Monte-Carlo Simulations of a N-bin PM-Sensor Model for Studying Particle Size Quantization Errors

Bernd Laquai,22.9.2017

The detection of particles in the measurement volume of a PM-Sensor is a statistical phenomenon that is time dependent. The appropriate mathematical description is a stochastic process. The computer based simulation of stochastic processes can be efficiently achieved with the Monte-Carlo methodology and helps to get a better insight in the behavior of systems that deal with random processes such as particle sizing and counting.

The Monte-Carlo methodology was used in this study to develop a model for an ideal PM-sensor that quantizes the size of particles into N-bins. This model was used to implement an ideal calibration scheme for determination of particle mass and to study the impact of size quantization on the measurement accuracy after calibration.

For the calculation of differential mass from the differential counts of particles in each size bin a simplified theory was used neglecting any Cunningham slip corrections or shape factor. The equation used is the simple relationship valid for perfect spherical particles:

$$dM = dN * \pi/6 * d^3 * \rho \qquad (1)$$

During Monte-Carlo simulation a random number generator that outputs numbers uniformly distributed between 0 and 1 was used to generate a large sequence of random numbers that were mapped using a well-determined function (y=1/sqrt(2*x)) to produce a sequence of particle size values that reflect particle sizes (diameters) distributed between 0.1um and 10um. As an example, fig. 1 shows a sequence of 1 million particle sizes generated by means of this Monte-Carlo simulation.

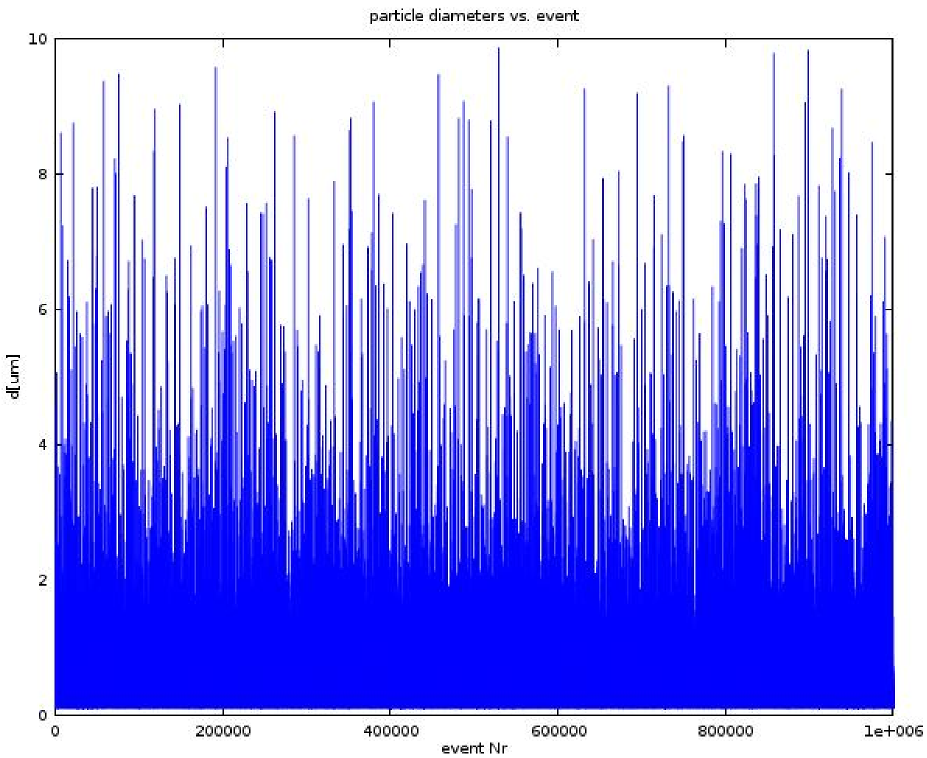

Fig. 1: A random sequence of 1 million particle sizes generated for simulation of size quantization

The mapping function for the random generator was particularly chosen such that the resulting particle size distribution reflecting the relative count of particles dN is proportional to 1/ d^3. The reasoning behind this approach is, that in such a case the mass distribution, reflecting the relative differential mass of the particles dM that is proportional to dN* d^3, then is uniformly distributed.

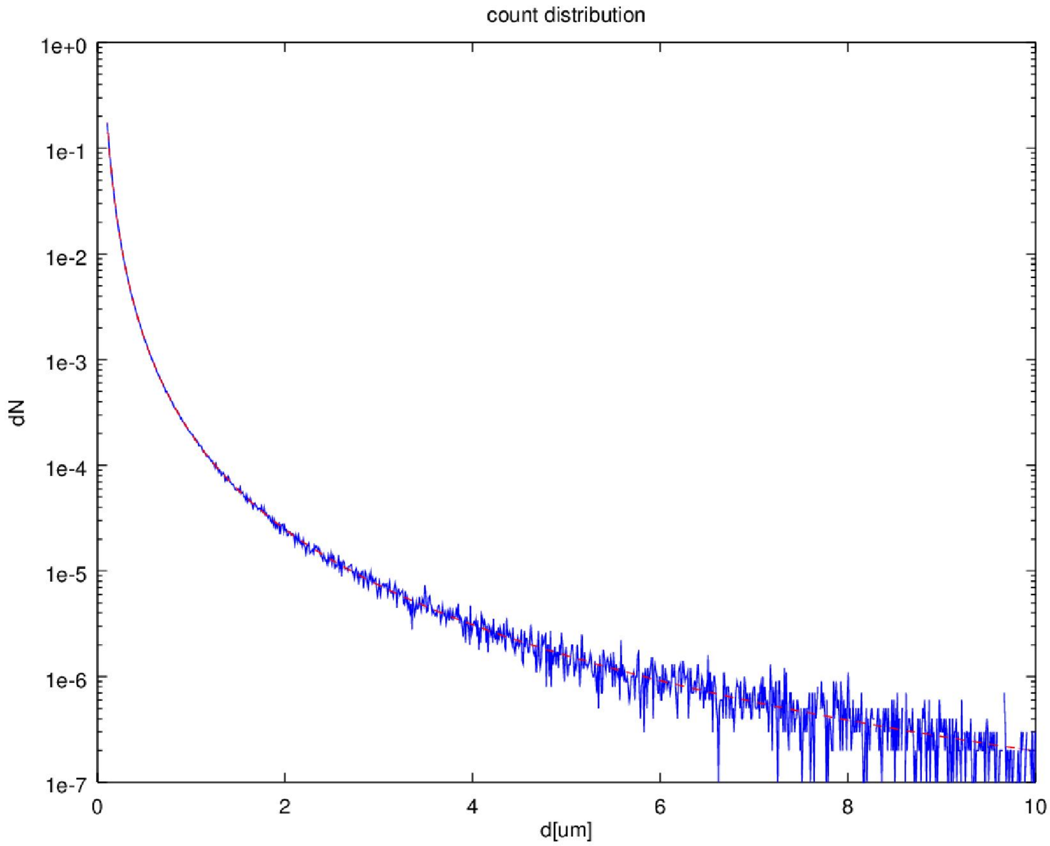

Fig. 2: Particle size distribution function from simulation (blue curve) and theoretical calculation (red dashed line, ~1/ d^3)

It quickly becomes obvious that the probability density for particles sizes larger than 1um quickly drops to values of less than 1e-4. That means most of the particles are less than 1um in size. This can also be seen in fig.1 where only few of the 1 million particles reach large sizes close to 10um.

In a next step a quantization into 32 equally spaced bins was performed to reflect a high-resolution PM-sensor. The differential dN counts shown on the y-axis are relative counts (actually dN/N when N is the number of particles simulated) to also reflect probability rather than absolute counts. As it can be seen, the relative count bars (drawn with blue lines) approximate well the theoretically expected values (red dashed line). However, for large sizes the approximation of the theoretical curve becomes more uncertain due to the much smaller absolute counts used in the simulation (1E6 particles). It can be seen that the relative count in the first bin close to 0.1um is filled to almost 100% (probability 1) and the last bin yields a relative count of only 10^{-5} which means only 10 of 1 million particles fall into this bin. This is the consequence of the 1/d^3 size dependency for the goal to achieve a constant mass distribution.

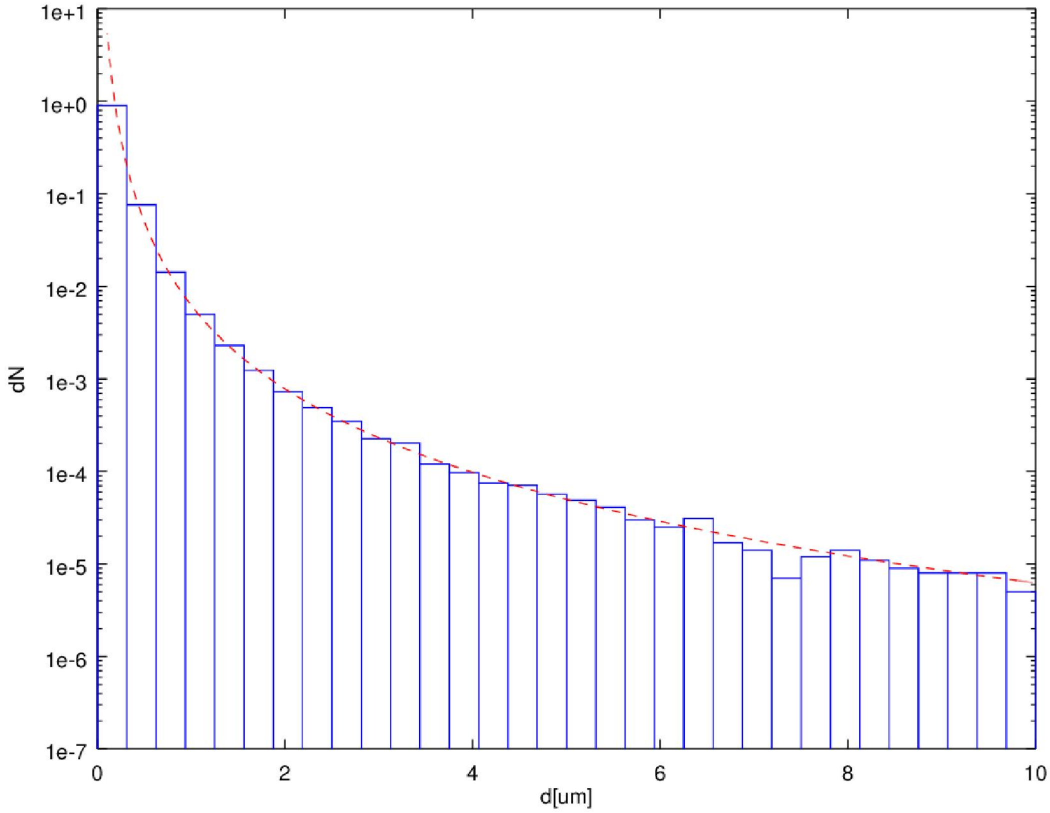

Fig. 3: Binning the particle sizes into 32 equally spaced bins of a PM-sensor, blue simulation, red expected theoretical values

When dN is adjusted by simulation such that dM is uniformly distributed, then a calibration factor $f_i=1/dN_i$ can be obtained for each bin i that translates dN_i into dM_i. This calibration factor f_i can later on be used for any other distribution dN, provided the quantization error is small enough to approximate the ideal relationship according to the above equation (1) precisely enough.

To simulate the mass distribution behavior, the calibration factors $f_i=1/dN_i$ were determined from a previous representative run. Then in first run with calibration, the calibration was applied in terms of calculating dM_i from dN_i*f_i without changing the particle size distribution. The resulting mass distribution is shown in fig. 4. Neglecting the constant factor of $\pi/6 * \rho$ (without loss of generality), dM is then expected to be uniformly distributed with $dM_i = 1$ for each bin i.

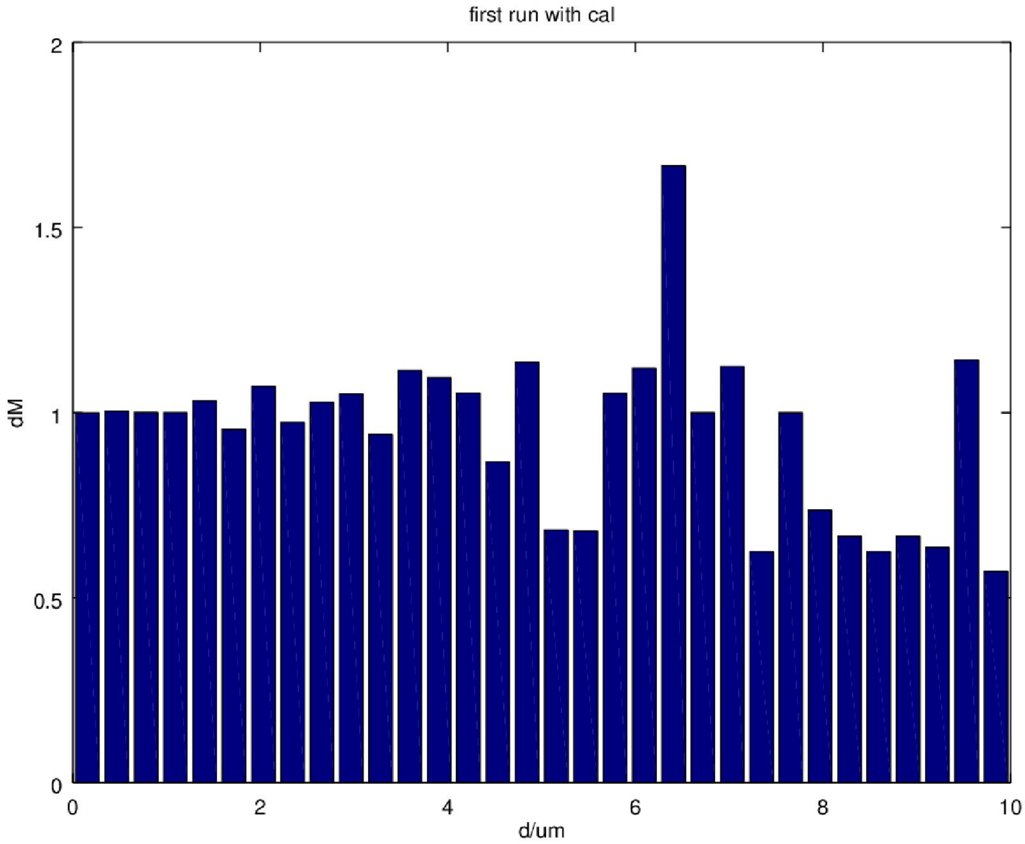

Fig. 4: Simulated mass calculation result with calibration applied to the same particle size distribution

As it can be seen, the result reflects well the expected constant mass distribution with a value of 1 for sizes below 2.5um. Above particle sizes of 2.5um the simulation using 1 million particles becomes increasingly noisy. For large particle sizes, close to 10um, the result shows a substantial variance for multiple executions due to the small number of particles in this bin that hardly can be statistically representative.

In the following second run, the particle distribution was changed, such that the particle sizes were uniformly distributed between 0.1um and 10um and no longer proportional to $1/d^3$. The simulation was executed keeping the previous calibration factors. We expect that the resulting mass distribution should then be proportional to d^3 according to equation (1). This is indeed the case. However, for large particle sizes the variance again becomes substantial due to the small sample set in these weakly populated bins for the very large particles.

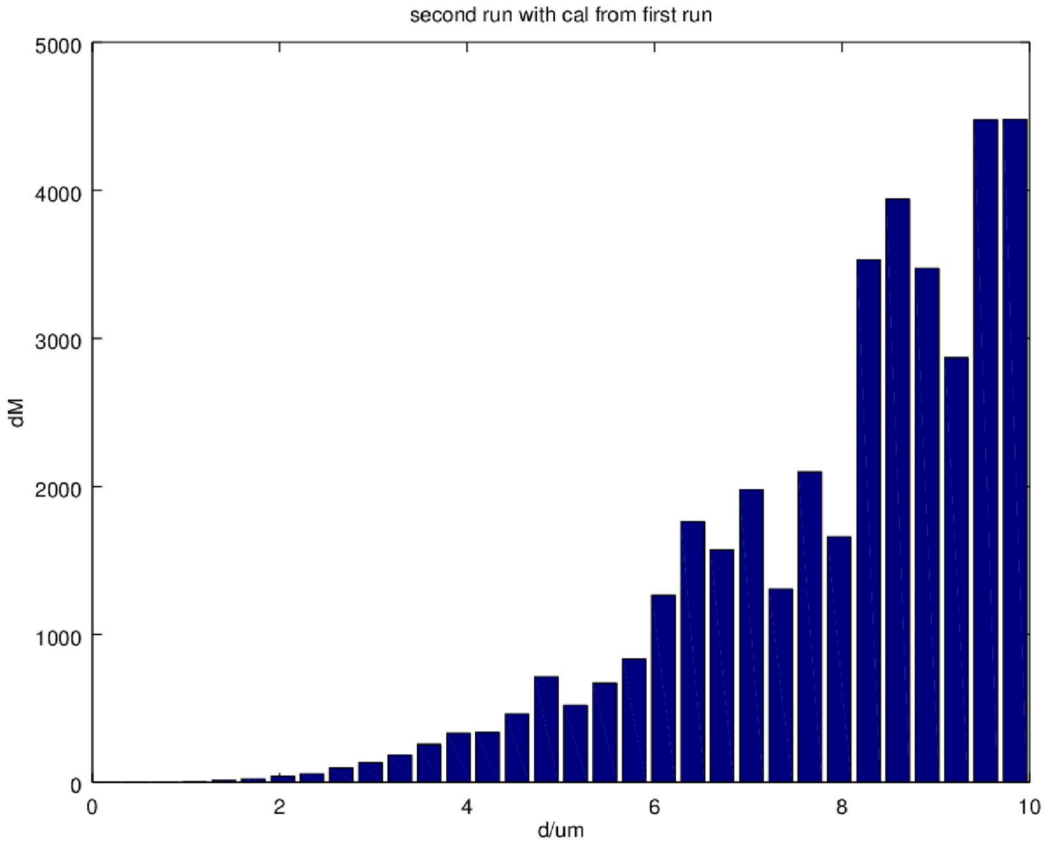

Fig. 5: Simulation of mass calculation with a calibrated PM-sensor model applied to a uniform distribution of particle counts dN

Interestingly, the whole calibration concept even works for 2 bins as long as the particle size is distributed across full bins, from the lower bin boundary to the upper bin boundary (see fig 6a). Even more, this is valid if a size distribution that is proportional to ~1/ d^3 just fills one single bin of the two bins available, i.e. when particles are distribute from 5-10um in the simulation example of fig. 6b. In this case $dN_1 = 0$ and $dN_2 = 1$, or in other words the particle size distribution fills the second bin to 100%. Just from theory f_1 then tends towards infinity and f_2 is 1, therefore dM_1 is 0 and dM_2 is 1, as expected.

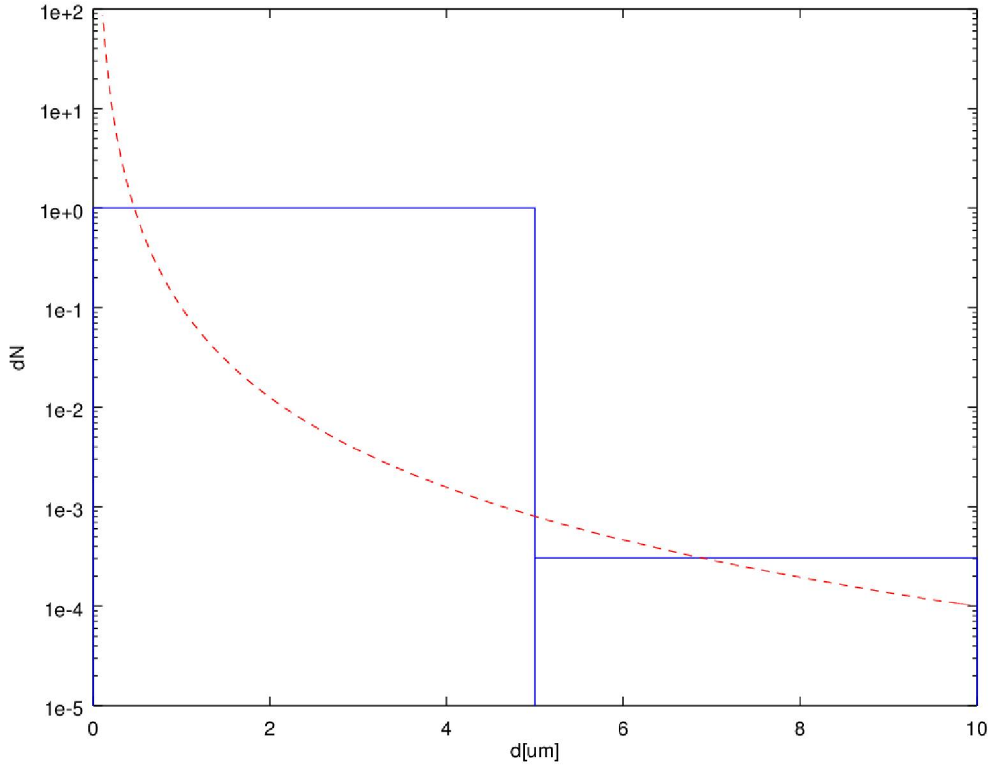

Fig 6a: Binning the count distribution 0.1-10um with a PM-sensor using only 2 bins for a working range of 0.1-10um

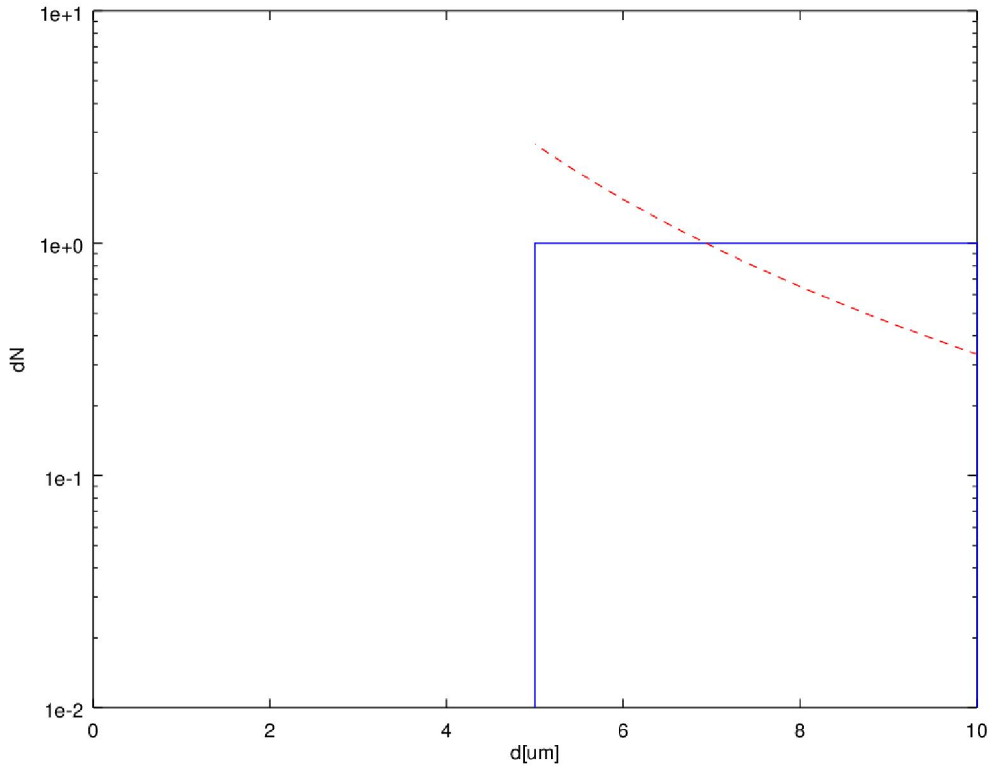

Fig 6a: Binning a count distribution reduced to 5-10um with a PM-sensor using only 2 bins for a working range of 0.1-10um

From this extreme 2-bin PM-sensor simulation examples, we can derive that in case the particle distribution is narrowly monodisperse and just fills one bin x, the calibration factor f_x is assigned to the relative count $dN_x = 1$ (filled completely) regardless of particle size as long as particle size is within the bin boundaries. As a consequence, the resulting differential mass of $dM_x = 1$ will always be reported regardless of particle size. This is obviously a massively wrong result caused by the coarse granularity of the 2-bin size classification. This error may become the larger, the larger the bin width is. The size distribution therefore must stretch at least across a whole bin to give correct results.

The requirement, that the size distribution must stretch across whole bins for obtaining correct results, is of course also valid for a PM-sensor that uses N-bins. This observation however, leads to the fundamental conclusion, that the particle size distribution may not change rapidly versus size. To be more precise, the change rate of the particle distribution function versus size must be such smooth, that each bin is filled with a well-balanced fraction of the size distribution that fits to the assigned calibration factor f_x.

There is a well-known analogy in digital information theory known as the sampling theory which is based on the fundamental Nyquist theorem (according to the inventor Harry Nyquist). This theorem basically says that when a time dependent signal is sampled for further digital processing, the sampling time interval must be shorter than half of the smallest change period contained in the time signal. In analogy to the Nyquist's theorem in the time domain, we can conclude that there is a similar restriction for a particle size distribution that must be fulfilled in the particle size domain, when the particles are binned with a limited number of bins versus size for further processing such as mass calculation. When we define a minimum change period in size for a continuous particle size distribution then we can state approximately that the bin width used for binning this size distribution into a size-discrete form must be smaller than half of the change period across the whole measurement range of a PM-sensor to yield correct results (equally spaced bin widths are assumed). More practically speaking, the bin width must be by a factor of at least 2 smaller than the width of any peak visible in the (reasonably averaged) size distribution of the measured aerosol.

This outcome is indeed not really surprising and appears to be pretty logical. It simply limits the use of low-cost PM-sensors that implement only a few size bins to particle spectra that are smoothly distributed versus size. In turn it prevents low-cost sensors from being successfully applied for measuring monodisperse particles or when a broadly distributed particle spectrum contains narrow peaks of significant height. However, we can also state, that a sufficiently smooth-distributed particle spectrum will be measured correctly even with a low-cost sensor using only few bins.

The key question that remains however is, if particle size distributions have to be measured in reality are they indeed distributed smoothly enough to allow the correct deployment of a low-bin-count PM-sensor? This question can only be answered when the particle spectrum is measured in advance with a high-bin-count instrument to ensure that the requirements of a low-cost low-resolution PM-sensor are met.

As an example, when a low-cost PM-sensor should be used for road dust measurements, a representative particle size distribution measurement at the specific measurement site must be done first. However, when we consider Arizona road dust (ARD) to be representative enough for this site, we can estimate the minimum change period in particle size of the distribution function from publications published in the internet (see fig. 8). Just from a coarse estimate when looking at the graph in fig. 8, we can see that the distribution was binned "representatively" with a minimum bin width of about 2um for the area where most of the mass is concentrated (between 1 and 10um). Transferring this to equally spaced size bins, discussed in the above consideration and assuming a measurement range of the PM-sensor from 0.1 to 10um this means that at least 10 bins are required to fulfill the criteria derived previously. As it can be seen from the measurement approach used for producing the cited graph, a logarithmical spacing of the bins is advantageous to resolve short distribution changes in the sub-um range and to adapt to wider change periods for particles with more than 10um in size by still maintaining a number of 9 size bins for the whole measurement range. This is the reason why most commercial instruments use non-equal spacing of size bins.

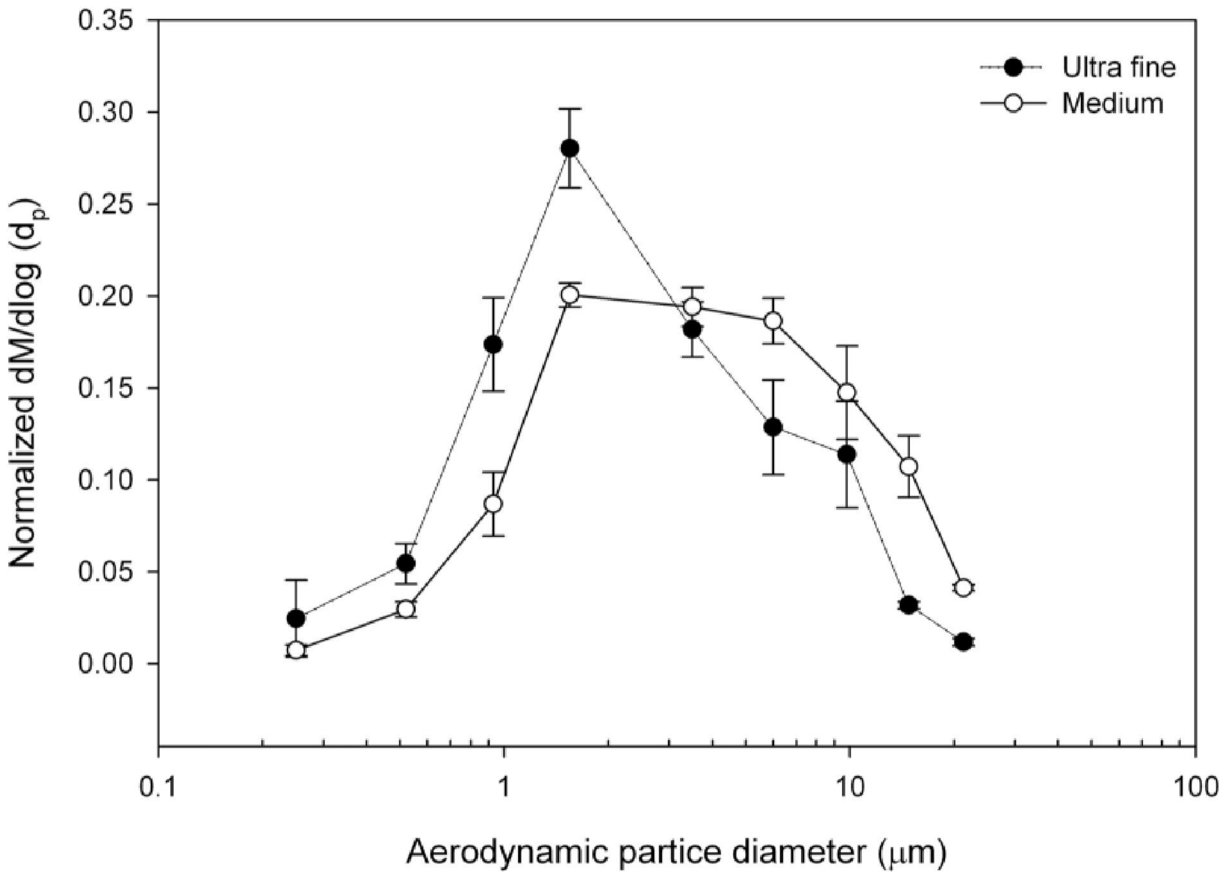

Fig. 8: Size distributions of two grades of Arizona road dust according to ISO 12103-1, 1997 (Source: Collection Efficiencies of High Flow Rate Personal Respirable Samplers When Measuring Arizona Road Dust and Analysis of Quartz by X-ray Diffraction, P. Stacey et al.; Ann. Occup. Hyg., 2014, Vol. 58, No. 4, 512–523)

Finally, we can conclude that low-cost sensors with reduced number of size bins (classes) can be used with sufficient accuracy provided the particle spectrum shows a broad and smooth distribution that matches the bin resolution. This again is highlighted illustratively in the 4-bin example shown in fig. 9. In contrast to the smooth $1/d^3$ distribution shown with the green curve that would work well, we would have to expect substantial measurement errors for the red particle distribution showing a short change period in size of less than a bin width during a substantial peak.

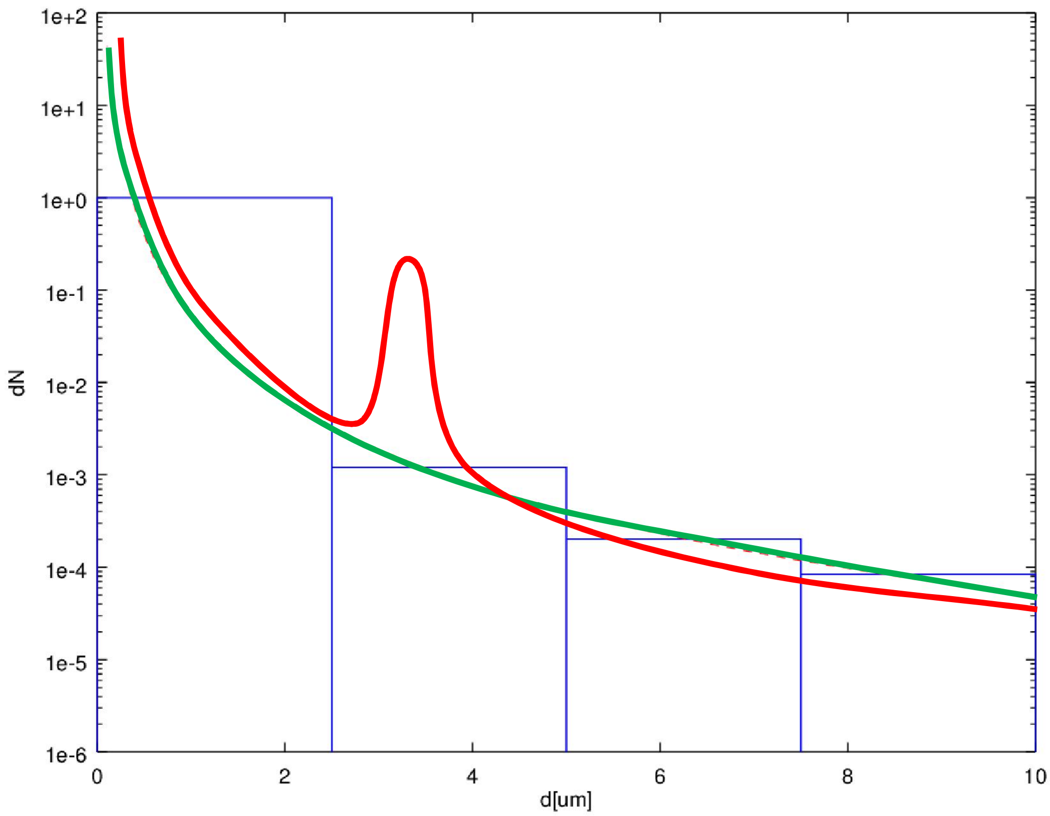

Fig. 9: Application of a 4-bin PM-sensor for a suitable particle size distribution (green) and for a pathological size distribution (red) that won't be resolved correctly by the sensor

Literature

/1/ Athanasios Papoulis; Probability, Random Variables and Stochastic Processes; Mc Graw Hill, 1965
/2/ Collection Efficiencies of High Flow Rate Personal Respirable Samplers When Measuring Arizona Road Dust and Analysis of Quartz by X-ray Diffraction, P. Stacey et al.; Ann. Occup. Hyg., 2014, Vol. 58, No. 4, 512–523

Simulation Code

The below code was developed under GNU Octave but also executes under Matlab from Mathworks

```
%count distribution dN between ofst and a+ofst, such that dM = dN*d^3 const with density 1
clear;
nrSamp=1000000; %number of particles generated
randNum=rand(1,nrSamp);
ymin = 0.1; % min particle size [um]
ymax = 10; %max particle size [um]
ofst= 1/(2*ymax^2);
a= 1/2*(1/ymin^2-1/ymax^2);
aux=(a*randNum+ofst);
nc=100;
[N,X]=hist(aux,nc);
Q=N/nrSamp;  %true aux
%plot(X,Q);hold on;
%plot(X,1/nc*ones(size(X)),'r'); hold off; %theoretical aux
%disp(sum(Q)); % != 1
%
y=1./sqrt(2*aux); % these are the particle sizes
figure; plot(y);
xlabel('event Nr');
ylabel('d[um]');
title('particle diameters vs. event');
nc=1000;
```

```
[N,X]=hist(y,nc);
Z=N/nrSamp; %true dN vs. d
figure; semilogy(X,Z,'-');hold on;
ymax=1./sqrt(2*ofst);
ymin=1./sqrt(2*(a+ofst));
P=1./(a*X.^3)/nc*(ymax-ymin); %theoretical dN
semilogy(X,P,'r--'); hold off;
xlabel('d[um]');
ylabel('dN');
title('count distribution');
%disp(sum(P)); % != 1 %check distribution
%
if (0) %check that mass density is constant
R=Z.*X.^3*a/(ymax-ymin); %true dM vs. d
figure;plot(X,R);
title('mass distribution');
%disp(sum(R)); % != 1 %check distribution
end

%low-cost sensor
rangeMin=0; %range of the sensor
rangeMax=10;
nrBins=4; % number of bins used
binWidth=(rangeMax-rangeMin)/nrBins;
binMid=binWidth/2:binWidth:rangeMax;
[N1,X1]=hist(y,binMid);
sLC = sum(N1);
binFac=binWidth*(nc/(ymax-ymin));
minLy=10^(floor(log10(min(P*binFac)))-1);
mySemilogBar(X1,N1/sLC,'b',minLy);hold on %actual count distribution
semilogy(X,P*binFac,'r--');hold off; %expected theoretical shape
xlabel('d[um]');
ylabel('dN');
tit=sprintf('%d-bin sensor vs. ideal sensor',nrBins);

function mySemilogBar(x,y,str,minLy);
len=length(x);
xgraph=zeros(1,4*len);
ygraph=xgraph;
i=1;
d1=(x(i+1)-x(i))/2;
d2=(x(i+1)-x(i))/2;
xgraph((i-1)*4+1)=x(i)-d1;
xgraph((i-1)*4+2)=x(i)-d1;
xgraph((i-1)*4+3)=x(i)+d2;
xgraph((i-1)*4+4)=x(i)+d2;
ygraph((i-1)*4+1)=minLy;%y(i);
ygraph((i-1)*4+2)=y(i);
ygraph((i-1)*4+3)=y(i);
ygraph((i-1)*4+4)=minLy;%y(i+1);
for i=2:length(x)-1
    d1=(x(i)-x(i-1))/2;
    d2=(x(i+1)-x(i))/2;
    xgraph((i-1)*4+1)=x(i)-d1;
    xgraph((i-1)*4+2)=x(i)-d1;
    xgraph((i-1)*4+3)=x(i)+d2;
    xgraph((i-1)*4+4)=x(i)+d2;
    ygraph((i-1)*4+1)=minLy;%y(i-1);
    ygraph((i-1)*4+2)=y(i);
    ygraph((i-1)*4+3)=y(i);
    ygraph((i-1)*4+4)=minLy;%y(i+1);
end
i=len;
d1=(x(i)-x(i-1))/2;
d2=(x(i)-x(i-1))/2;
xgraph((i-1)*4+1)=x(i)-d1;
xgraph((i-1)*4+2)=x(i)-d1;
xgraph((i-1)*4+3)=x(i)+d2;
xgraph((i-1)*4+4)=x(i)+d2;
ygraph((i-1)*4+1)=minLy;%y(i-1);
ygraph((i-1)*4+2)=y(i);
ygraph((i-1)*4+3)=y(i);
ygraph((i-1)*4+4)=minLy;%y(i);
figure;semilogy(xgraph,ygraph,str);
```