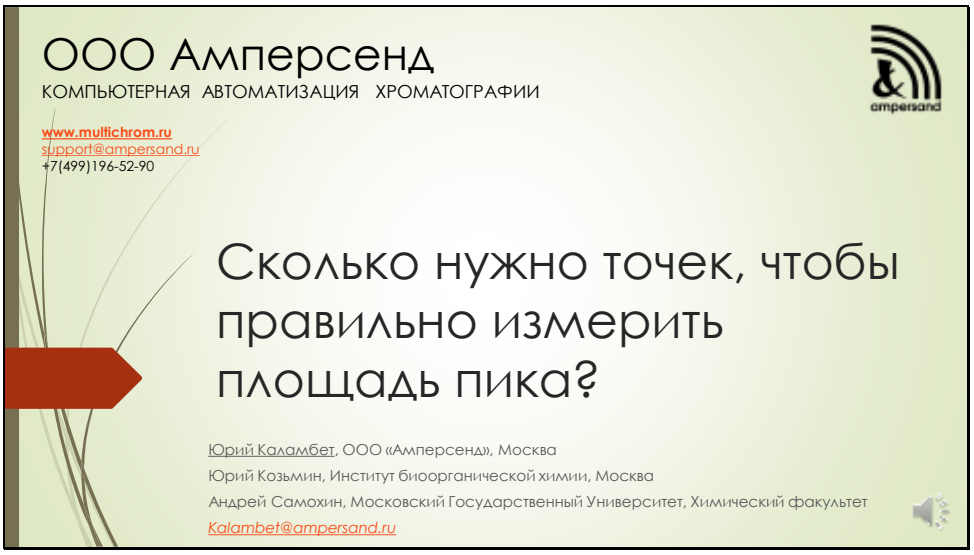
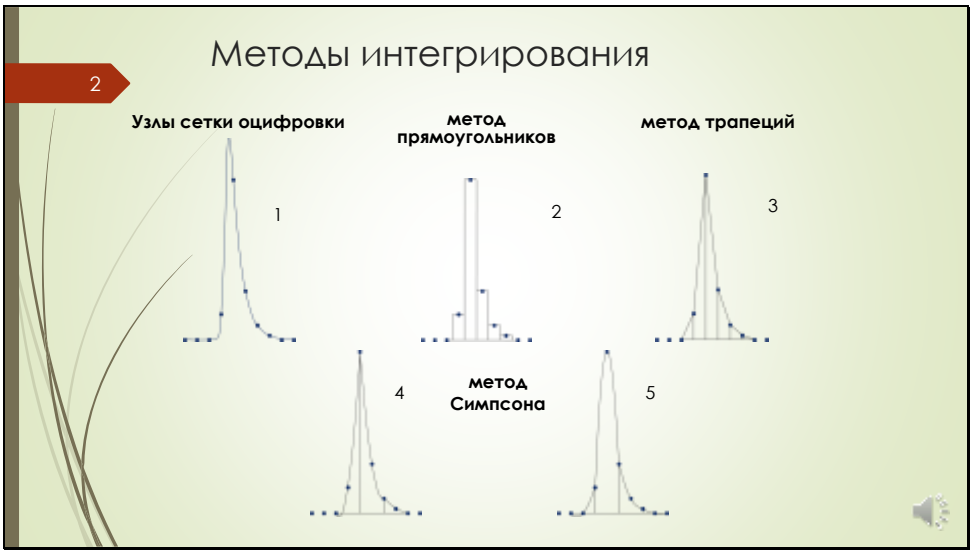
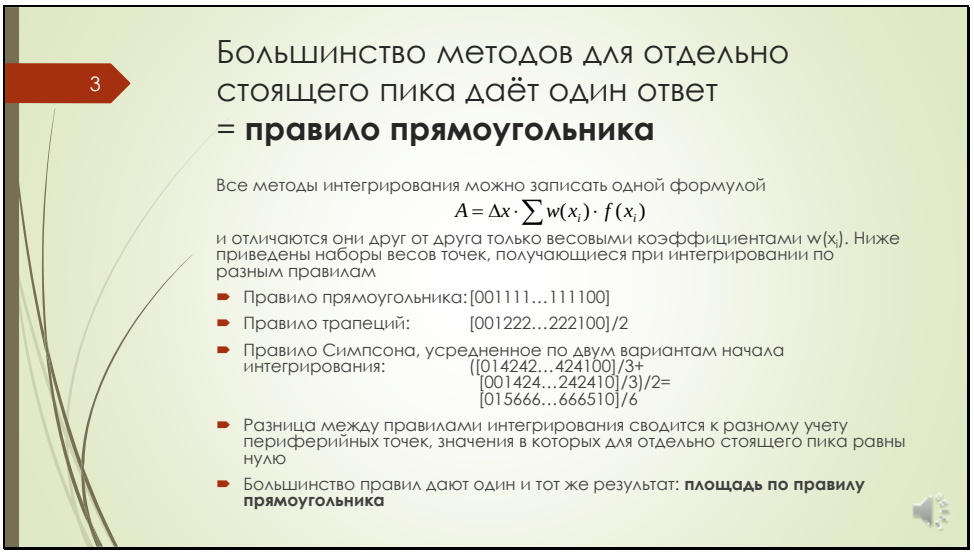


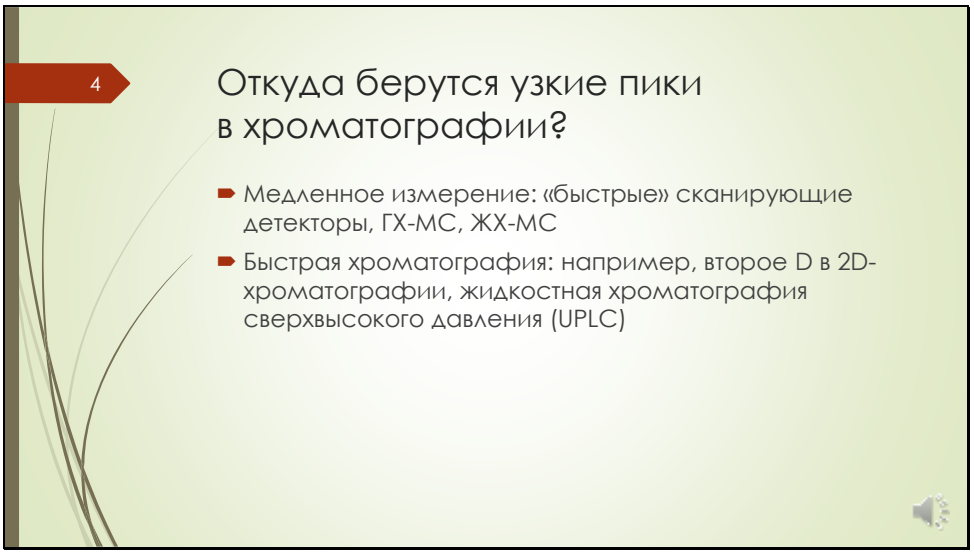
Здравствуйте! Меня зовут Юрий Каламбет, и сегодня расскажу вам о том, сколько точек нужно для правильного измерения площади пика. Этот вопрос касается далеко не только хроматографии, но разберем мы его на примере хроматографических пиков, поскольку авторы публикации работают в области хроматографии.



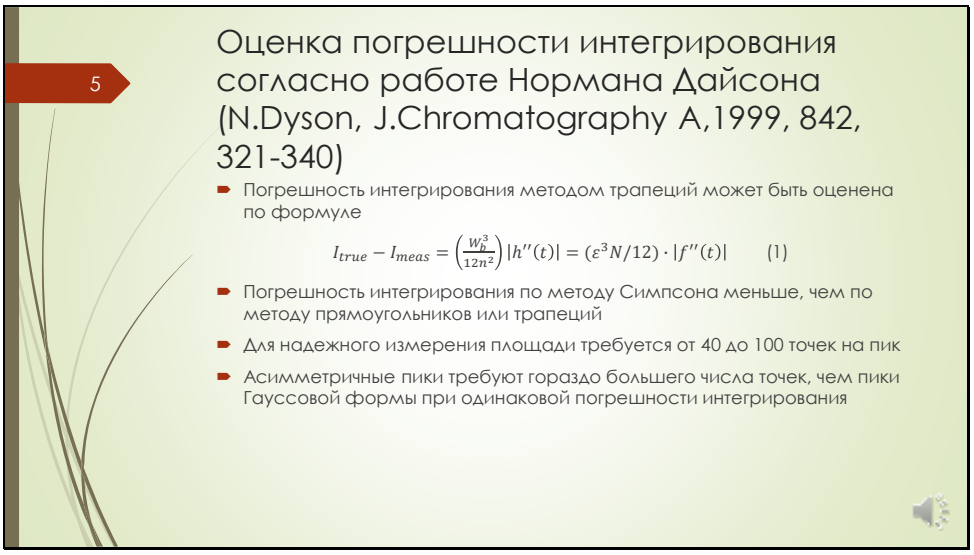
Начнем с некоторых понятий теории измерения. Обычно сигнал измеряется (переводится в цифровую форму) через равные промежутки времени. На рисунке 1 показана модель характерной формы хроматографического сигнала с отмеченными координатами оцифрованных точек. По ординате отложен отклик детектора, по абсциссе – время. Одна из основных задач обработки данных – оценить площадь под кривой по координатам измеренных точек, процесс оценки называется интегрированием.
На рисунке 2 показано интегрирование методом прямоугольников. Каждая оцифрованная точка задает прямоугольник, площадь равна сумме площадей прямоугольников. На рисунке 3 проиллюстрировано интегрирование методом трапеций. В этом методе все оцифрованные точки соединяются ломаной линией и оценкой площади считается площадь под ломаной. В методе Симпсона, проиллюстрированном на рисунках 4 и 5, каждые три соседние точки соединяются параболой, и площадь пика оценивается как площадь под соединенными параболами. Обратим внимание, что в методе Симпсона можно построить две разных оценки площади, начав конструирование парабол с четной или нечетной точки.



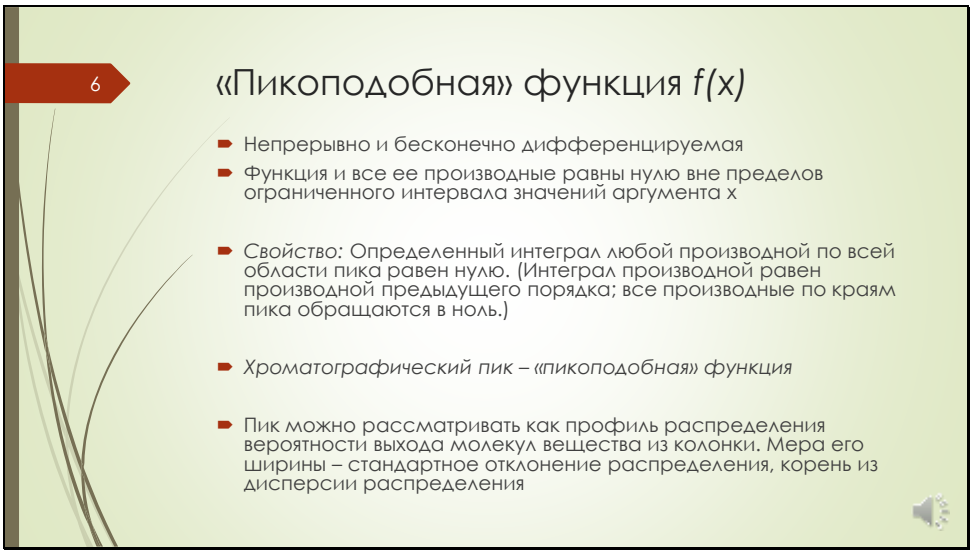
Площадь пика вычисляется по измерениям как сумма их ординат, взятая с некоторыми весами. Разные методы интегрирования отличаются набором весов. Аккуратные расчеты показывают, что три рассмотренных метода – прямоугольников, трапеций и Симпсона – для отдельно стоящих пиков дадут одинаковый результат, поскольку все отличия сконцентрированы на краях пика, где функция близка к нулю.



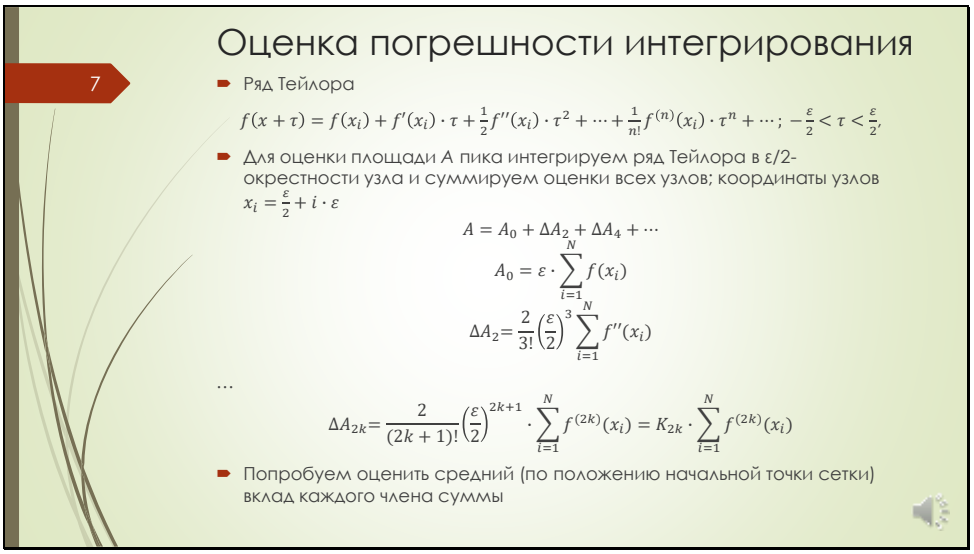
Собственно, откуда берутся узкие пики в хроматографии? И что значит узкие? Типовая задача в аналитической химии – измерение площади пика для оценки концентрации вещества в растворе. Число измерений в единицу времени определяется природой детектора. Какое бы ни использовалось оборудование, у него есть свой предел, так что при ширине пика, сравнимой с постоянной времени измерения, результаты, полученные с использованием прибора, становятся недостоверными. В качестве примера можно привести газовую хроматографию с квадрупольным масс-спектрометрическим детектором. Этот детектор делает развертку масс по времени, и чем шире диапазон масс, тем реже происходит измерение каждой индивидуальной массы. Задача повышения информативности (диапазон масс) вступает в конфликт с задачей точного измерения площади.



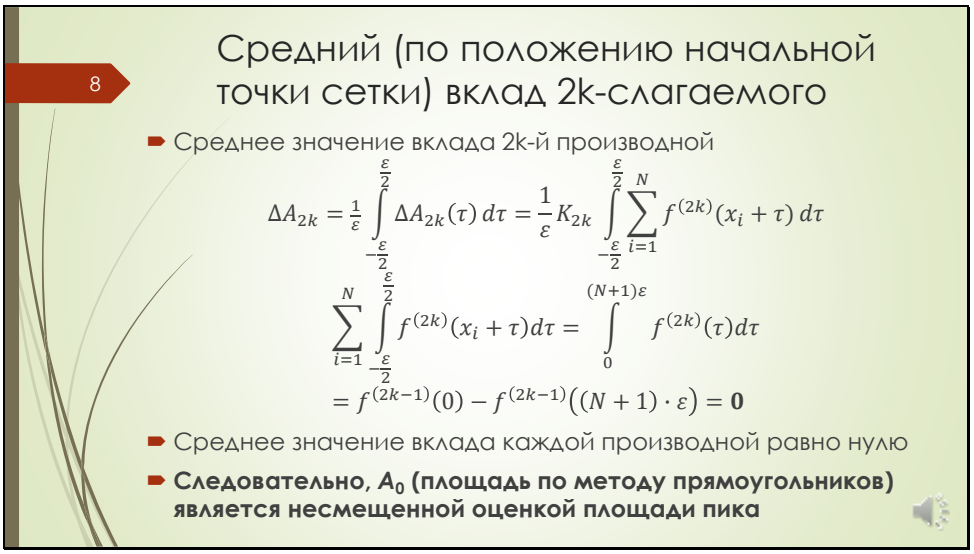
Норман Дайсон в середине 90-х годов написал очень полезную и популярную книгу «Методы интегрирования в хроматографии». Эта книга выдержала несколько переизданий. К сожалению, он также написал статью, посвященную интегрированию «очень узких» хроматографических пиков, некоторые утверждения из которой приведены на слайде. Все утверждения, кроме последнего, неверны. Самый главный и неприятный вывод, имеющийся в этой статье, состоит в том, что для правильного измерения площади нужно иметь очень высокую частоту оцифровки, не ниже 40 точек на пик.
Есть работы других авторов с отличающимися оценками, но четкого вывода оценки необходимой частоты измерения мы не нашли.



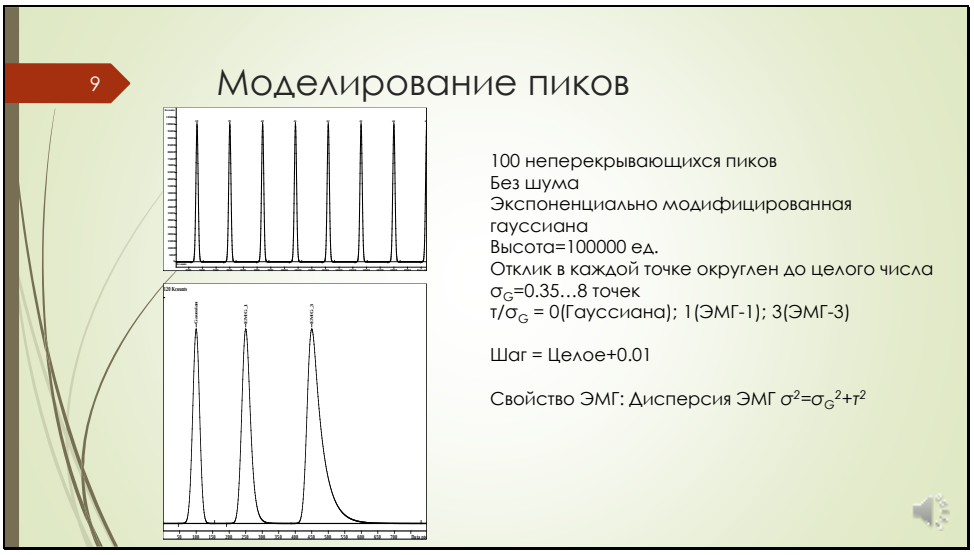
Не буду вдаваться в детали математических вычислений – желающие могут приостановить демонстрацию и почитать текст слайдов подольше. Обратим внимание только на то, что частоту измерений мы сравниваем не с шириной пика, поскольку у нас нет устоявшегося удобного определения ширины пика, а со стандартным отклонением соответствующего пику распределения вероятностей, которое обозначается греческой буквой сигма. В зависимости от определения, ширина пика составляет от 4 до 8 сигма.



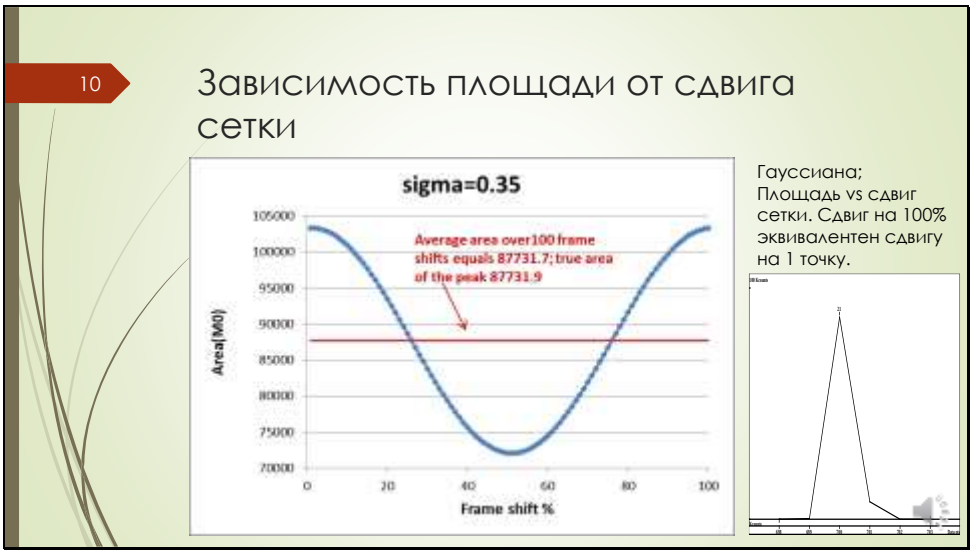
Мы попытались вывести правильную оценку погрешности измерения площади,



И выяснили, что оценка площади по методу прямоугольников является несмещенной. Это утверждение означает, что среднее значение площади, посчитанной методом прямоугольников, по всем возможным положениям первой точки сетки оцифровки, равно истинной площади пика.

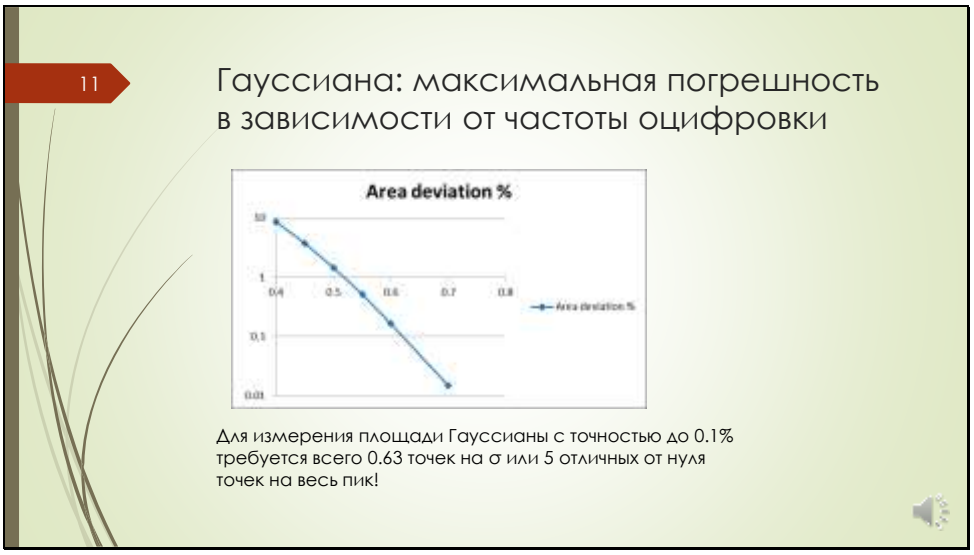


Теорию несложно проверить простым модельным экспериментом. Делаем 100 отдельно стоящих пиков, у каждого следующего вершина сдвинута относительно предыдущего на целое число плюс 1/100 точки, и меряем их площади. У нас должна получиться средняя площадь, равная истинной площади пика. Истинная площадь пика известна из его формулы. Процесс был повторен три раза для пиков разной формы: симметричного (Гауссиана), слабо асимметричного (ЭМГ-1) и сильно асимметричного (ЭМГ-3).

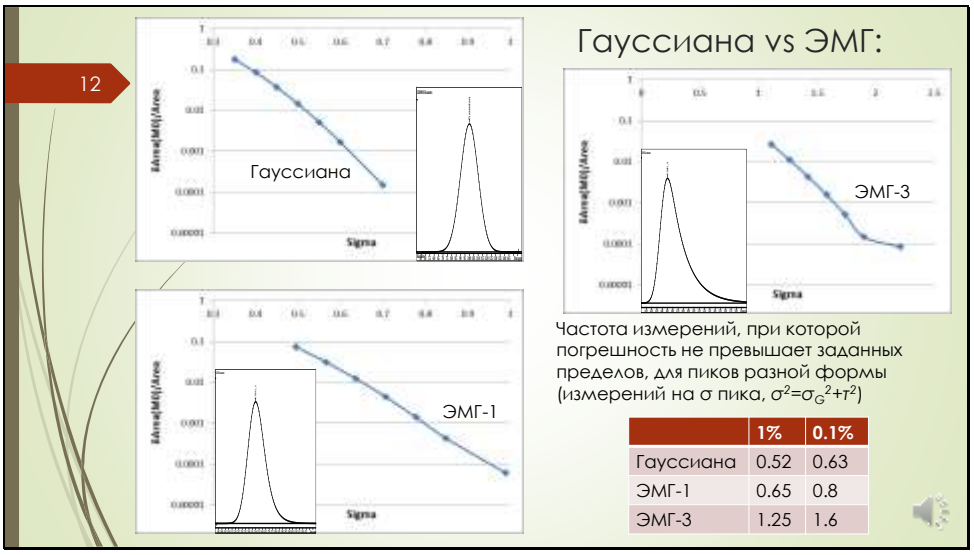


На графике приведена зависимость площади пика (по вертикали) от сдвига первой точки сетки оцифровки. Сдвиг на 100% соответствует сдвигу по времени на 1 точку. Генерировались пики Гауссовой формы, частота оцифровки равна 0.35 точек на сигму. Коричневая линия соответствует истинной площади пика. Форма одного из сгенерированных пиков приведена справа.

Как и ожидалось, среднее значение площади по 100 точкам равно ожидаемой площади пика. Разница в последнем (шестом) знаке вызвана округлением отклика до целых чисел. График отклонений от среднего очень напоминает синусоиду. Отклонение измеренной площади от истинной никогда не будет больше, чем амплитуда этой синусоиды. Поэтому естественный следующий шаг – построить зависимость амплитуды синусоиды от частоты измерений.



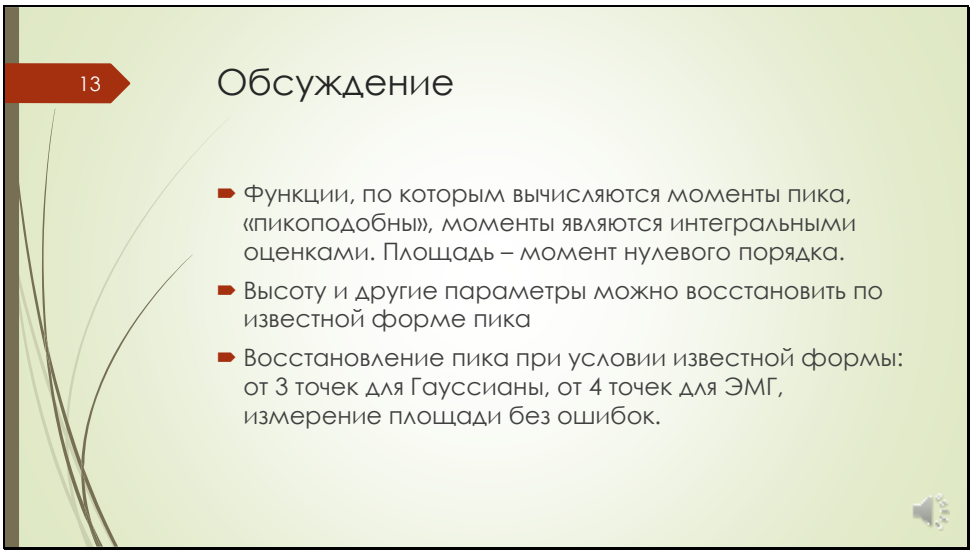
На этом слайде показана зависимость максимальной погрешности интегрирования (абсцисса) от частоты оцифровки, выраженной в измерениях на сигму. По мере роста числа точек максимальная погрешность падает очень быстро, график построен в логарифмическом масштабе. Наиболее интересные уровни погрешности соответствуют одному проценту площади и одной десятой процента. Результаты приятные, по сравнению с оценкой Дайсона позволяют поднять скорость анализа в 8 раз.



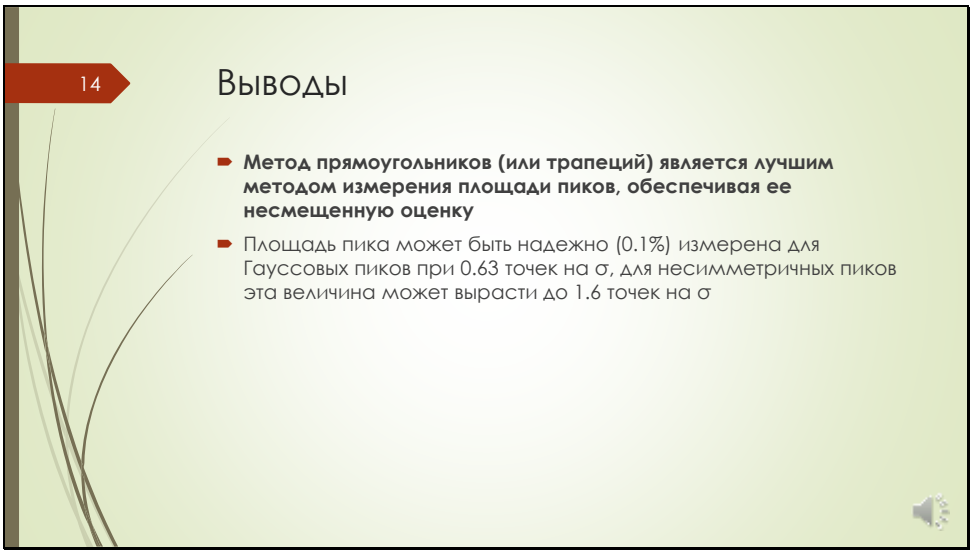
	1%	0.1%
Гауссиана	0.52	0.63
ЭМГ-1	0.65	0.8
ЭМГ-3	1.25	1.6

На рисунках этого слайда приведена зависимость максимальной погрешности от частоты оцифровки для пиков разной формы. Форма модельных пиков приведена на каждом рисунке; граничные частоты сведены в таблицу.

В реальном анализе вклад ошибок, происходящих от частоты оцифровки, суммируется с ошибками, причиной которых является шум. В большинстве случаев уровень погрешности измерения определяется именно шумом. Приведенную на этом слайде таблицу можно использовать для принятия решения о том, можно ли пренебречь погрешностью, связанной с частотой оцифровки по сравнению с погрешностью, порожденной электронным или химическим шумом. Выведенные закономерности указывают нижний предел частоты измерения и дают определение термина "очень узкий пик". Если принять форму сильно асимметричного пика ЭМГ-3 за худший из приемлемых вариантов пика в хроматографии, то очень узкий пик – это такой, у которого на сигму приходится менее полутора точек.



Кроме площади, есть другие свойства пика, вычисляемые интегрально. К примеру, дисперсия (квадрат сигмы) – второй центральный момент, и ее можно оценивать по небольшому числу точек аналогично площади.
Есть много других параметров пика, которые нужно оценивать. Интегрирование не может оценить, к примеру, высоту пика. Для определения высоты чрезвычайно полезно иметь представление о форме пика. Если такие представления есть, форма может быть восстановлена аппроксимацией. Для восстановления Гауссова пика потребуется три точки, ЭМГ-четыре.



Основной вывод этой презентации - то, что метод прямоугольников является несмещенной оценкой площади пика. Как ни странно, несмотря на его тривиальность, мы не смогли найти его в литературе. Искать можно начиная с конца 19 века…
Спасибо за внимание!