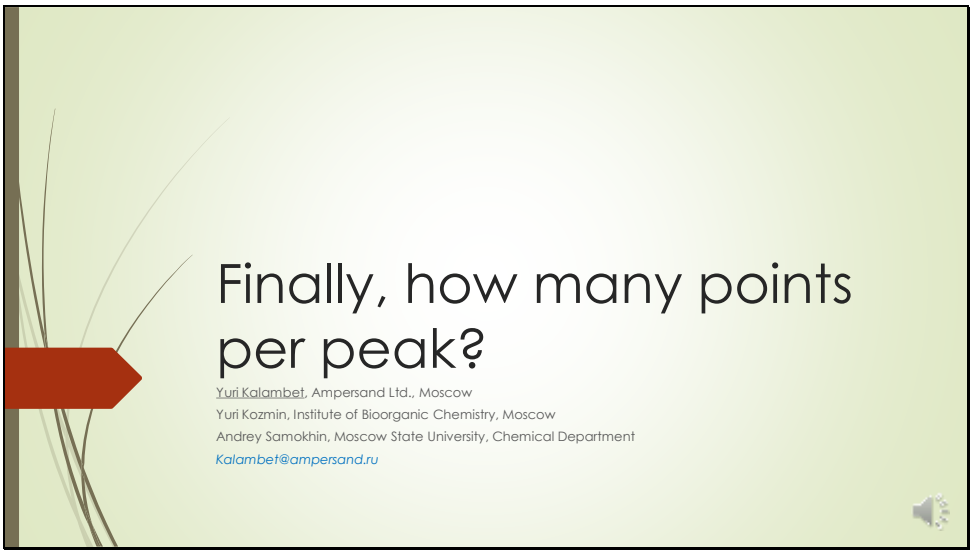
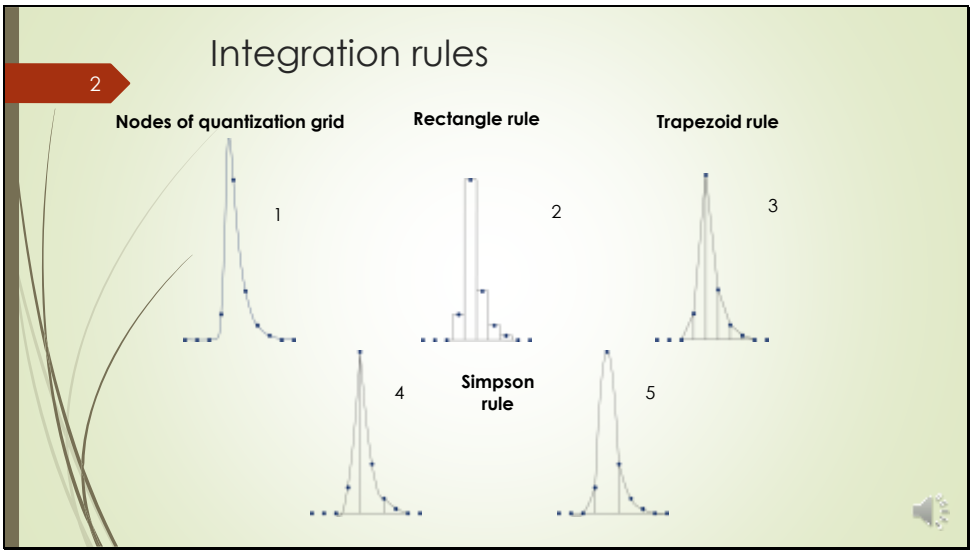
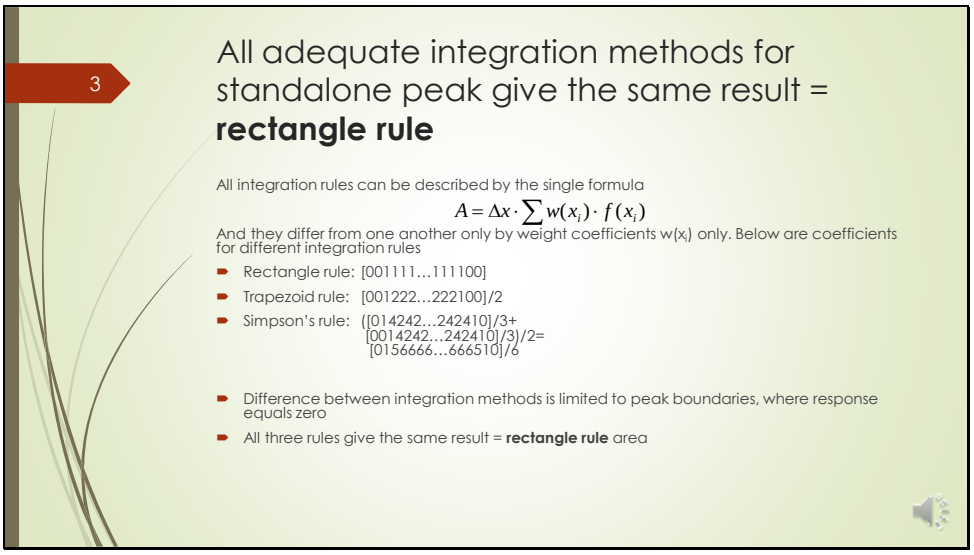


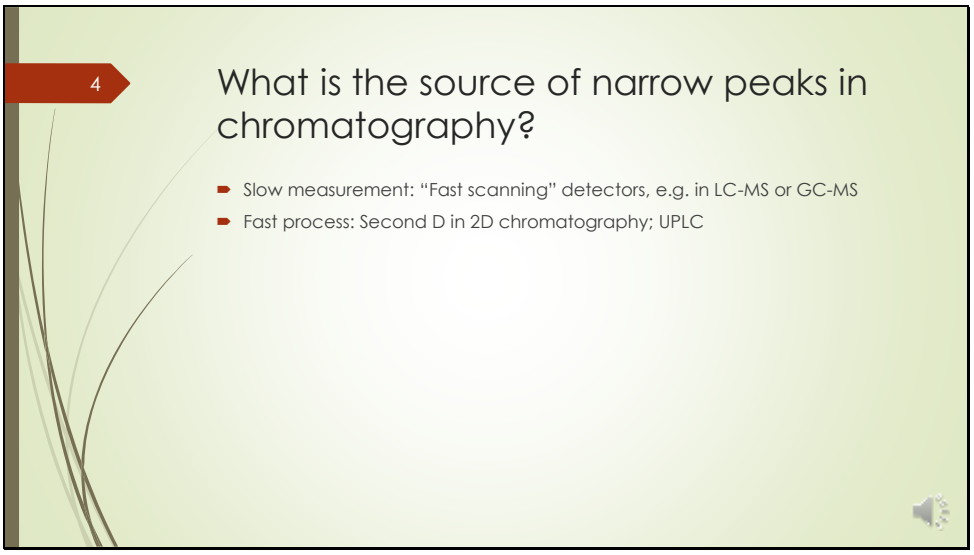
Good time of day! My name is Yuri Kalambet and I will tell you, how many points per peak are required for the purpose of correct measurement of area. This topic was very popular twenty to thirty years ago, but nevertheless today we feel enough courage to open the theme once again. This question concerns a lot of areas, but we will discuss it using chromatographic terminology, as authors are working in the field of chromatography data processing.



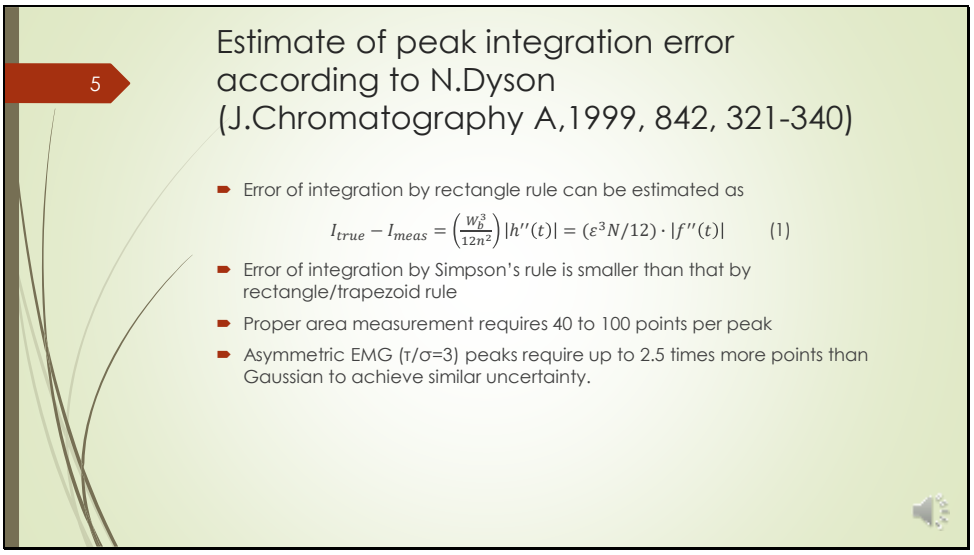
We will start from several notions from the measurement theory. Usually signal is measured (digitized) using fixed time slice. Fig.1 shows typical chromatographic signal (peak) profile with digitized points marked by dots. Ordinate axis stays for detector response, abscissa – for time. One of the main purposes of data processing – evaluation of the area under the original signal profile using measured points, the process of evaluation is called integration.
Fig.2 shows integration by rectangle rule. Every digitized point defines rectangle, area is evaluated as the sum of rectangle areas. Fig.3 illustrates trapezoid rule. In this method all digitized points are connected by broken line and area is estimated as an area under the broken line. In the case of Simpson rule, illustrated on figures 4 and 5, every three adjacent points are connected by parabola, and peak area is evaluated as area under connected parabolas. Please pay attention, that in the case of Simpson's rule we can get two different estimates of area, starting process of parabola construction from even or odd point.



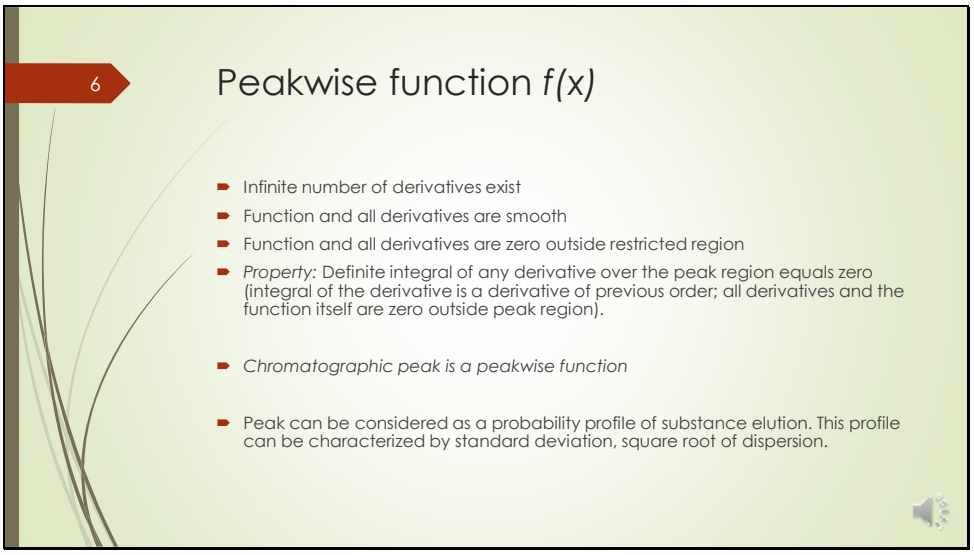
In the case of any rule area can be calculated using weighted sum of responses. Different integration rules are different in the weights only. Accurate calculations show, that all three discussed integration rules being applied properly will give the same answer, as their differences are located far from the top of the peak, and response at points of difference is zero.



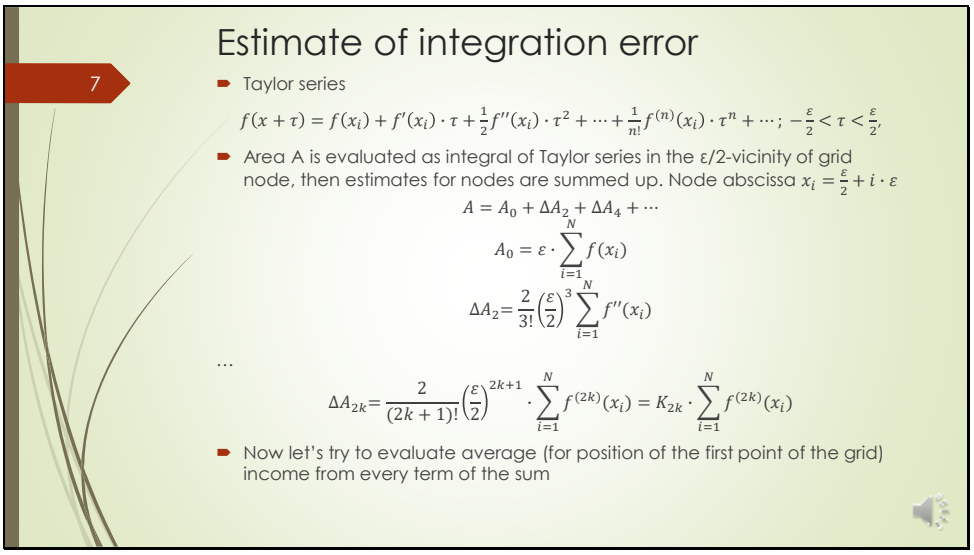
What is the reason of narrow peaks in chromatography and what is narrow?
Typical task of analysis in chromatography is measurement of peak area for the purpose of evaluation of concentration of some substance in the solution. Number of detector response measurements per second is defined by the physics of the detector. Any equipment has its limit, so when measurement time constant becomes comparable with the width of the peak, results of measurement can be considered as unreliable. We can use Gas Chromatography with quadrupole mass-spec detection as an example. This detector makes mass scan in time, and the wider is the mass range, the less frequently every individual mass is measured. The task of increased informativity (mass range) conflicts with the task of accurate measurement of area for the mass.



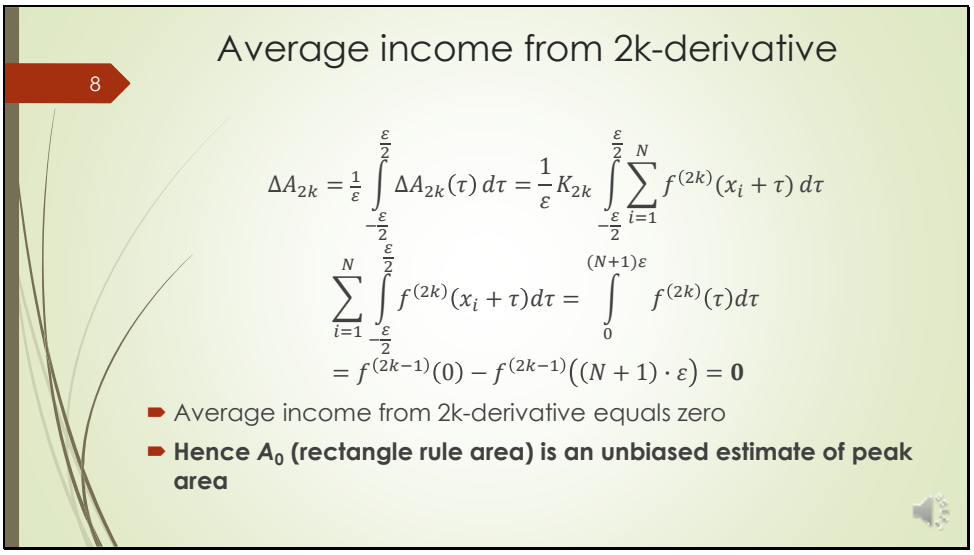
In mid-90's Norman Dyson wrote a very useful and popular book "Chromatographic Integration Methods". This book was published in several editions. Unfortunately, he also wrote a review article on measurement of very narrow chromatographic peaks. Some conclusions from this article are on the slide. All statements except the last one are misleading. As we will show later, the reason for that is wrong estimate of peak integration error using formula 1 taken from the textbooks. Main and most unpleasant conclusion is that for proper measurement of the peak one needs quite high measurement rate, not less than 40 points per peak. Despite the fact that there are other works with different estimates, we found none with clear answers.



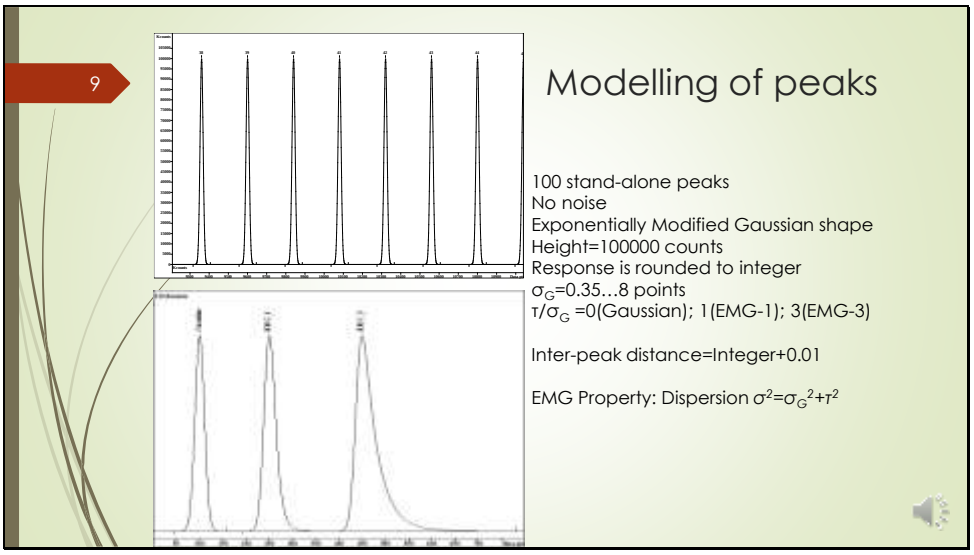
I will not go deep into details of computations – those interested may pause the demonstration and make a closer look at slides. We will only attract attention to the fact, that we do not compare data rate with peak width, as we do not know commonly accepted definition of the width. We compare data rate with standard deviation of the probability distribution function, corresponding to the peak. Typically this standard deviation is designated with Greek letter sigma. Depending on definition, peak width may be from 4 to 8 sigma.



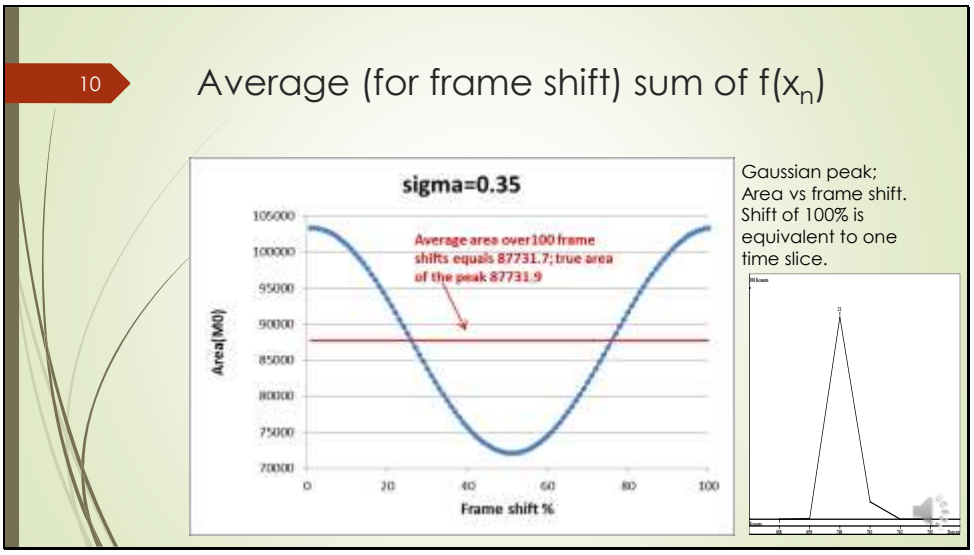
We made an attempt to evaluate integration error from the scratch,



And found out, that rectangle rule estimate of the area of stand-alone peak is unbiased. This statement means, that rectangle rule area being averaged over position of the first point of quantization grid, gives true peak area.

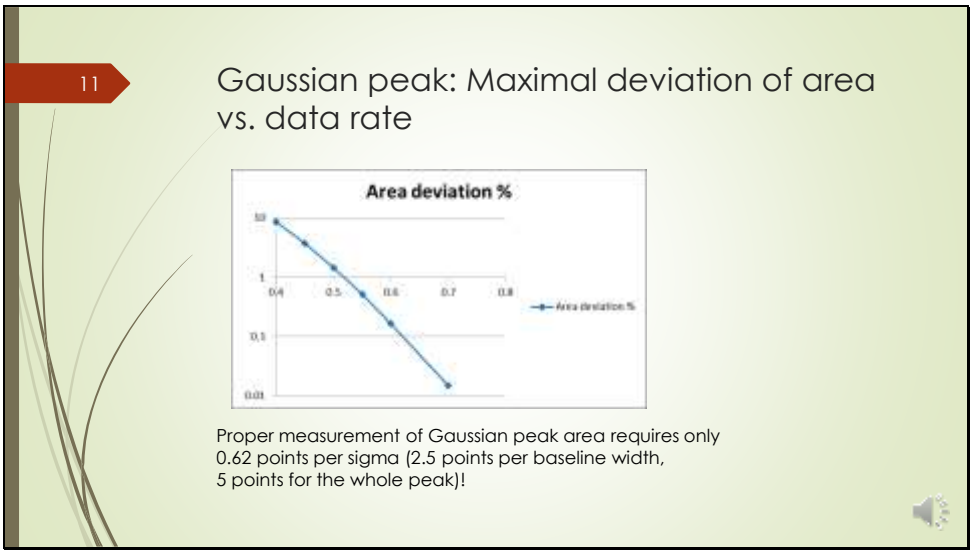


Theory can be easily verified by model experiment. We made 100 stand-alone peaks, every next one has its apex shifted integer number plus 1/100th of point with respect to previous one. Then we calculated their areas. Average area of all peaks should be equal to true area of original peak known from the peak's formula. The process was repeated three times for symmetric (Gaussian), slightly asymmetric (EMG-1) and highly asymmetric (EMG-3) peak shapes.

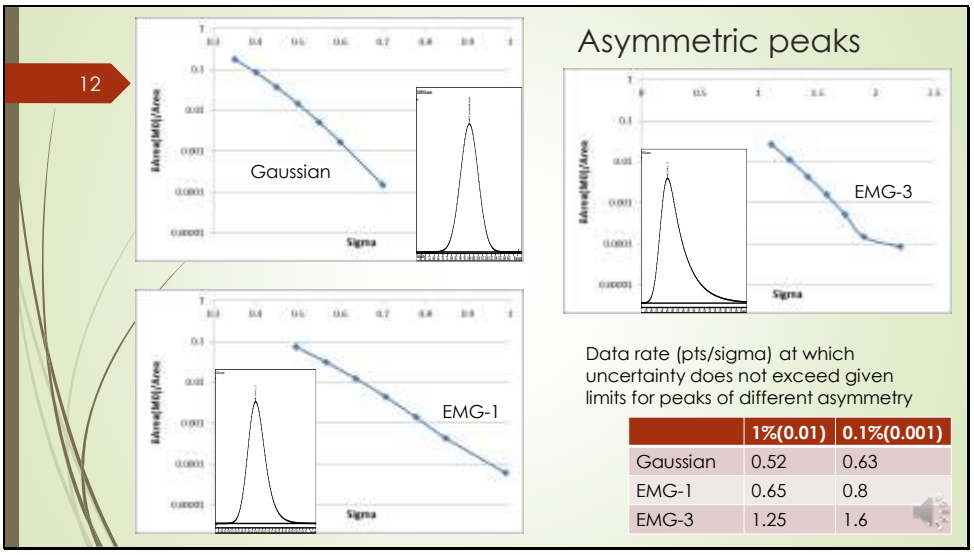


100 Gaussian peaks were generated, data rate equals 0.35 points per sigma. Shape of one of generated peaks is on the right.

The central graph of the slide shows dependence of rectangle rule peak area on frame shift percent. Hundred percent shift corresponds to one time slice. Brown line corresponds to true peak area. As expected, average area estimate for 100 peaks equals true area. The difference in 6^{th} digit is caused by discrete step and rounding errors. The graph resembles sine wave around true peak area level. Deviation of measured area from true area will never be bigger, than amplitude of this wave. So natural next step – do draw dependence of wave amplitude on data rate.

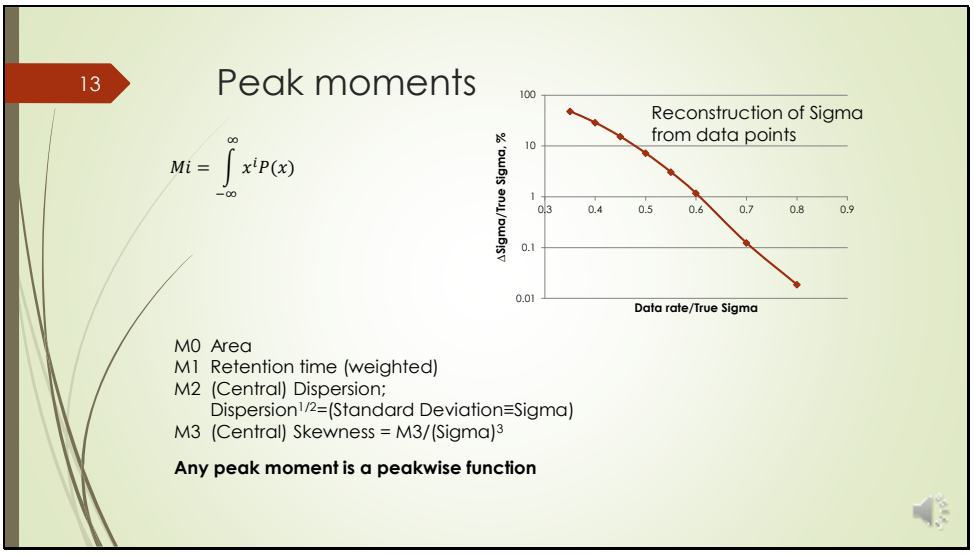


This slide illustrates how maximal deviation from true area depends on data sampling rate. As already mentioned, data rate is measured in points per sigma. As data rate increases, maximal deviation drops down very fast, ordinate axis is logarithmic. Most interesting error levels are one percent and one-tenth percent, these levels are often considered in analytical chemistry. Results are encouraging, compared with Dyson's estimate they allow to make 8 times faster analysis.

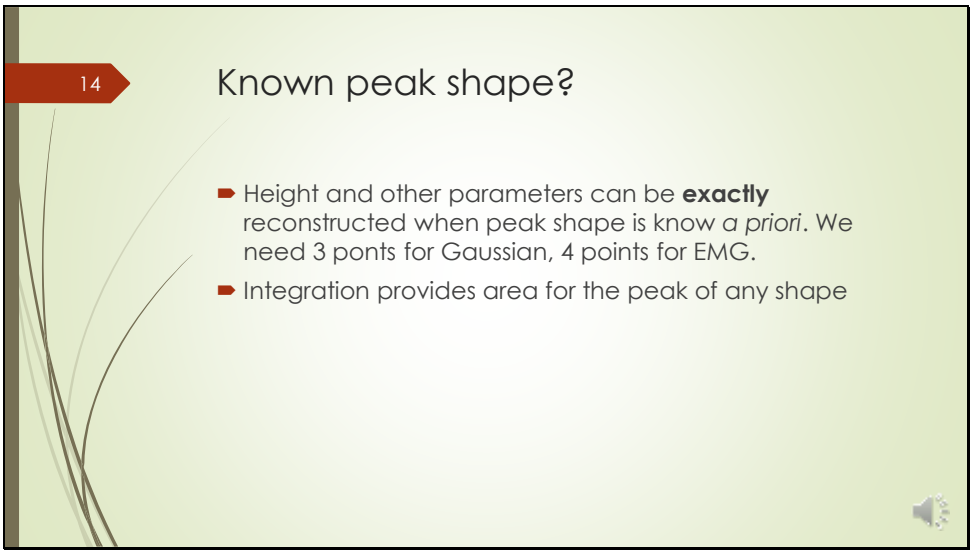


Pictures from this slide illustrate dependence of maximal area deviation on data rate for peaks of different shape. Shape of every peak is illustrated on the corresponding graph; critical frequencies are summarized in the table.

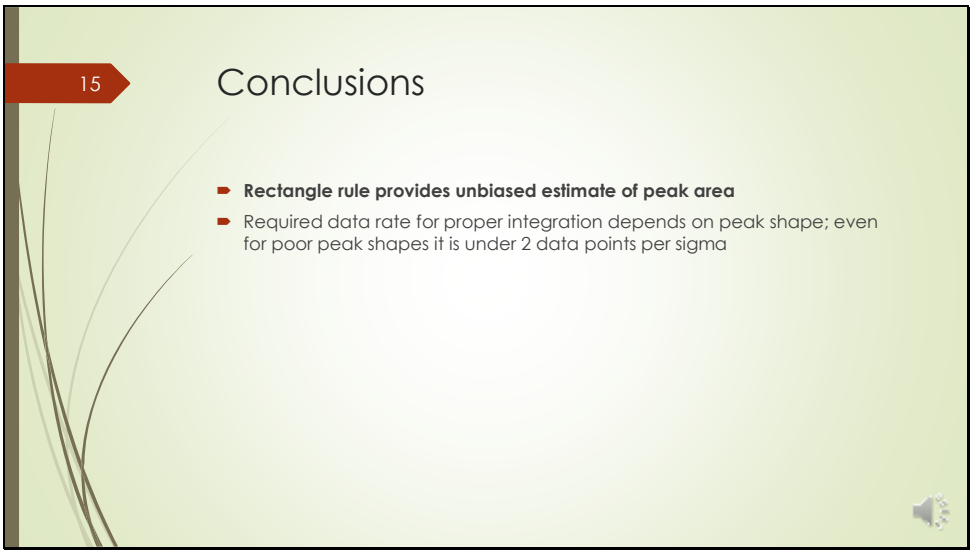
In the case of real analysis uncertainty, originated from data rate, are summed up with uncertainties, originating from noise. In most cases major part of uncertainty comes from noise. The table from this slide can be used for making a decision whether data rate uncertainty can be neglected compared to noise uncertainty. Derived rules indicate lower level of data rate and are a definition of “very narrow peak”. If we assume, that EMG-3 is the worst acceptable peak shape in chromatography, then “very narrow peak” is the peak, which has less than 1.5…2 measurements per sigma.



There are other peak properties that are calculated as an integral. For example, dispersion is a second central moment, it can be evaluated along with its square root (apparent sigma) using few data points.



The more we know about our system, the better. If we know peak shape, exact integration becomes possible starting from 3 points for Gaussian and 4 points for EMG. Conventional integration discussed in this presentation has such a benefit that it evaluates area of unknown peak shape.



Main conclusion of this talk is the fact that rectangle rule is an unbiased estimate of peak area. It's rather strange that, despite being trivial, this conclusion is not met in textbooks or other literature. The search for available publications on the subject should start from mid-19th century…

Thank you for your attention!