According to Hamamatsu 2007)

Instead of the originally counted pulses (corresponding to the photons), this linearization, just like widening, yields fractions as well as integers in the computation. This in turn impacts the histogram, which naturally plots only integers such as gray values. In any case, the widths of these energy channels are normally always constant, and it is always possible to convert any fractions into integers, e.g. by beginning at 0 and then simply continuing to count. Linearization therefore leads to the occurrence of beat effects in the histograms, so that individual energy values protrude far beyond the envelope of the histogram. Although irritating at first, it is an exact representation of the statistically corrected photon numbers.

3. The Avalanche Photodiode

Of course, light-sensitive components that can be used to measure intensities have meanwhile been produced in semiconductor technology, too. The instrument that can be most easily compared to a photomultiplier is the avalanche photodiode (APD).

3.1 Turning Photons Into Charge Pairs

To measure light with semiconductors, the inner photoelectric effect is used. This entails generating a charge pair (free electron and diffusible “hole”) in a very weakly doped semiconductor material after the absorption of a photon. The absorbing layer is embedded between a p- and an n-doped layer, thereby having an effect like the enlargement of the depletion region that would form at a direct pn junction. The layer is called ‘i’ layer because it only demonstrates intrinsic conductivity. These semiconductor systems are called pin diodes due to the order in which the layers are arranged. If a reverse-biased voltage is applied to the electrodes of such a photodiode, a current will flow that over a wide range is proportional to the intensity of the incident light. However, such a diode is not sensitive enough for weak signals and has too much noise.

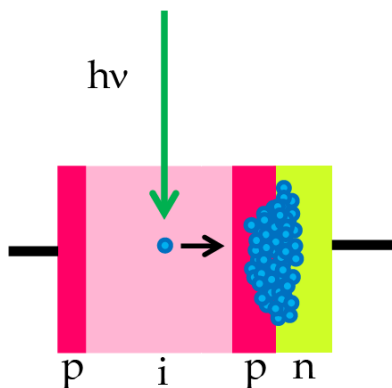


Fig 13: In an avalanche photodiode, an intrinsic (i) layer is sandwiched between the outer p- and n-doped layers. The photons are absorbed here and trigger the production of a charge pair. For the sake of simplicity, only one charge is illustrated here. An applied field moves this charge to the multiplication zone (pn junction on the right). Here, very many new charge pairs are generated that enable a pulse to be measured at the output.

3.2 The Avalanche Effect

To be able to measure even weak signals, an extra layer is added to the pin arrangement: another highly doped p layer is inserted between the i and the n layer. The extremely high field strengths formed here cause the charges to accelerate rapidly and emit their energy as they knock against the grid elements, generating further charges (the reason for the comparison with the PMT). These charges are accelerated, too, causing the process to swell like an avalanche and lead to a pulse in the measurement circuit. If the charges are moderate, the avalanche stops by itself, and a gain of up to 10^2 can be attained. At extremely high reverse voltages, massive gain is achieved (up to 10^8), but the current has to be actively interrupted to avoid destroying the component. This latter case is called “Geiger mode”, as it is suitable for detecting single photons without additional amplification. The disadvantage of the Geiger mode is the long deadtime period after a measured event. Generally, therefore, this mode is not suitable for image recording.

3.3 How It Differs From the PMT

Unlike photomultiplier tubes, avalanche photodiodes only have a small dynamic range, so one always has to make sure that the intensity of the light being measured is not too high. The gain in non-Geiger mode is between 100x and 1000x, depending on the voltage applied. On the other hand, this type of sensor has extremely slight dark current, making it suitable for weak signals that would be lost in background noise if the dark current was too high. The spectral sensitivity covers a wide range of 300 nm to over 1000 nm, which explains why, for red fluorescence emissions in particular, APDs are used for image recording, too. In spite of the extremely high sensitivity in the red range, the thermal noise is still relatively small. This noise depends on the surface area of the sensor, which is 0.1 mm² at most for an APD, whereas the surface area of PMT sensors can be up to 10 mm².

APDs generate very narrow pulses, an important property for photon counting. This can only be utilized in non-Geiger mode, though, as otherwise the advantage is lost due to the very long deadtime periods (several 10s of ns).

4. The Hybrid Detector

It would be good to be able to unite the high dynamics of a PMT with the speed and low noise of an APD. This is exactly what has been achieved with a chimera technology consisting of a vacuum tube and a semiconductor component. The result is a hybrid detector (HyD; also HPD: Hybrid Photo Detector or HPMT: Hybrid Photo Multiplier Tube).

The combination of the two technologies in this component really has resulted in a sensor that combines the best characteristics of a PMT and an APD.

4.1 Design

The 'input' of an HyD is identical to that of a PMT: a photocathode in an evacuated tube can release a photoelectron by absorption of a photon. As a rule, GaAsP cathodes are used – these are also the most suitable cathodes for PMTs used in biomedical fluorescence applications. Unlike the PMT, however, this photoelectron is massively accelerated in one single step rather than in stages, passing through a potential difference of over 8000 volts.

Here already, we can see that the differences in the relative pulse height are much less than for a PMT, for which the acceleration voltage as far as the first dynode is about 100 volts. As described in Chapter 1, the number of electrons released at the first dynode is between 2 and 4, with an expected variance of 1 to 2 electrons. This is why the pulse heights there usually vary by a factor of 3-5. By contrast, the photoelectrons in a hybrid detector are accelerated by 8 kV. From the kinetic energy, approx. 1500 secondary electrons are obtained. There are therefore variations of 40 electrons ($\sqrt{1500}$). Although this is, of course, more than 2-4 electrons, it is only 3% of 1500. The output pulses are therefore extremely homogeneous and a much more accurate representation of the illumination energy.

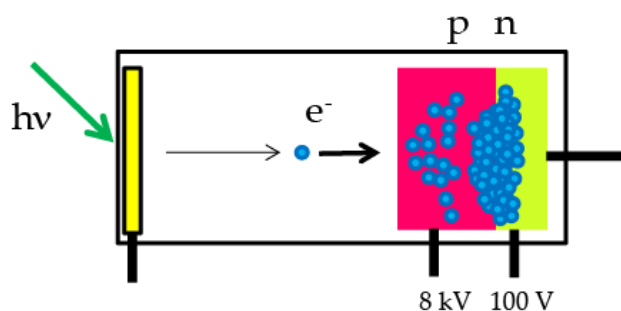


Fig 14: Design of a hybrid detector (HyD). As in a photomultiplier tube, a photon ($h\nu$) is absorbed at a photocathode and a free electron (e^-) is generated. This is accelerated in a vacuum at an extremely high voltage (approx. 8 kV). The energy is converted into charge pairs in the semiconductor material and amplified again in the multiplication layer (pn).

Now, the highly accelerated photoelectron does not hit a dynode (which could not cope with such high energies), but a semiconductor. Here the energy is converted into a great number of charge pairs (approx. 1500, as mentioned above). An applied voltage makes this charge move in the direction of a multiplication layer as described for the APD (avalanche effect). Here, the signal is amplified again by about 100x and thus becomes a measurable pulse.

4.2 Data Recording With a HyD

Due to the single-step acceleration and the subsequent direct amplification, there is much less variance at the trajectories of the charges, which means that the pulses are much sharper than with a PMT: the pulse width is reduced by about 20x and is about 1 ns (currently close to $\frac{1}{2}$ ns).

Furthermore, due to the absence of dynodes and the much smaller size of the photocathode, there are far fewer dark events. Compared to a photomultiplier tube, background noise is also greatly reduced. Besides improving the contrast of the recorded images, this also allows a greater number of images to be averaged or accumulated. As the background is almost totally black, it still remains black when a large number of images are combined into one, which makes it possible to significantly reduce signal noise at high contrast.

The hybrid detector therefore offers myriad advantages. Its homogeneous and narrow pulses allow photon counting with light currents that would soon oversaturate a PMT. It therefore makes sense to use hybrid detectors in photon counting mode for 'ordinary' image recording, too. After all, as described in Chapter 2.3, photon counting has a lot of advantages compared to analog detection techniques.

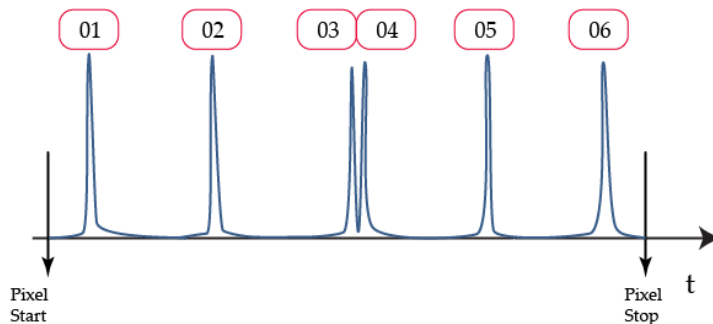


Fig 15: The same course of events as in Figs. 7, 10 and 11, but with a sensor with a far higher bandwidth and constant pulse height as provided by a hybrid detector compared to a normal PMT. The significantly improved pulse resolution allows more closely spaced pulses to be resolved (i.e. higher light intensities can be recorded in photon counting mode). Also, small pulses are not missed.

Consequently, the evaluation does not yield histograms 'smudged' with noise and quasi-continuous histograms, but, as is to be expected, well defined frequencies of only a few photons. Fig. 9 actually shows a histogram of this type. At the top you see the Poisson distribution for a mean value of 10 photons per pixel.

Utilizing hybrid detectors, the Leica TCS SP8 system offers a free choice of three different data recording techniques. First, the raw data of the photon counting can be used as image information ("Photon Counting" mode). As linearization methods are not used here, the maximum counting rate is lower than for linearized data, but still significantly higher than for conventional photomultipliers. A rate of around 60 Mcps is obtained (cf. Chapter 2.3, Photon Counting, for an explanation).

If the counted events are corrected by the mathematical linearization method described above, the maximum counting rate is improved by a factor of 5 and brightnesses of up to 300 Mcps can be recorded. Here, photons can be converted into values that may not necessarily be integers, so it is a good idea to carry out a scaling process right away to make the brightnesses comparable with conventional sensors and suitable for a balanced display on a monitor. This method is applied as standard. Note that the histograms may look 'odd' due to widening and rounding effects. The data measured when evaluating image brightness are still correct, though.

For better perceptibility of the total dynamic range used, the values can also be changed to reduce the effect of bright pixels and increase the effect of dark pixels. This corresponds to the results of HDR rendering. Designated 'Bright R', this mode is used to display images that exhibit great brightness differences so that there is no glare in bright image areas and dark areas remain visible. An example of this would be neurons with a great amount of dye in the cell soma and extremely thin dendrites.

Table 01:

Mode	Scaling	Linearization	Augmented Dynamics
Photon counting	no	no	no
Standard	yes	yes	no
Bright R	yes	no	yes

5. Literature

Becquerel AE, 1839: „Mémoire sur les effets électriques produits sous l'influence des rayons solaires.“ Comptes Rendus, 9, 561-567.

Einstein A, 1905: „Ueber einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt.“ Annalen der Physik. 322, No. 6, 1905, p. 132–148.

Hallwachs WLF, 1888: „Ueber die Electricisierung von Metallplatten durch Bestrahlung mit electricischem Licht.“ In: Annalen der Physik, volume 34, p. 731-734.

Hamamatsu Photonics KK, 2007: „Photomultiplier Tubes – Basics and Applications“, Edition 3a, 310pp.

Hertz HR, 1887: „Ueber den Einfluss des ultravioletten Lichtes auf die electriche Entladung.“ Annalen der Physik 267 (8), p. 983-1000.

Newton I, 1704: „Opticks. Or, A Treatise of the Reflections, Refractions, Inflections and Colours of Light.

Planck M, 1900: „Ueber irreversible Strahlungsvorgänge“ Ann. Phys., 1, 69-122