Assessment of Measurement Uncertainties for a SDS011 low-cost PM sensor from the Electronic Signal Processing Perspective

Bernd Laquai, 4.10.2017

Low-cost PM measuring devices are meanwhile frequently used as indicators for air quality in cities that suffer from high air pollution by particulate matter emitted from traffic, industrial combustion as well as from office and residential heating units. These devices typically make use of cheap but intelligent laser scattering sensor elements that provide PM10 and PM2.5 values in terms of ASCII data, readily prepared for a microcontroller, that serves for displaying data, managing of other sensors like temperature and humidity, an alert system, the power supply and sometimes providing even networking connectivity. Since the data sheet of such a low-cost sensor is rarely providing detailed data with respect to accuracy, the key question is, what measurement uncertainty has to be taken into account when measurement values of such a device have to be judged.

The most realistic check for evaluating the remaining measurement uncertainty is a comparative measurement to a well calibrated reference instrument in real life experiments. However, in case of PM measurements the results are influenced by many parameters such as the particle mass distribution versus size or the physical and chemical composition of the particles or simply the shape. Covering the whole n-dimensional space of parameter variations with realistic experiments, just to assess the measurement uncertainty is a tedious job. Therefore, a more pragmatic approach is often desired to find out what can be expected from such a sensor in terms of accuracy.

A low-cost PM-sensor typically uses the laser scattering technique for measurement. The particles are guided into a measurement chamber with an air stream from a fan where they scatter the light of a laser-diode beam. The signal processing is composed of an electro-optical front end to covert the scattered light into electrical pulses, with the pulse height or area reflecting the particle size. A microcontroller based postprocessing unit for digitizing the pulse information is finally used for calculation of PM values. The electro-optical analog front end senses the scattered light with a photo diode that converts light intensity into a photo current. The photo current is amplified to match the input range of the A/D converter that in many cases is already embedded into a microcontroller being part of the intelligent sensor device.

According to the theory of ideally spherical particles, the scattering light intensity is proportional to the square of the size ($I \sim d^2$) in a particle size range where Lorenz-Mie scattering dominates (sizes larger than the wavelength of the used laser light). In the Rayleigh scattering range, applicable for smaller particles, the proportionality is according to $I \sim d^6$ and thus can be used for diameters near or below the wavelength. However, Rayleigh scattering in such is often neglected.

The relationship between mass of an ideal spherical particle and its size is proportional according to $M \sim d^3$ and therefore it can be stated, when combining the mass calculation and the mapping of the light intensity distribution function, that the resulting proportionality of mass distribution to light intensity is also according to $M \sim I^3$. Electronically, the light intensity can easily be measured with a PIN-photodiode because the photo current is proportional to the light intensity I. Therefore, the analog front end just uses a current-to-voltage converter with amplification, a so called transimpedance amplifier (TIA) to provide the light intensity information to the microcontroller in terms of a voltage.

From this functional perspective, it is obvious that a more pragmatic approach for assessing a low-cost sensor is to first check the measurement uncertainties on the electronic signal processing side by measuring the mapping of the scattering light intensity in terms of the amplified photo current with respect to the calculated PM mass output. Such a check is equivalent to using ideally monodisperse

particles for a test with the optical unit included but simply assuming that the optical scattering process works ideally according to theory. In order to mimic a monodisperse particle distribution it is sufficient to substitute the signal from the TIA by a signal from an external, highly precise pulse generator adjusted to a given amplitude and duration. This pulses can be tailored before the actual experiment to resemble the true scattering pulses obtained from the optical system. This tailoring is achieved when the output signal of the TIA is first analyzed when using either a LED for injecting artificial scattering light pulses into the measurement chamber or from exposing the sensor to real dust with a desired particle size distribution. In both cases an oscilloscope can be used to monitor the TIA output.

This approach was used in this document for judging possible measurement uncertainties for the example of the SDS011 low-cost PM sensor from Nova Fitness. This was done with the goal in mind to identify potential options for compensating accuracy deficiencies in a further step.

Due to the high amplification required to even detect the very small photocurrents from the particles below 1um, the sensor uses an additional shielding that needs to be removed for accessing the active electronics. It would be nearly impossible to measure the photo currents directly at the PIN-diode, therefore only the amplified current can be investigated. The duration of the photo current pulses is in the range of several tens to hundreds of microseconds and the current amplitudes are in the nano-Ampere range. As a consequence, a further assumption has to be made: It is assumed that the TIA shows a ideally linear transimpedance gain $R_f=V_{out}/I_{in}$ that is independent of the input current. Assuming the TIA behaves linearly, test pulses for this investigation can be fed directly into the microcontroller input that is connected to the TIA output and the digital algorithm can be tested.

As an implementation of the test, it would be a natural choice to remove the TIA and to feed the pulse generator pulses directly into the PCB trace interconnecting the TIA footprint with the microcontroller at a suitable location (e.g. at the capacitor C4). However, experiments have shown that removing the TIA completely forces the microcontroller of the SDS011 to output maximum PM-values. However, feeding in pulses from a pulse generator without removing the TIA means to work against the output impedance of the amplifier. Since the output of a precision pulse generator is typically 50 ohm and the TIA output impedance is in the same range, the correct voltage has to be measured simultaneously during parallel operation with an oscilloscope monitoring the voltage at the microcontroller input.

Before setting up the pulse generator substituting the scattering light pulses, the TIA output was analyzed with the scope in a real application to understand what pulse width, amplitudes and waveform shapes occur in reality. In a first test, a LED was used to create light pulses with different duration. It turned out that pulses as small as 50us do not allow the electrical pulse to reach a high level equal to what is achieved as settling value from a step signal. On the other hand a 500 us optical pulse forces the electrical pulse to settle at a peak height but the TIA output decays again to zero after about 400us. This means the amplifier actually implements a bandpass characteristic limiting the high frequencies and blocking DC-levels. The frequency of a sine-wave signal transmitted without attenuation was determined to be approximately 3kHz which reflects the approximate bandwidth of the TIA amplifier.

When exposing the sensor to the smoke of a soldering iron, the signal that appears at the TIA output is shown in fig. 4. From the pulse visible it can be estimated that a realistic average pulse width is in the range of 200us. This width will be smaller for pulses with less height as the average and will be larger for higher pulses. Due to the bandpass filtering characteristic, it is assumed that the area is proportional to the scattered light intensity rather not only the pulse height. Since very bright LED pulses produce waveforms clipped at 3.3V, it was assumed the maximum high level may reach the positive voltage rail of the sensor which is at 3.3V.

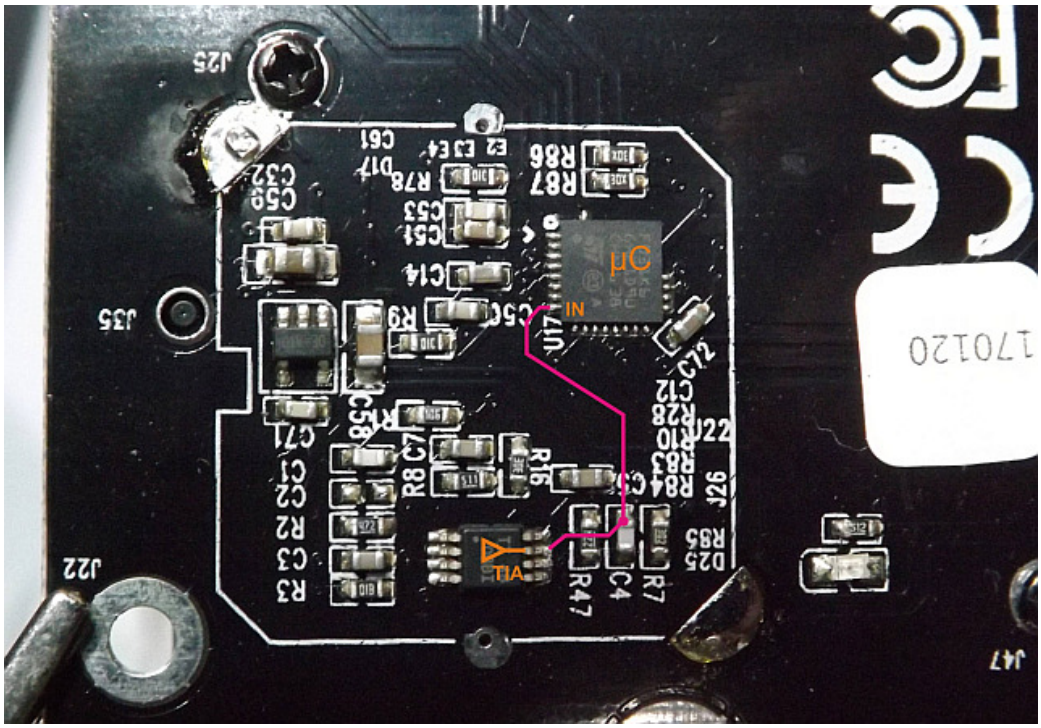


Fig. 1: Assembly side of the SDS011 PCB after removing the shielding cover. The graphic overlay indicates the signal path from the TIA to the analog input of the microcontroller

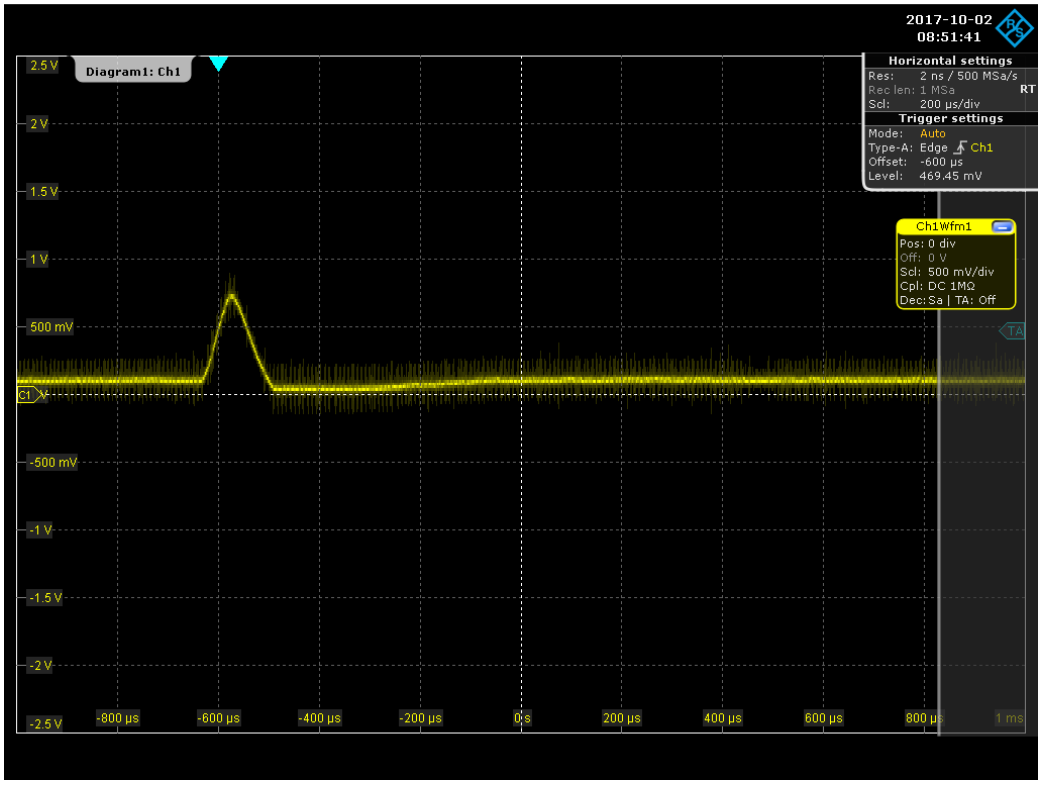


Fig. 2: The TIA output for a pulse generated by a pulsed LED focused into the air intake with a pulse width of 50us

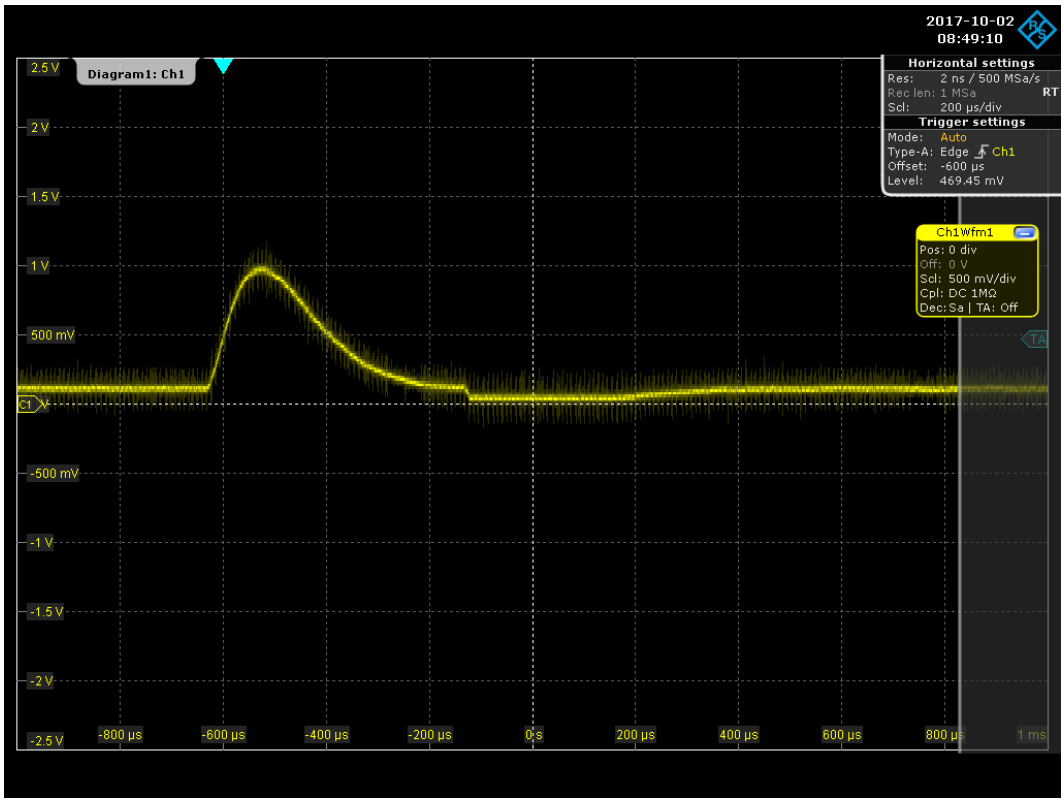


Fig. 3: The TIA output for a pulse generated by a pulsed LED focused into the air intake with a pulse width of 500us

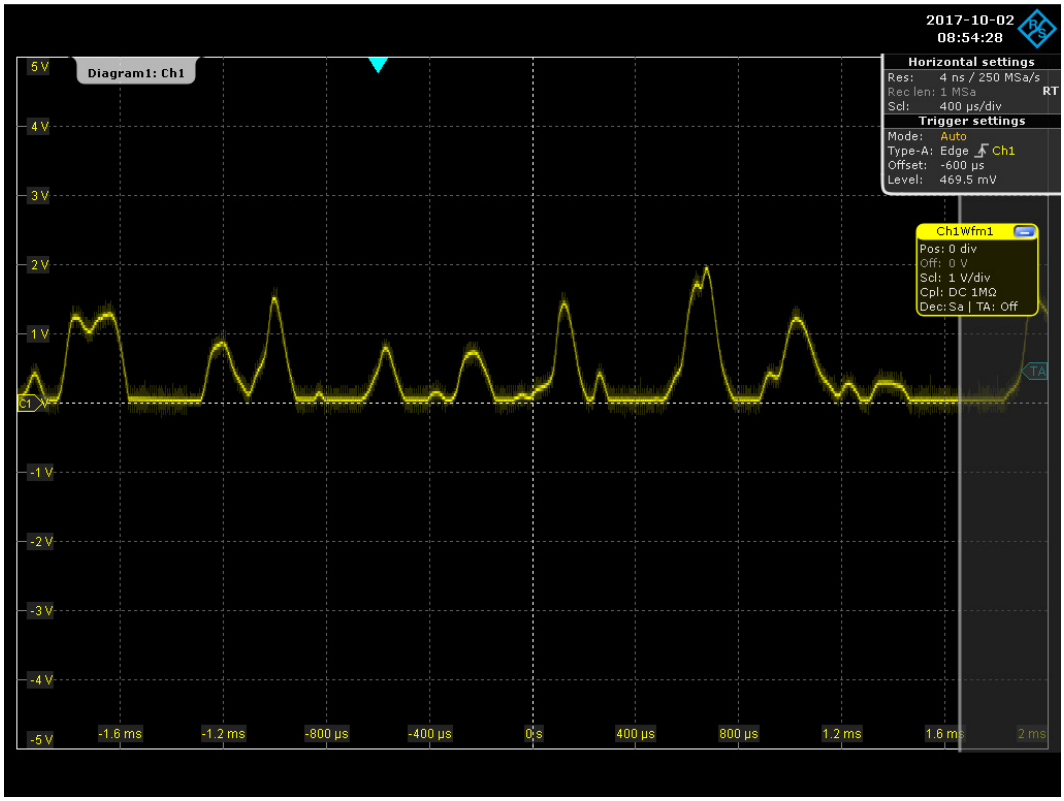


Fig. 4: The TIA output signal as a reaction of the sensor being exposed to the smoke of a soldering iron

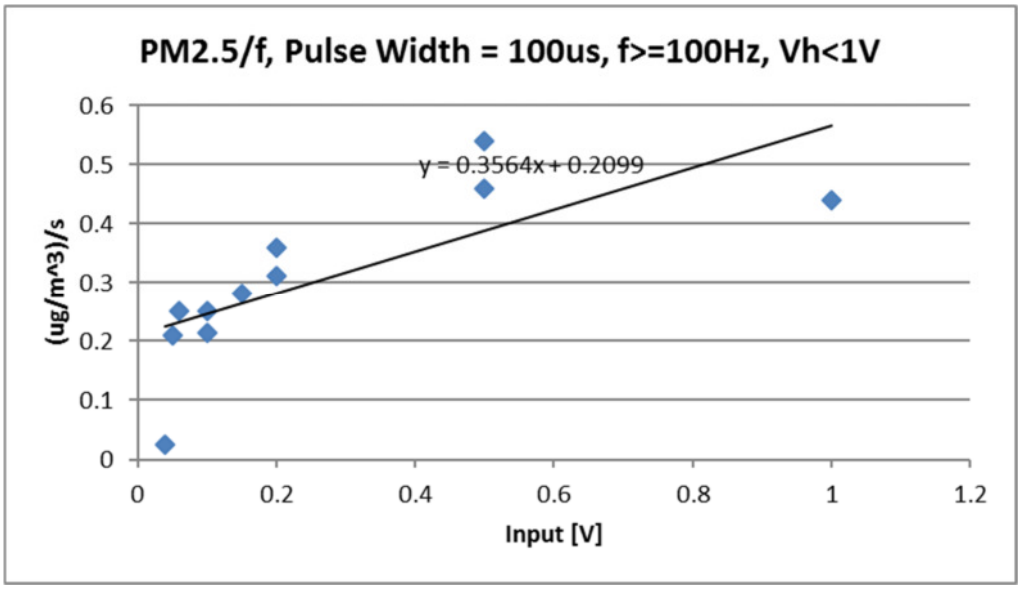
In the following experiment a Keysight 33600A pulse generator was used to generate the pulses that substitute the scattering light pulses. The TIA input was shorted to force the device to output zero PM values when no input from the pulse generator was given. The pulse generator was connected to the capacitor C4 and to sensor ground. The oscilloscope was used to measure the high level of the pulse at the microcontroller input (the low level was set to zero) and to check the pulse width and frequency. Having setup the pulse generator, an Arduino microcontroller was used to read out the sensor PM data and to display the value readings. The PM-values were recorded along with frequency, pulse width, and high level of the input signal (see also appendix 2).

Since the PM values were intended to be plot against input voltage considered to reflect the scattered light intensity and thus particle size, there are still frequency and pulse width as further variable parameters actually creating a 4-dimensional parameter space. Scanning this space at a regular grid was considered ineffective, therefore a trace was stepped through this space that was considered reasonable with respect to the idea of checking the sensors functionality. With this approach, one goal was to see a transition in PM-values when the simulated particle size increases beyond 2.5um or any other possibly implemented size boundary. Another goal was to see if very small particles create equal PM-values for PM2.5 as well as for PM10. And finally a goal was to see the mathematical dependency of the PM10 and PM2.5 values on light intensity and to visualize the PM10/PM2.5 ratio versus light intensity.

In the following, when stepping through the parameter space, the graphs from various track segments from data in appendix 2 are shown. From these graphs, it quickly became evident that the sensor reacted correctly on changing the pulse frequency, reflecting a change in particle count. Therefore, the PM-value/frequency ratio was plotted to reflect the differential mass contribution of a single particle instead of using the absolute mass.

The PM10/PM2.5 ratio in the plots also indicates that no real threshold is implemented that sorts large particles only into a bin dedicated to larger particles and not taken into account for PM2.5 calculation. Instead, after being close to 1 for voltages between 60 and 80mV, the PM10-to-PM2.5 value ratio follows a kind of quadratic relationship with respect to the scattered light intensity. This however means, that PM2.5 readings will depend on PM10 values even for large particles. Or in other words, no real size binning is actually performed in the sensor.

For all data points approached, it could be seen that slight changes in the parameters also mean slight changes in the PM values reported. Furthermore, it could be validated that the pulse width always had a significant effect when being changed. It therefore can be assumed, that the sensor samples the input signal at a high rate and converts the sampled data with an A/D converter that has at least 8bit. It seems to be very likely that the sensor algorithm somehow evaluates the area under the pulses. The area below the pulses could then be used for a subsequent particle size estimation. Since no proportionality to the power of 3 is visible, a likely implementation could be that just the average of the sampled signal is calculated. For PM2.5 the deployed strategy in the algorithm could be to process the samples in a non-linear way that boosts the sample amplitudes of the small pulses and to subsequently subtract the PM10 signal before averaging it for approximating a PM2.5 value.



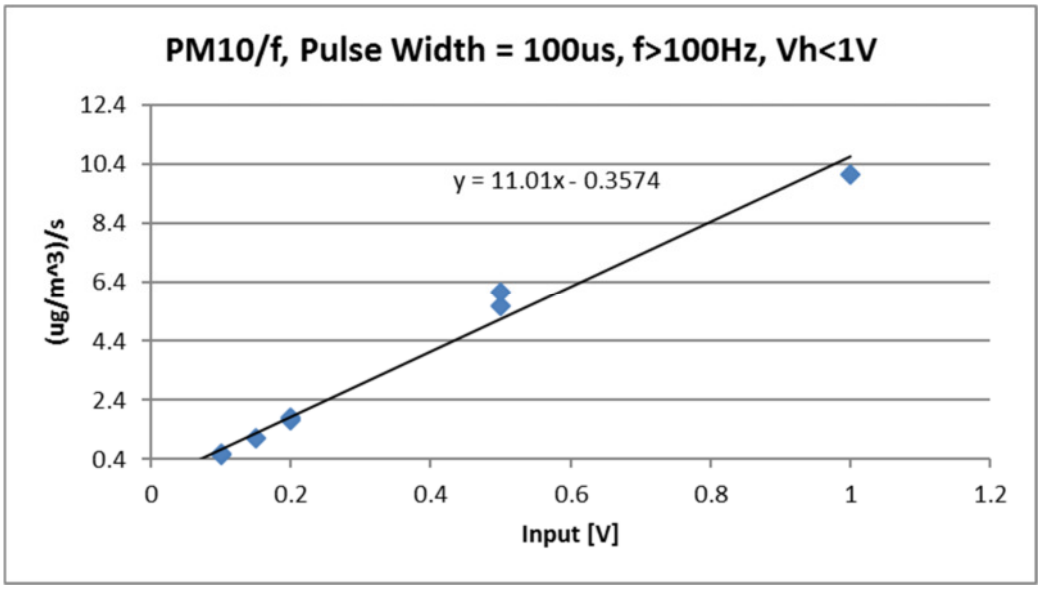
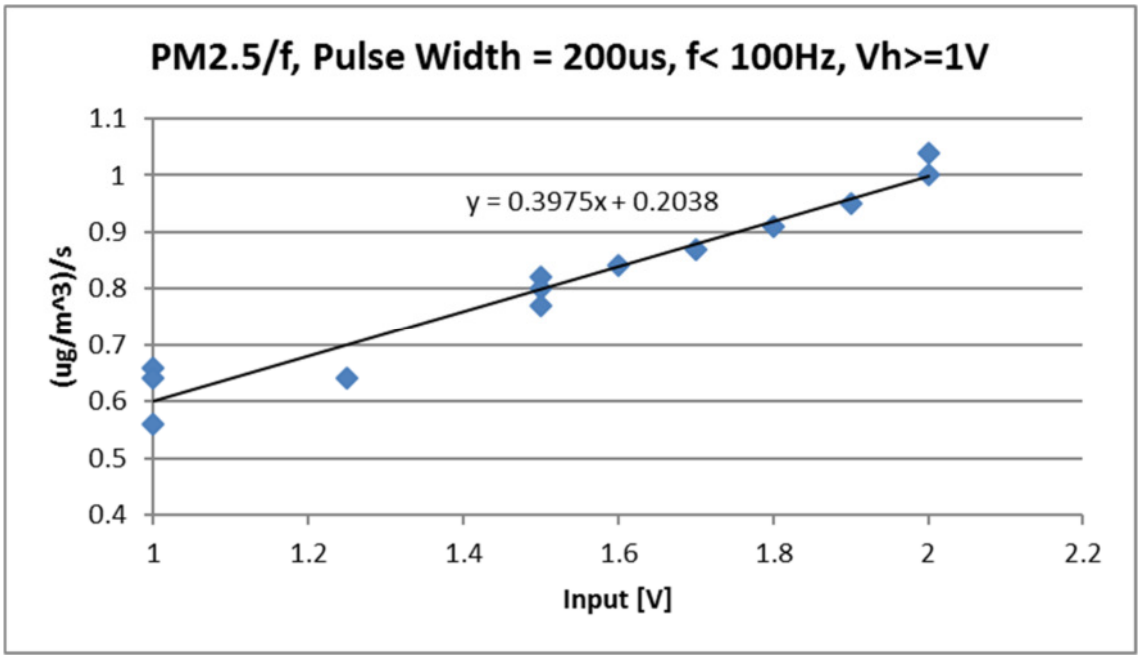


Fig. 5a, b: PM values per particle for small pulses (100us) and frequencies between 100 and 1000Hz

The graphs also show a linear dependency of PM10 and a nearly linear dependency for PM2.5 for the small particles. The PM2.5 characteristic shows an outlier near zero for 40mV input. It was later determined that below 50mV input, the PM values drop to zero regardless of duration, pulse width and frequency. Thus, 50mV input seems to be the lower limit for evaluation. The PM2.5 characteristic deviates slightly from linearity with a maximum at 0.5V input



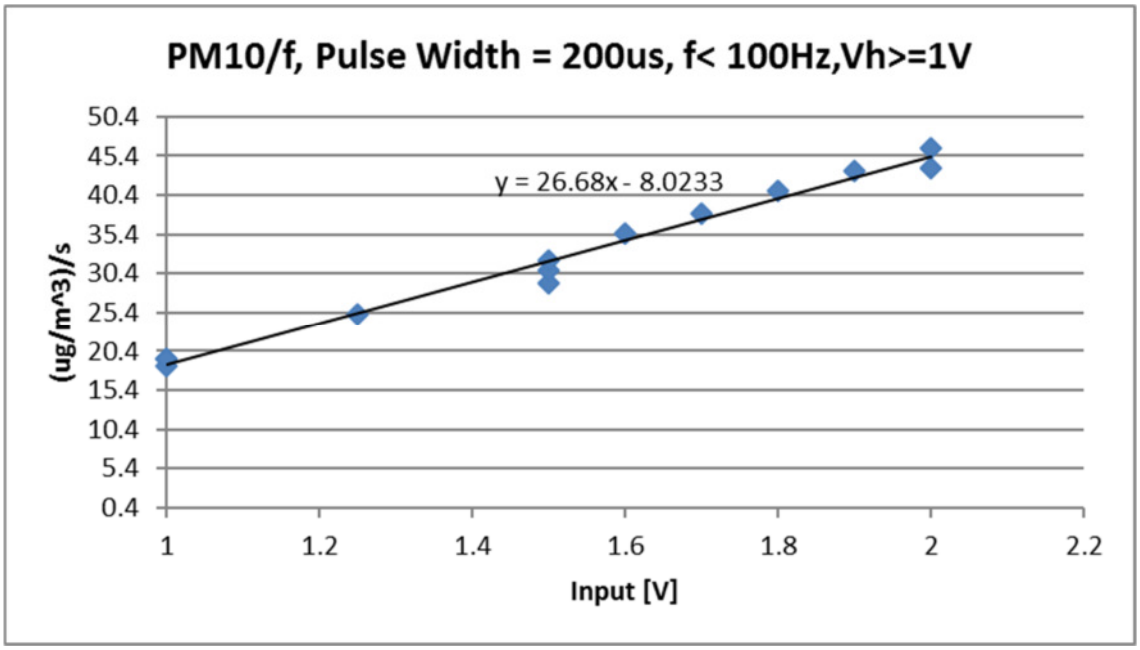
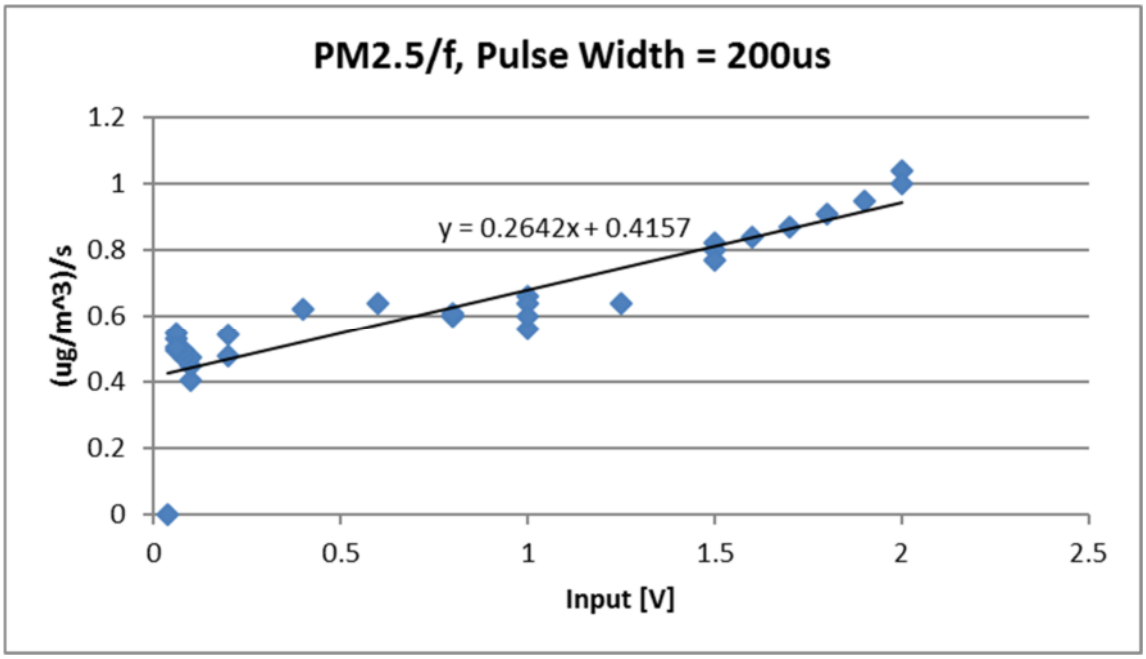


Fig. 6a, b: PM values per particle for larger pulses (200us) and frequencies below 100 Hz

The graphs in fig. 6 also show a linear dependency of PM10 and PM2.5 for the larger particles.



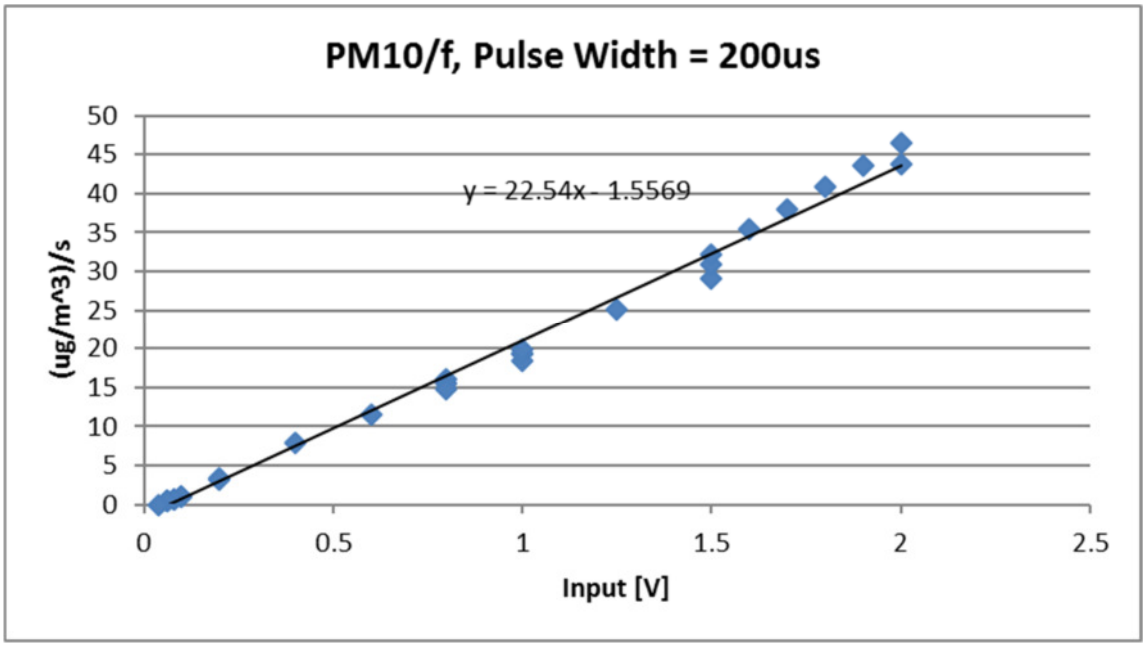


Fig. 7a, b: PM values per particle for pulses with 200us pulse widths covering all frequencies between 10Hz and 1.5kHz

In contrast to the PM10 behavior the PM2.5 behavior in the graphs in fig 7 shows a slight ripple around a linear behavior. In contrast, the PM10 behavior is again almost linear over the whole range of particle sizes.

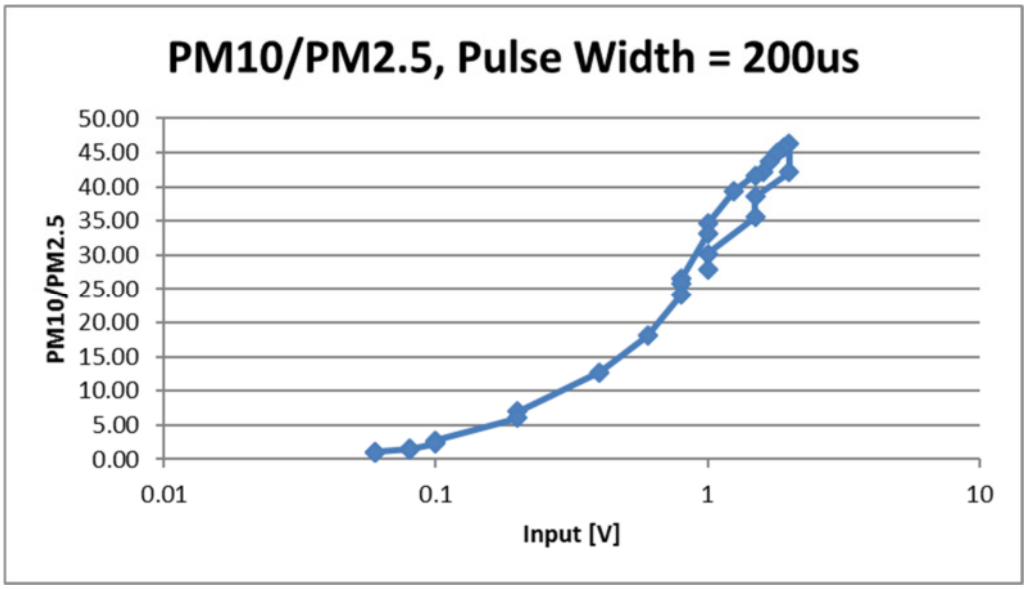


Fig 8: PM10-to-PM2.5-ratio for the data shown in fig. 6a, b. It does not show a clear separation between PM2.5 and PM10 when increasing scattering light intensity

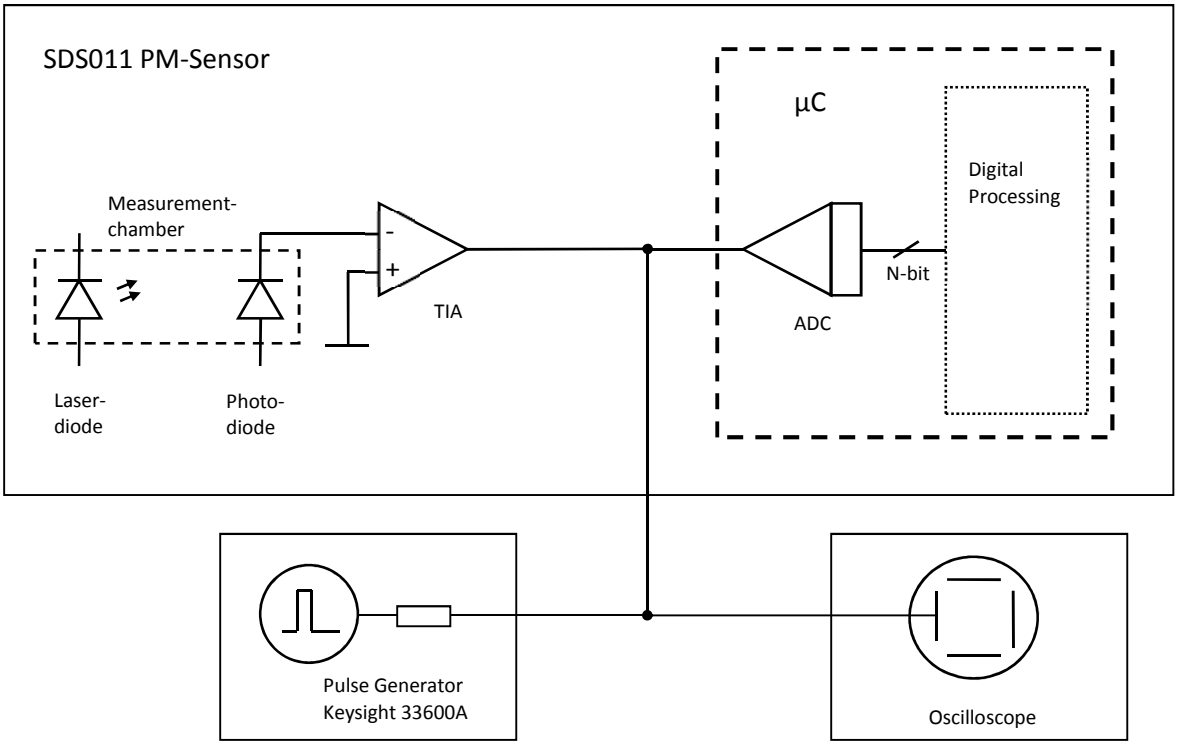


Fig. 9: Measurement setup for testing the SDS low-cost PM-sensor

In order to compare the subsequent processing of the information the sensor takes from the pulses, an idealized simulation of a sensor's behavior, when calculating PM-values as expected from a histogram, was performed (also see appendix 1). In this simulation it was assumed the dependency of size on the scattered light intensity I is according to $d \sim I^{1/2}$ and the (differential) mass of particles is depending on the size according to $dM \sim dN * d^3$. The simulation was run for (dimensionless) light intensities between 0.1 and 1. A PM-value representing PM2.5 was extracted for d<=0.3 and a PM value representing PM10 was extracted at d<=0.9.

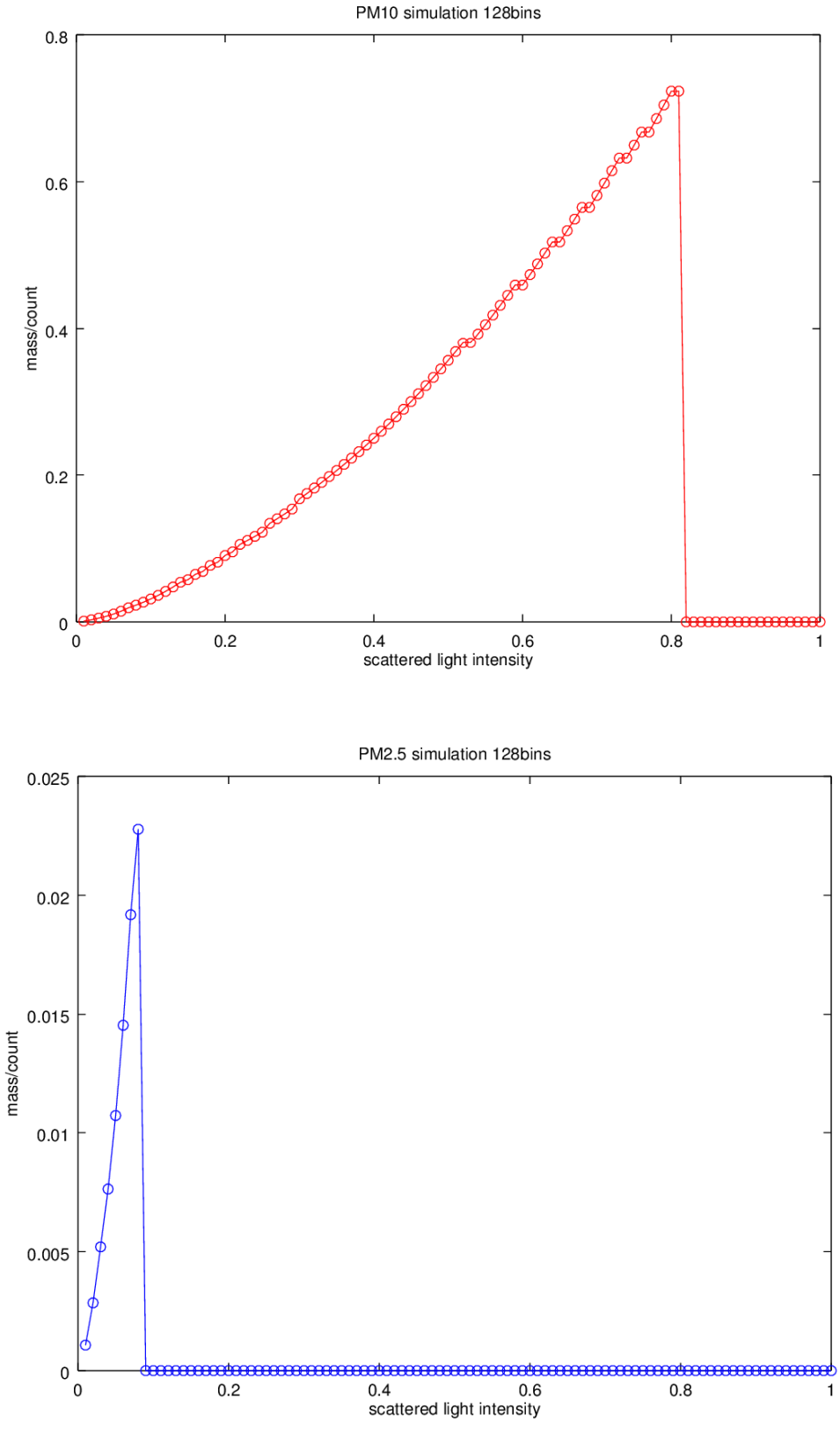


Fig. 10a, b: Simulation of an ideal histogram based PM calculation with 128 intensity bins for PM2.5 (assumed to be 30% of the intensities, blue) and PM10 (assumed to be 90% of the intensities, red)

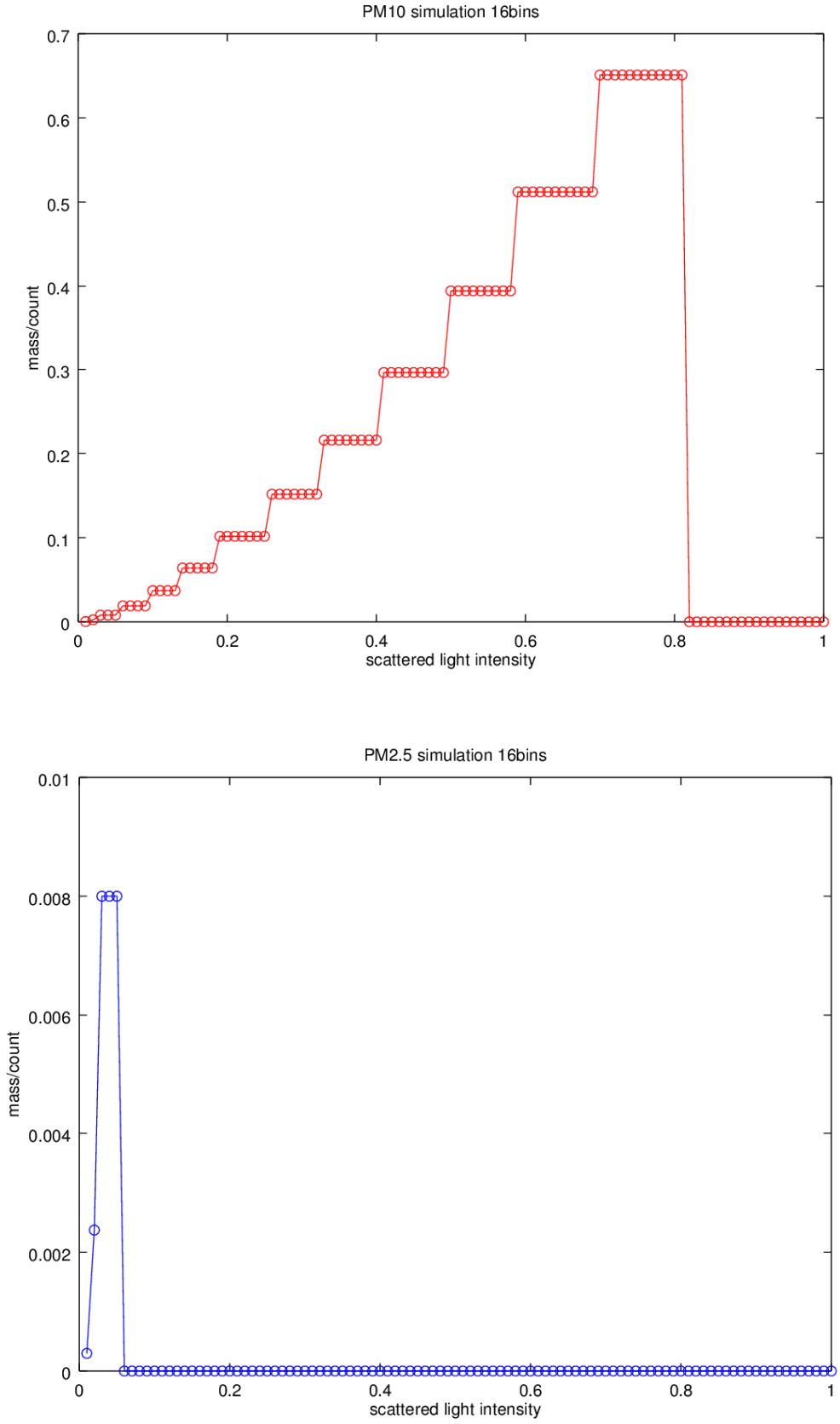


Fig. 11a, b: Simulation of an ideal histogram based PM calculation with 16 intensity bins for PM2.5 (assumed to be 30% of the intensities, blue) and PM10 (assumed to be 90% of the intensities, red)

In a first simulation, the number of equally spaced light intensity bins was set to 128 to better visualize the resulting mass distribution. In the subsequent simulation, the number of size bins was set to 16 to reflect a lower cost approach (fig. 8 and fig. 10). Clearly, the quantization effect of using only few size bins becomes visible in terms of a staircase approximation of the d^3 relationship.

As it can be seen from the simulation, the theoretical particle mass distribution is directly mapped to the result because monodisperse particles with increasing light intensities in each run just ideally sample the theoretical mass distribution at one point. This mapping stops when the size of the highest intensity bin contributing to the mass calculation is reached (ignoring any size separation efficiencies). For larger sizes, the PM result should report a zero result because particles are too large to fall into a bin taken into account for the respective PM calculation.

When these simulation results are compared to the measurements with the SDS011 low-cost sensor experiment, the first striking aspect is that indeed no size limit for a particular PM intensity bin is visible. Even PM2.5 is showing monotonously increasing values with an increasing light intensity signal over the whole light intensity range. This more or less proves that no histogram bins are used for PM calculation. For PM10 the assumption of the sensor designers could have been that there is a limit, given by the TIA amplifier maximum output, linked to the maximum input value of the A/D converter. If this limit reflects the maximum light intensity for a particle taken into account for PM10 calculation, then all information from the peaks just could be summed up or averaged and mapped jointly to a PM mass according to a previous calibration. However, it does not seem that only peak information up to a certain point in the light intensity range is summed up for PM2.5 mass calculation. Otherwise a significant drop of PM readings should be visible when the light intensity signal exceeds a given point at which only PM10 readings should further increase.

Furthermore, it becomes obvious, that the sensor behaves differently than expected with respect to the ideal cubic dependency on increasing light intensity as used for simulation. However, a possible reason for the SDS011 behavior could be that the TIA already compensates the respective power function dependency by implementing the inverse of it using a certain analog circuitry. A further assumption could be that the implemented linear behavior is just a linearization of the non-linear behavior of true particle physics.

Conclusion

The experiment strongly suggests that the SDS011 low-cost sensor does not use a histogram based algorithm for PM calculation. It further must be assumed that PM2.5 readings are significantly incorrect since the SDS011 sensor reports PM2.5 values always with dependency on the PM10 readings for large scattering light intensities rather than with a zero value. When PM2.5 values indeed match the reality for a given particle mass distribution over size then it must be taken into account that the results may significantly change when the mass distribution changes. All these assumptions would make the SDS011 a sensor that may be only accurate for a given reference particle mass distribution. The PM-values would be inherently incorrect, with pretty unpredictable results for extreme cases of mass distribution. The final proof of these assumptions could be done in an experiment that includes the optical sensor part and uses narrowly distributed large particles with size d=5um and d=10um. For such an experiment the predicted outcome is that the SDS011 sensor still outputs a non-negligible PM2.5 reading and the PM10 readings for the two sizes would not reflect the cubical dependency on the light intensity.

The most probable functionality of the used algorithm for PM2.5 calculation is some sort of averaging of non-linearly amplified samples of the TIA signal whereas the PM10 calculation uses almost linearly

amplified samples both avoiding any histogram calculations. If this is indeed the case, PM10 readings of the SDS011 low-cost sensor still could approximate histogram based results with an appropriate calibration due to the limiting effect of the upper ADC limit. For the PM2.5 readings however, it will be very difficult to calibrate to histogram based PM2.5 values independent of the particle spectrum the sensor is exposed to, particularly in presence of a significant mass contributions from large particles. The only exception imaginable would be a particle spectrum with a mass distribution inherently limited to particle sizes below 2.5um.

However, it needs to be noted explicitly that the above conclusions assume an ideally working optical unit of the sensor and it only reflects the electronic signal processing. Nevertheless, it is evident that the limitations from the electronic post-processing can hardly be compensated by the optical-unit. It is therefore expected, that the optical processing will contribute to further measurement uncertainties and thus will further reduce accuracy.

Appendix 1

Listing of the simulation program for PM calculation from scattered light intensity, written for GNU octave (portable also to Matlab of Mathworks):

```
clear;
nrPartRuns=1000;
dx=0.01;
xMax=1;
nrXsteps=11;
nrBins=16;
x0=0.1;
%loop around experiments
x0arr=dx:dx:xMax;
A=0.3; %simulates PM2.5
B=0.9; %simulates PM10
pmAarr = zeros(1,length(x0arr));
pmBarr = zeros(1,length(x0arr));
for i=1:length(x0arr)
  x0 = x0arr(i);
  %generate particles in n runs
  sig = [];
  for n=1:nrPartRuns
    if (0) %equally distributed
      x=xMax*rand(1,nrXsteps);
    else % monodisperse
      x=x0*ones(1,nrXsteps);
    end
    y=x.^(1/2); %size calculated from scattered light (photo current)
    sig = [sig y];
  end
  nrParticles=length(sig);
  maxSig=max(sig);

  bins=linspace(0,xMax^(1/2),nrBins);
  [dN,di]=hist(sig, bins); %dN
  dM=dN.*di.^3;

  pmAarr(i)=sum(dM(1:nrBins*A))/nrXsteps/nrPartRuns;
```

```
  pmBarr(i)=sum(dM(1:nrBins*B))/nrXsteps/nrPartRuns;
end
figure; plot(x0arr,pmAarr,'o-')
xlabel('scattered light intensity');
ylabel('mass/count')
tit = ['PM2.5 simulation ' num2str(nrBins) 'bins'];
title(tit);
figure; plot(x0arr,pmBarr,'or-');
xlabel('scattered light intensity');
ylabel('mass/count');
tit = ['PM10 simulation ' num2str(nrBins) 'bins'];
title(tit);
```

Appendix 2

Measurement results from the experiment obtained for the SDS011:

Pulse High Level[V]	Pulse Width [us]	Frequency [Hz]	PM2.5 [ug/m^3]	PM10 [ug/m^3]	PM2.5/Freq [ug/m^3/s]
0.04	100	1000	24	24	0.024
0.05	100	1000	210	220	0.21
0.06	100	1000	250	260	0.25
0.1	100	1000	250	545	0.25
0.1	100	500	107	290	0.214
0.15	100	500	141	560	0.282
0.2	100	500	180	863	0.36
0.2	100	250	78	452	0.312
0.5	100	250	135	1396	0.54
0.5	100	100	46	605	0.46
1	100	100	44	1008	0.44
1	200	100	66	1846	0.66
1	200	50	32	967	0.64
1.5	200	50	41	1459	0.82
1.5	200	25	20	771	0.8
2	200	25	26	1095	1.04
2	200	10	10	464	1
1.9	200	10	9.5	435	0.95
1.8	200	10	9.1	409	0.91
1.7	200	10	8.7	380	0.87
1.6	200	10	8.4	355	0.84
1.5	200	10	7.7	321	0.77
1.25	200	10	6.4	252	0.64
1	200	10	5.6	194	0.56
1	200	25	15	497	0.6
0.8	200	25	15	400	0.6
0.8	200	50	30	777	0.6
0.8	200	100	61	1473	0.61
0.6	200	100	64	1157	0.64
0.4	200	100	62	787	0.62
0.2	200	100	48	334	0.48
0.2	200	200	109	658	0.55
0.1	200	200	81	221	0.41
0.1	200	300	134	331	0.45
0.1	200	400	189	428	0.47
0.08	200	400	189	297	0.47
0.08	200	500	249	364	0.50
0.06	200	500	249	262	0.50
0.06	200	750	399	419	0.53
0.06	200	1000	546	573	0.55
0.06	200	1500	755	793	0.50
0.04	200	1500	0	0	0.00