

Симуляция шумов детектора одномолекулярного секвенатора при моделировании процесса детектирования флуоресценции меченых нуклеотидов в реальном времени

© Е.К. Васильева, В.В. Манойлов, И.В. Чубинский-Надеждин

Институт аналитического приборостроения РАН,
198095 Санкт-Петербург, Россия
e-mail: virtkat@mail.ru

Поступило в Редакцию 2 марта 2026 г.

В окончательной редакции 8 июля 2026 г.

Принято к публикации 10 июля 2026 г.

Представлена разработка алгоритмов стохастического имитационного моделирования первичных данных для одномолекулярного SMRT (Single-Molecule Real-Time) секвенирования в реальном времени. Предложен подход к генерации виртуальных сигналов, учитывающий особенности оптической системы и процессов, протекающих в волноводах нулевой моды секвенатора. Проведено исследование шумов детектора секвенатора, включая темновые шумы, шум считывания, неоднородность фотоответа и фотонный дробовой шум. Получена функциональная зависимость среднеквадратичного отклонения шума от величины регистрируемого сигнала. Установлено, что теоретическая оценка шума соответствует физическому „нижнему пределу“, ниже которого значения реальных данных не опускаются. Сравнительный анализ модельных и реальных первичных данных показал высокую статистическую сходимость среднеквадратичного отклонения шумов. Отмечено, что оптимальное соотношение сигнал/шум достигается при длительности импульсов, близкой ко времени экспозиции кадра. Полученные результаты позволили создать и оптимизировать алгоритм генерации реалистичного шума для программного симулятора SMRT-секвенатора.

Ключевые слова: симулятор последовательности, симулятор шумов, шум видеокамеры, одномолекулярное секвенирование, волновод нулевой моды.

DOI: 10.61011/JTF.2026.11.63779.57-26

Введение

Одномолекулярное секвенирование фрагментов молекул ДНК, иммобилизованных на дне специализированных наноструктур — волноводов нулевой моды (ВНМ, ZMW), основано на детектировании флуоресцентного излучения от меченых нуклеотидов, которые комплементарно встраиваются в реальном времени в одиночные цепочки ДНК при помощи полимеразы. Это интересный, сравнительно редко используемый метод третьего поколения, в англоязычной литературе называемый Single-Molecule Real-Time sequencing (SMRT sequencing) [1–3]. Секвенирование по данной методике может обеспечить чтение длинных, до нескольких тысяч базовых пар, последовательностей (ридов).

Одной из проблем начального этапа разработки одномолекулярного секвенатора, работающего по методу SMRT, является отсутствие достаточного количества качественных первичных данных, при помощи которых можно было бы начать работу над созданием алгоритмов обнаружения баз (basecalling). Для разработки действительно эффективных программных средств обнаружения требуются большие репрезентативные наборы данных, отражающие такие характеристики регистрируемых полезных сигналов, как амплитуда импульса, соответствующего событию встраивания нуклеотида, от-

ношение сигнал/шум в разных условиях детектирования, длительность события встраивания, частота реализации событий встраивания в секунду, наличие паразитных ложных сигналов, мешающих обнаружению и т. п. Такие данные могут быть получены только в течение длительного времени работы прибора в разных условиях детектирования и, что самое важное, для этого необходим отлаженный, проверенный метод самого секвенирования, осуществляемого по данной методике. При разработке прибора „с нуля“ получение реальных первичных данных, заслуживающих доверия, представляет определенные сложности, избежать которых можно при помощи стохастического имитационного моделирования всего процесса детектирования.

На сегодняшний день известны программные симуляторы буквенных последовательностей [4–7], применяемые для разработки и отладки алгоритмов вторичной обработки данных NGS секвенирования и нанопорового секвенирования [8], но отсутствуют в свободном доступе симуляторы первичных данных для разработки алгоритмов basecalling для метода SMRT. Восполнить этот пробел можно с помощью разработки алгоритмов виртуального секвенирования, учитывающих особенности нового прибора — его детектора, оптической системы, реакционной ячейки и процессов, протекающих в ВНМ во время встраивания меченых нуклеотидов.

При формировании модельного цифрового сигнала, имитирующего процесс детектирования оптического сигнала, одной из ключевых задач является построение реалистичной модели шумов детектора. В настоящей работе исследованы все виды шумов детектора секвенатора, получена функциональная зависимость величины среднеквадратического отклонения (СКО, SD) шумов от величины регистрируемого сигнала и проведены сравнительные исследования параметров реального и модельного шума.

1. Метод детектирования флуоресценции и формирование первичных данных

Одномолекулярное секвенирование по методу SMRT проводится в специальных реакционных ячейках, содержащих большое количество ВНМ, где протекают реакции секвенирования для одиночных фрагментов ДНК. Массив ВНМ представляет собой наноструктуры цилиндрической формы, полученные путем травления в тонкой пленке алюминия, нанесенного на кварцевую подложку реакционной ячейки. Глубина и ширина каждого цилиндра не превышает 100–120 нм. Массив из нескольких тысяч ВНМ расположен внутри ячейки таким образом, что сверху в него может подаваться раствор с реагентами через специальный резервуар, а снизу кварцевая подложка засвечивается лазерами одновременно на длинах волн 532 и 639 нм. Сверху ячейка может быть подсвечена светодиодом для визуализации ВНМ на просвет, что обычно используется в дальнейшем для их быстрого поиска. Реакционная ячейка пассивируется¹ и биотинилируется². Одиночные молекулы полимеразы предварительно закрепляются на дне ВНМ при помощи стрептавида. В ячейку добавляют фрагментированную одноцепочечную ДНК в такой концентрации, чтобы в каждом ВНМ находилось не более одного фрагмента. Реально этого можно добиться только для части ВНМ. Структуры, куда попало больше или меньше одного фрагмента ДНК, не должны учитываться при дальнейшей обработке данных. В отечественной литературе общие подходы к реализации такого метода рассмотрены в [9,10].

Рис. 1 упрощенно иллюстрирует процесс встраивания меченных нуклеотидов в ДНК в одном ВНМ. В ходе секвенирования в реакционную ячейку с ВНМ 1 подается раствор, содержащий меченные нуклеотиды 2. Нуклеотиды по одному комплементарно встраиваются в одиночную, замкнутую в кольцо цепь ДНК 3 при помощи полимеразы 4, иммобилизованной на дне

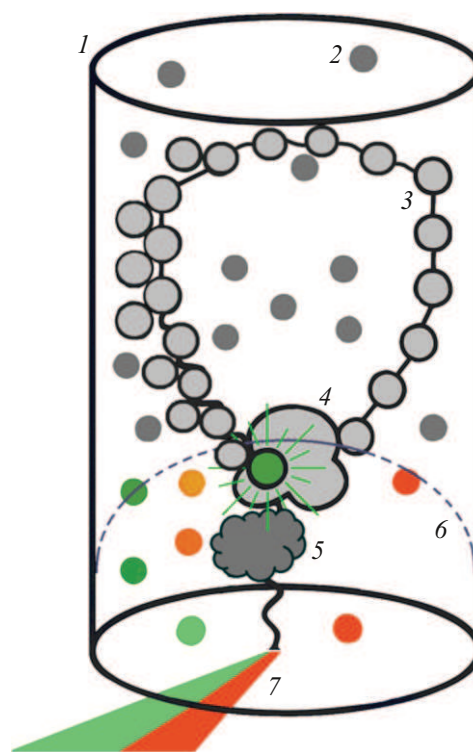


Рис. 1. Схематичное представление волновода нулевой моды с иммобилизованной на дне полимеразой и фрагментом ДНК в процессе встраивания меченного нуклеотида: 1 — ВНМ; 2 — раствор с мечеными нуклеотидами; 3 — одиночная замкнутая цепь ДНК; 4 — молекула полимеразы; 5 — молекула стрептавида; 6 — зона эффективной засветки лазером внутри ВНМ; 7 — пучок лазерного излучения от двух лазеров 532 и 639 нм.

ВНМ при помощи молекулы стрептавида 5. В ходе встраивания нуклеотид с флуоресцентной меткой на короткое время, от десятков до сотен мс, попадает в зону эффективной засветки 6 (зона вблизи дна ВНМ, выделенная на рис. 1 штриховой линией) лазерами 7, и в это время излучает свет на соответствующей длине волны в одном из четырех спектральных диапазонов (по числу нуклеотидов). После встраивания нуклеотид выходит из зоны эффективной засветки и на его место приходит следующий. При однократном чтении кольцевого фрагмента ДНК качество обнаружения баз не превышает 90 %, при многократном чтении кольца и последующем усреднении данных, качество прочтений может быть повышено [11,12]. Рид, полученный при многократном чтении, носит название полимеразного.

Рис. 2 иллюстрирует процесс детектирования. Исходящий от нуклеотидов свет 1 попадает на дисперсионную призму 2, где преломляется в соответствии с длинами волн и образует вытянутое световое пятно в области детектирования на светочувствительной матрице видеокамеры 3, которое передается в компьютер в виде изображения 4. Для одного ВНМ такое пятно занимает четыре смежные спектральные области, соответствующие месту падения на матрицу видеокамеры света от

¹ Пассивирование алюминия — технологический процесс обработки реакционной ячейки, направленный на снижение вероятности неспецифического связывания белковых молекул со стенками волновода нулевой моды.

² Биотинилирование кварцевой поверхности — процесс, направленный на создание слоя биотина на дне волновода нулевой моды, на который впоследствии будет прикреплена молекула стрептавида.

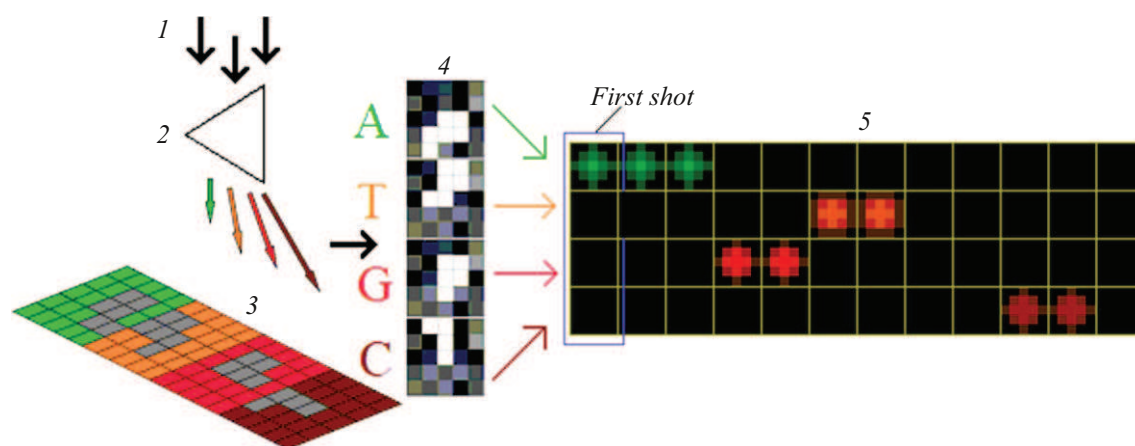


Рис. 2. Схематичное представление процесса детектирования оптического сигнала детектором секвенатора: 1 — свет, выходящий из ВНМ; 2 — дисперсионная призма; 3 — светочувствительная матрица видеокамеры; 4 — сформированное изображение; 5 — схема формирования фрагмента данных типа муви (movie) для свечения четырех разных нуклеотидов.

четырёх нуклеотидов А, Т, G и С после разложения в спектр 5.

В течение времени экспозиции регистрируется кадр, который за заданное, относительно короткое время, меньшее, чем время встраивания нуклеотида полимеразой, фиксирует картину свечения дна ячейки. Многократно зарегистрированные кадры дают нам полное изображение процесса секвенирования во времени. Выстроенные кадр за кадром в соответствующей временной последовательности изображения для одного ВНМ показывают процесс встраивания нуклеотидов для фрагмента ДНК, находящегося конкретно в этой наноструктуре, образуя муви (формат movie) для одного ВНМ. На рис. 2 справа представлена модельная ситуация 5, когда в выбранном ВНМ в зеленом канале длительность свечения соответствует трем кадрам, в красном, оранжевом и темно-красном (показан коричневым цветом) каналах — двум кадрам, между свечением в оранжевом и темно-красном каналах есть пауза в два кадра. Между разными спектральными каналами может присутствовать перекрестная флуоресценция (spectral crosstalk), не показанная здесь, как менее яркое свечение в соседнем канале.

Время экспозиции выбирается таким образом, чтобы для каждого ВНМ свечение одного нуклеотида регистрировалось в течение нескольких кадров, затем, пока не начнется встраивание следующего нуклеотида, регистрируется фоновый сигнал ячейки, после чего снова идет регистрация светящегося нуклеотида. Таким образом формируются последовательности изображений свечения каждого ВНМ одновременно в четырех спектральных каналах. В дальнейшем при обработке данных должно быть установлено, в каком из четырех спектральных каналов зарегистрировано свечение для каждого кадра.

На рис. 3 флуоресцентная эмиссия нуклеотида 1 формирует на светочувствительной матрице камеры элек-

тронный сигнал, при этом уровень яркости изображения в одном пикселе соответствует числу фотоэлектронов, накопленных в соответствующем сенсоре за время однократной экспозиции. Перед началом секвенирования изображение ячейки предварительно размечается в соответствии с местом расположения каждого ВНМ, чтобы впоследствии анализировать сигналы яркости в заранее локализованных областях. Сумма яркостей всех пикселей в области интегрирования сигнала 2 (область светового пятна для каждого из четырех каналов) формирует интегральный сигнал яркости, который становится единичным отсчетом трейса (формат trace) 3 в одном спектральном канале.

Таким образом, регистрация флуоресценции, излучаемой из большого числа ВНМ в ходе встраивания меченных нуклеотидов, осуществляется при помощи sCMOS³ видеокамеры, на сенсорах которой в течение времени экспозиции кадра формируются электронные сигналы, эквивалентные количеству света, полученному от каждого ВНМ и характеризующие процесс встраивания во времени. Изображение каждого кадра передается в компьютер, где происходит обнаружение ВНМ и определение сигналов яркости для каждого из них. Область интегрирования сигнала яркости на изображении обычно представляет собой квадрат 3×3 пикселя для каждого ВНМ, но может быть меньше или больше в зависимости от настройки оптической системы секвенатора. Для учета этих особенностей при моделировании обеспечивается возможность выбора размера области интегрирования от 2×2 до 5×5 пикселей. Суммарный сигнал яркости, полученный в области интегрирования, представляет собой отсчет дискретного сигнала. Множество таких отсчетов формируют трейс — зависимость яркости от времени регистрации сигнала. На рис. 4 показаны данные в формате трейс 1, где каждый трейс

³ sCMOS — scientific Complementary Metal–Oxide–Semiconductor, или КМОП комплементарная структура металл–оксид–полупроводник.

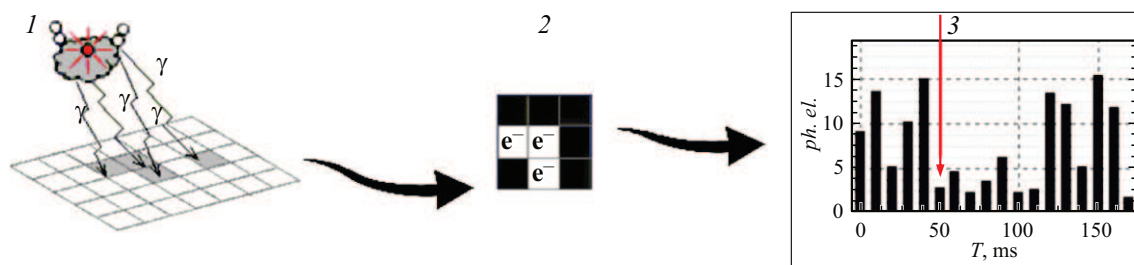


Рис. 3. Схематичное представление формирования отсчетов трейса: 1 — формирование электронного сигнала в пикселях светочувствительной матрицы видеокамеры; 2 — область интегрирования сигнала яркости на изображении; 3 — формирование данных типа трейс (trace).

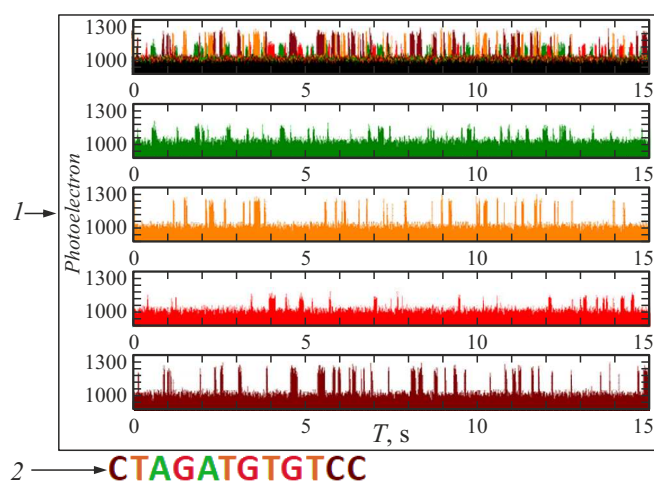


Рис. 4. Трейсы и буквенная последовательность нуклеотидов: 1 — общий трейс для всех каналов регистрации сверху и четыре трейса для каждого канала в отдельности; 2 — буквенная последовательность нуклеотидов.

представлен соответствующим своему спектральному каналу цветом. Суммарный трейс формируется для всех четырех каналов регистрации и совмещает графики четырех цветов. По трейсам осуществляется обнаружение и распознавание баз — basecalling, результатом которого является формирование на выходе прибора fasta-файла, содержащего последовательность буквенных обозначений нуклеотидов 2, которая называется ридом. Данные в формате муви и трейс являются первичными данными секвенатора.

2. Методика исследования шумов детектора

Шумы видеокамеры, выбранной в качестве детектора секвенатора, складываются из шумов, зависящих от уровня освещенности светочувствительной матрицы, и шумов, не зависящих от него.

К зависящим относятся два вида шума: PRNU (photo response non-uniformity) — неоднородность фотоответа

и дробовой шум. PRNU обусловлен неоднородностью реагирования на свет различных сенсоров матрицы и в среднем, как указывает производитель в документации к камере, не превышает 0.3 % от величины сигнала, накопленного в единичном сенсоре. Дробовой фотонный шум возникает вследствие флуктуации количества фотонов, упавших в единицу времени на сенсор камеры, и может быть приблизительно оценен, как корень квадратный из величины сигнала, полученного в этом сенсоре.

К независящим от уровня освещенности шумам относятся также два вида — DSNU (dark signal non-uniformity) и readout (шум считывания). Первый характеризует темновые токи в сенсорах камеры и не превышает 0.2 photoelectron на сенсор в данной камере. Readout-шум возникает при считывании сигнала через схему, при этом каждый сенсор в sCMOS-камере соответствует своей схеме считывания, поэтому шумы для разных сенсоров матрицы незначительно отличаются [13]. В документации к камере указана величина шума считывания 1.2 photoelectron. Одной из задач работы стала проверка соответствия данных, указанных в документации, реальным возможностям камеры.

Программное формирование реалистичных шумов, соответствующих по своим статистическим характеристикам реальным шумам детектора, является принципиально важным этапом для качественного моделирования первичных данных секвенирования. Одной из ключевых задач является разработка общей модели шума, учитывающей совокупность всех его перечисленных видов для выбранного режима работы камеры детектора. Для построения такой модели были зарегистрированы и исследованы серии изображений рабочей поверхности реакционной ячейки секвенатора.

Исследование проведено для sCMOS-камеры секвенатора Dhyana 400BSI V3 Tucsen Photonics Co Ltd для режима регистрации High Sensitivity с оцифровкой 12 бит на пиксель и смещением нулевого уровня на 100 единиц. Для засветки рабочей поверхности реакционной ячейки использованы встроенные лазеры секвенатора: зеленый (532 nm) с мощностью 120 mW и красный (639 nm) с мощностью 75 mW. Для формирования импульсов света заданной длительности, имитирующих излучение

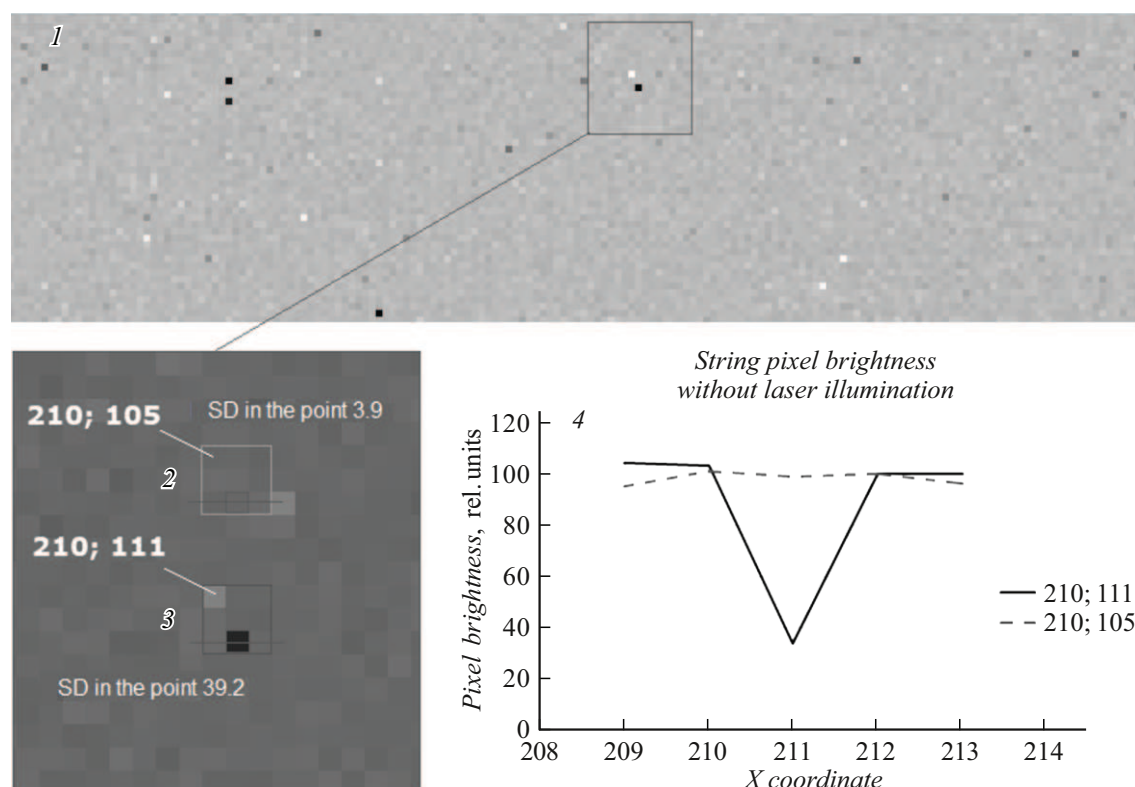


Рис. 5. Фрагмент изображения темного кадра, примеры областей интегрирования 3×3 пикселя и профилей сигнала яркости для „хороших“ и „плохих“ пикселей: 1 — фрагмент изображения темного кадра, на котором сверху в центре выделена квадратная область, рассмотренная ниже в увеличенном масштабе; 2 — область интегрирования сигнала яркости 3×3 пикселя, содержащая „хорошие“ пиксели, с координатой левого-верхнего пикселя {210; 105}; 3 — область интегрирования сигнала яркости 3×3 пикселя, содержащая „плохие“ пиксели, с координатой левого-верхнего пикселя {210; 111}; 4 — профили яркости фрагментов изображений в нижней строке для областей 2 и 3, построены по пяти пикселям.

меченных нуклеотидов, использован встроенный светодиод. Время экспозиции кадров в разных экспериментах составило: 0.5, 5, 10, 20, 40, 60, 80, 100, 150, 200 и 250 ms.

При наличии возбуждающего лазерного излучения рабочая поверхность реакционной ячейки сама излучает свет, что позволяет исследовать шумы, зависящие от уровня освещенности (PRNU и дробовой шум). При фиксированной мощности лазерного излучения время экспозиции кадров варьировалось, обеспечивая формирование сигналов яркости изображения ячейки во всем диапазоне для выбранного режима — от 100 до 4096 photoelectron/pixel. При исследовании шумов, не зависящих от уровня освещенности (DSNU и шум считывания), были сделаны серии изображений, полученные в таких же условиях, но при перекрытом входящем световом потоке.

2.1. Исследование шумов, не зависящих от уровня освещенности

Среднее значение сигнала интенсивности в одном пикселе для темного кадра составляет 100 единиц, эта

величина соответствует отсутствию внешнего оптического сигнала и формируется вследствие принудительного сдвига нулевого уровня в видеокамере. На рис. 5 приведено изображение части темного кадра 1, где для выделенного фрагмента в увеличенном масштабе показаны две области интегрирования сигнала размером 3×3 пикселя. В одной из них только „хорошие“ пиксели 2, вторая содержит „плохой“ пиксель с большим СКО шума 3. Рядом справа показаны профили сигнала 4 для выделенных на изображении в этих областях пикселей. Очевидно, что попадание одного или нескольких пикселей с более высоким СКО в область интегрирования сигнала интенсивности может увеличить СКО шума на трейсе и ухудшить обнаружение полезного сигнала флуоресценции от ВМ.

В экспериментах по определению СКО шумов для темных кадров были получены значения СКО сигналов яркости для темного шума в пикселе для всех пикселей камеры, гистограмма распределения значений СКО показана на рис. 6. Эта гистограмма характеризует независимые от уровня освещенности шумы детектора. Максимум распределения соответствует величине 2.7 photoelectron (округленно 3 photoelectron), 89 %

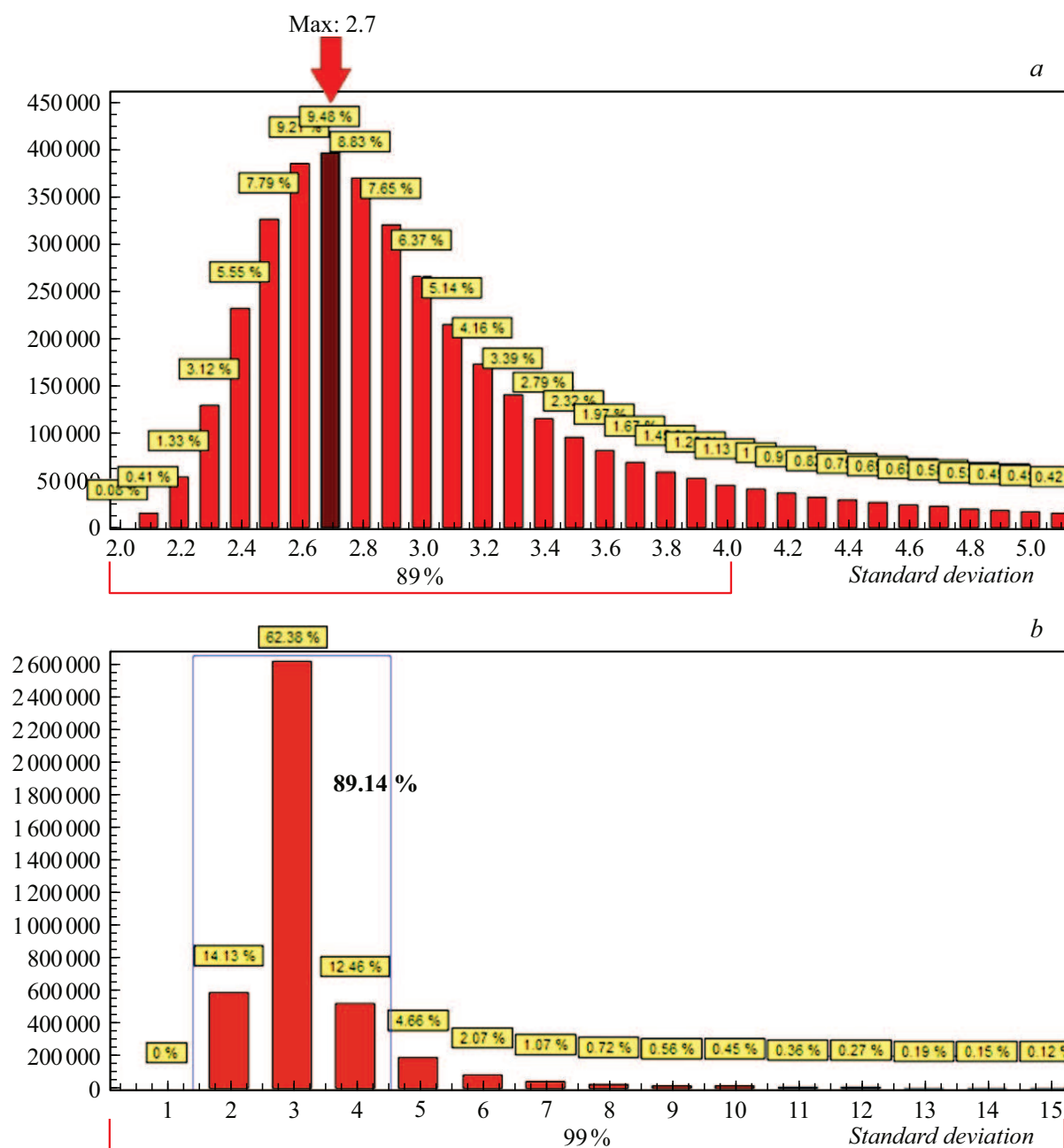


Рис. 6. Гистограмма распределения СКО темновых шумов (СКО шума в пикселе для всех пикселей кадра); *a* и *b* соответствуют разным масштабам представления интервалов (бинов) для столбцов.

пикселей имеют СКО темнового шума, не превышающее 4 photoelectron/pixel. При формировании первичных данных секвенатора в формате трейсов (зависимость величины сигнала от времени регистрации) происходит интегрирование сигнала яркости на изображении в области пятна 3×3 пикселя, что характеризует изменение яркости внутри ВМ. Таким образом, для области 3×3 пикселя уровень, соответствующий отсутствию сигнала (нулевой уровень), составит суммарно по 9 пикселям 900 единиц. СКО шума для такой квадратной области, согласно правилу сложения по-

грешностей, составит 9 photoelectron для большинства „хороших“ пикселей, т.е. для пикселей с СКО шума 3 photoelectron

2.2. Исследование шумов, зависящих от уровня освещенности

СКО шумов детектора, обусловленное влиянием PRNU и дробового шума, определяется величиной зарегистрированного сигнала и увеличивается пропорционально росту сигнала. Под термином „СКО шума“ трей-

са в дальнейшем будем понимать среднее квадратичное отклонение величины зарегистрированного сигнала от среднего уровня этого сигнала на трейсе, т.е. ту флуктуацию сигнала, которая мешает отличать импульс, соответствующий флуоресцентной вспышке, от фонового свечения.

Для получения функциональной зависимости величины СКО шума от величины регистрируемого сигнала были обработаны данные измерений свечения рабочей поверхности ячейки при засветке ее лазерами. Меняя экспозицию кадра, мы получали все более яркий фон на изображении ячейки до тех пор пока не зарегистрировали полный динамический диапазон значений яркости на пиксель, которые может выдавать камера в выбранном режиме регистрации. В этом же эксперименте в течение всего времени регистрации через одинаковые промежутки времени автоматически включался светодиод, обеспечивая вспышки света одинаковой интенсивности длительностью 50 ms, которые подсвечивали все ВНМ сверху, визуализируя их (рис. 7). Светодиод, вспыхивая, формирует дополнительный к фоновому свечению поверхности ячейки периодический оптический сигнал в каждом ВНМ. Область интегрирования сигнала была выбрана в зоне максимальной яркости свечения ячейки в центре кадра.

Зарегистрированные изображения были объединены в один набор и обработаны. Таким образом, были получены уровни яркости, увеличивающейся ступенчато пропорционально времени экспозиции (рис. 8). Средний уровень для каждого фрагмента трейса с заданной экспозицией соответствует интегральному сигналу яркости изображения поверхности ячейки в области 3×3 пикселя, а импульсы эквивалентны оптическому сигналу, зарегистрированному в момент вспышек светодиода. Над „ступеньками“ графика цифрами указано соответствующее время экспозиции в ms. Для каждой „ступеньки“ определены средние значения сигнала интенсивности свечения и СКО для этих значений.

Используя полученные результаты и учитывая вклад темновых шумов, мы построили семейство кривых (рис. 9), отражающих зависимость величины СКО сигнала, который может быть зарегистрирован камерой, от величины этого сигнала в области интегрирования. Разными цветами показаны графики, отображающие разный вклад величины темнового шума в общий шум — от 2 до 15 photoelectron/pixel. Наиболее часто реализуемое значение показано коричневой кривой, ее наименование на графике „3 (9) 62 %“ означает, что данная кривая построена для пикселей, у которых СКО сигнала яркости в одном пикселе равно 3, СКО интегрального сигнала в области 3×3 пикселя равно 9, и таких пикселей всего 62 % от общего количества пикселей в кадре⁴; аналогично следует понимать обозначения для других кривых.

⁴ Это следует из результатов, представленных на гистограмме рис. 6, b.

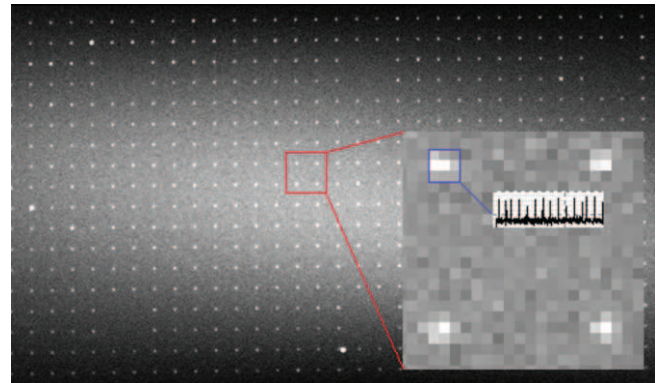


Рис. 7. Фрагмент изображения поверхности реакционной ячейки. Крупно показаны 4 ВНМ в центре пятна засветки лазером и фрагмент трейса в одном из них.

Штриховой линией на рис. 9, 10 показана аппроксимация полиномом пятой степени значений для коричневой кривой „3 (9) 62 %“, характеризующей поведение шумов для большей части пикселей, уравнение описывается формулой:

$$SD_{det.noise\ exper} = 1 \cdot 10^{-18} S^5 - 9 \cdot 10^{-14} S^4 + 2 \cdot 10^{-9} S^3 - 2 \cdot 10^{-5} S^2 + 0.0781 S - 46.827. \quad (1)$$

Темно-синим цветом показана кривая, соответствующая теоретической оценке шумов камеры, сделанной по формуле (2) на основе сведений о составе и характеристиках шумов, полученных из паспортных данных камеры. Формула (2) позволяет оценить СКО шумов для сигнала яркости в области интегрирования 3×3 пикселя и дает нам нижнюю границу существования всех возможных шумов детектора:

$$SD_{det.noise\ theor} = \sqrt{SD1^2 + SD2^2 + SD3^2}, \quad (2)$$

где $SD1 = 3 \cdot \sqrt{3^2} = 9$ — СКО темнового шума в области интегрирования, где 3 — величина СКО темнового шума на пиксель (суммарно для DSNU и ReadOut); $SD2 = 3 \cdot \sqrt{S^2 \cdot 0.003^2} = 0.009 \cdot S$ — теоретическая оценка СКО PRNU шума для области интегрирования, где величина СКО сигнала на пиксель равна 0.3 % от величины сигнала в этом пикселе; $SD3 = \sqrt{S - 900}$ — корень квадратный из амплитуды интегрального сигнала (теоретическая оценка СКО дробового шума в области интегрирования), где S — интегральный сигнал яркости, полученный в области интегрирования, 900 — суммарное смещение нулевого уровня сигнала в области интегрирования.

Особый интерес для моделирования представляет величина шумов в области слабых сигналов, поскольку сигналы флуоресценции, излучаемые одиночными молекулами флуорохромов, имеют малую амплитуду. На рис. 9 пунктирной прямоугольной рамкой выделен

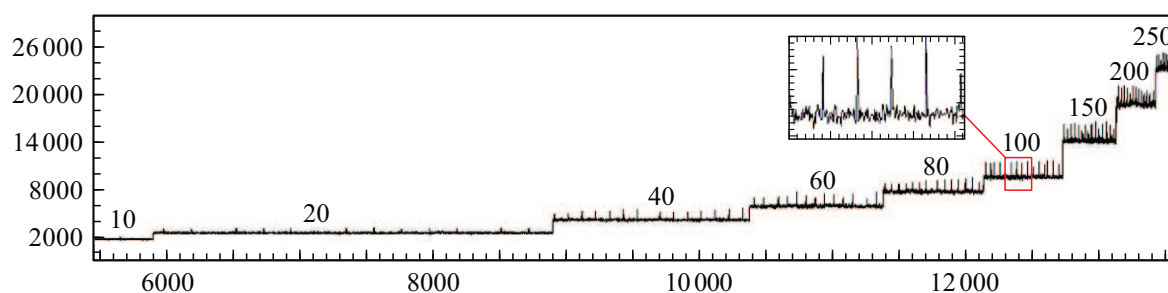


Рис. 8. Визуализация общего трейса, где по оси абсцисс показан номер кадра, а по оси ординат интегральный сигнал яркости в области интегрирования 3×3 пикселя.

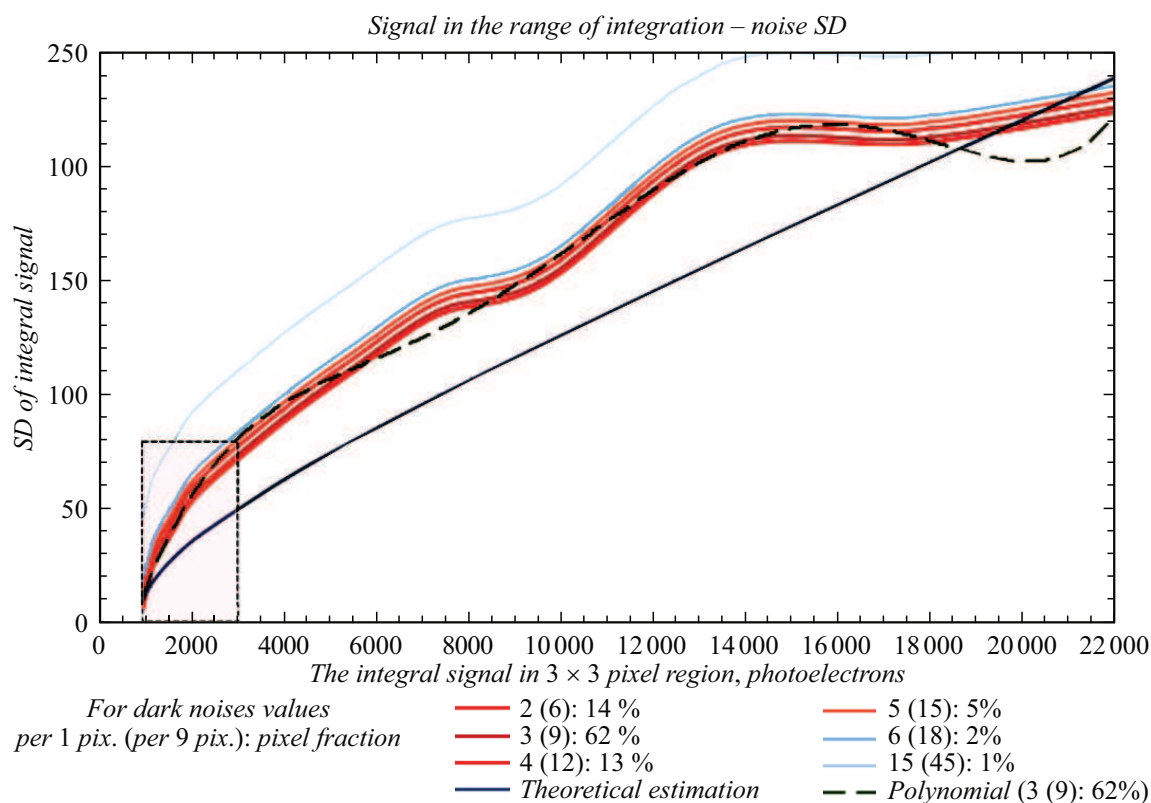


Рис. 9. Зависимость СКО шума детектора от интегрального сигнала для области интегрирования 3×3 пикселя, весь диапазон регистрируемых сигналов.

фрагмент, внутри которого интенсивность интегрального сигнала яркости лежит в диапазоне от 900 до 2900 photoelectron, где 900 соответствует нулевому уровню сигнала трейса. Этот фрагмент графика в увеличенном масштабе показан на рис. 10. Для этого диапазона были проведены измерения интенсивности излучения на очень малых выдержках. Получены серии кадров с экспозициями 500 μ s и 5 ms. Засветка осуществлялась зеленым лазером. Регистрация кадров велась с фиксированной экспозицией, но при изменении (увеличении, а затем уменьшении) мощности лазера от 30 до 1000 mW и снова до 30 mW.

Рис. 10 показывает хорошее совпадение результатов, полученных при различных реальных измерениях фоно-

вых сигналов для большого числа ВМН, для изображений, зарегистрированных нами при работе с разными реакционными ячейками (серия с зелеными кружками) с кривыми, построенными по результатам описанных экспериментов и с кривой аппроксимации (зеленый пунктир).

Кривая, полученная на основе теоретических знаний о составе шума детектора (темно-синяя), проходит на графике заметно ниже, что свидетельствует о наличии дополнительных не учтенных при математической оценке шумов, возникающих в процессе регистрации и передачи данных в компьютер. В то же время реальные данные не попадают в область, лежащую ниже этой кривой.

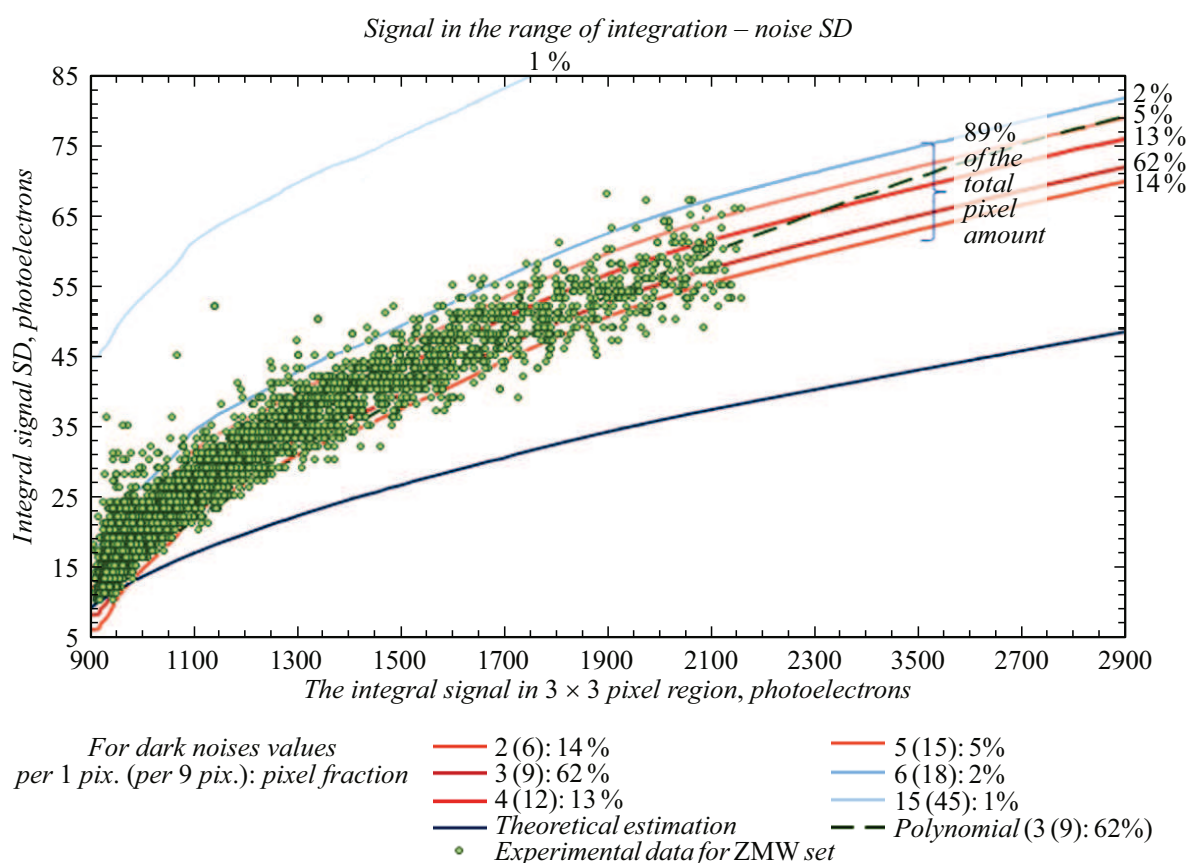


Рис. 10. Зависимость СКО шума детектора от интегрального сигнала для области интегрирования 3×3 пикселя, область слабых сигналов от 900 до 2900 photoelectron

Кривая аппроксимации, описанная уравнением (1), хорошо согласуется с данными различных экспериментов, проведенных нами с реакционными ячейками, особенно в зоне слабых сигналов, характерных для сигналов от одиночных молекул флуорохромов. Рассчитанный по результатам экспериментов для этих данных и данных, полученных в результате аппроксимации, коэффициент детерминации близок к единице, что показывает их высокую статистическую сходимость. Таким образом, уравнение (1) может быть использовано для моделирования шумов детектора при симуляции первичных данных.

3. Формирование модельных трейсов и сравнение с экспериментальными данными

При помощи уравнения (1) были сформированы шумы для модельных трейсов, имитирующих данные, полученные в эксперименте со светодиодом и меняющейся экспозицией, описанном ранее. Результаты моделирования сопоставлены с реальным экспериментом. На рис. 11–13 показаны трейсы для реальных сигналов, полученных с разной экспозицией (черные графики),

и модельные трейсы (синие графики). Реальные трейсы построены для одного и того же ВНМ, расположенного в центре пятна засветки лазером (как показано на рис. 7 в выделенном фрагменте). Сигнал яркости интегрируется в области 3×3 пикселя. Для полученных трейсов программно определены: средний уровень фона в области интегрирования, СКО сигнала фона и амплитуда полезного сигнала, сформированного светом от светодиода. Эти параметры использованы для формирования модельных трейсов (см. таблицу).

Рис. 11 иллюстрирует результаты сравнения для экспозиции 10 ms, рис. 12 — для 100 ms, рис. 13 — для 250 ms.

Амплитуда импульсов, сформированных при помощи модуляции светодиода, остается примерно постоянной для экспозиций, равных или превышающих длительность свечения, поскольку независимо от времени экспозиции мы накапливаем одно и то же дополнительное к фону количество фотонов на сенсорах камеры. Только в случае, когда время экспозиции меньше времени высвечивания диода, событие высвечивания дробится на несколько экспозиций с соответствующим уменьшением амплитуды сигнала.

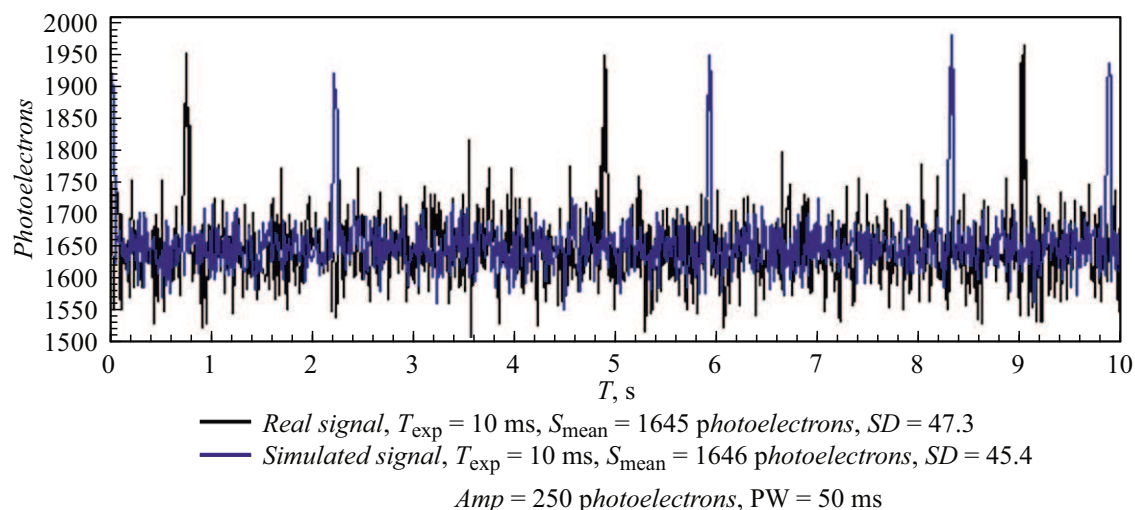


Рис. 11. Моделирование трейса с сигналами амплитудой 250 photoelectron, уровнем фона 1645 photoelectron, СКО фона 47.3 photoelectron (сигнал/шум = 5.3), длительностью 50 ms и экспозицией 10 ms.

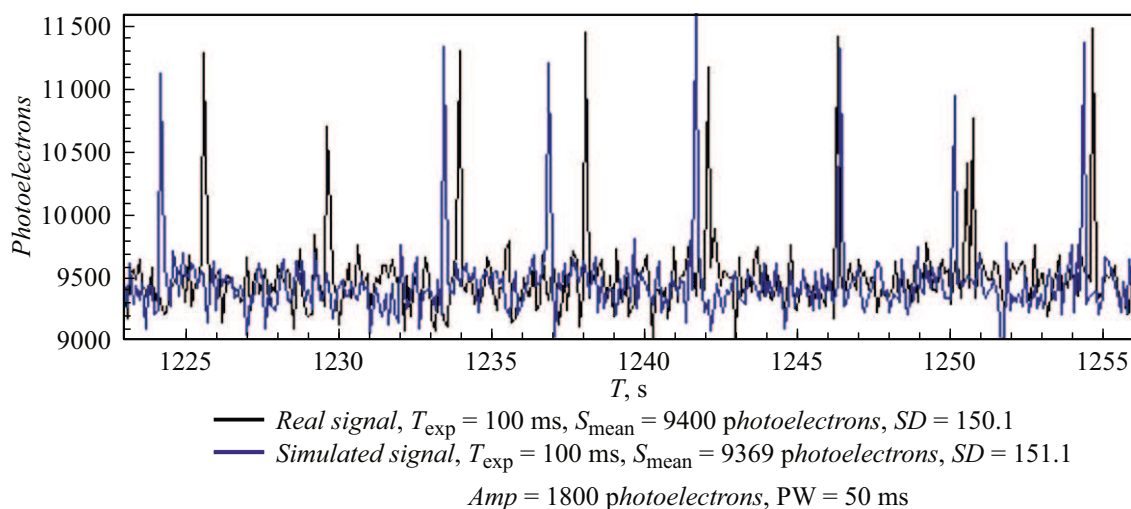


Рис. 12. Моделирование трейса с сигналами амплитудой 1800 photoelectron, уровнем фона 9400 photoelectron, СКО фона 150.1 photoelectron (сигнал/шум = 12), длительностью 50 ms и экспозицией 100 ms.

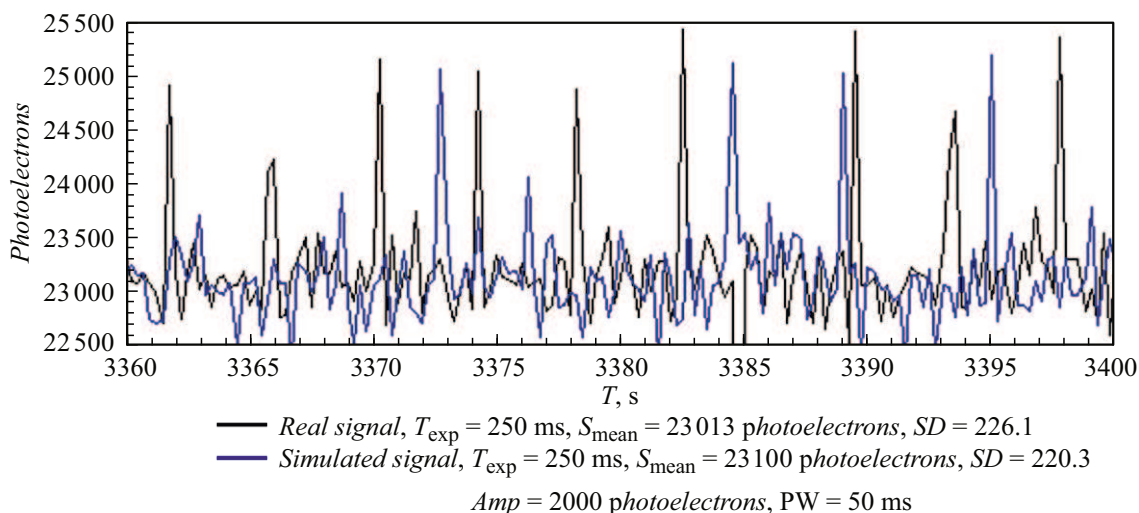


Рис. 13. Моделирование трейса с сигналами амплитудой 2000 photoelectron, уровнем фона 22013 photoelectron, СКО фона 226.1 photoelectron (сигнал/шум = 8.8), длительностью 50 ms и экспозицией 250 ms.

Параметры реальных и модельных трейсов

Время экспозиции Texp, ms	Длительность импульса PW (pulse width), ms	Амплитуда импульса, photoelectron	Средний уровень фона трейса Smean, photoelectron		SD фона трейса, photoelectron	
			Реальные	Модельные	Реальные	Модельные
10	50	250	1645	1646	47.3	45.4
100	50	1800	9400	9369	150.1	151.1
250	50	2000	23 013	23 100	226.1	220.3

Заключение

Проведенное исследование шумов детектора секвенатора позволило получить функциональную зависимость СКО шума детектора от величины регистрируемого сигнала, описанную формулой (1). Были учтены фундаментальные составляющие, формирующие шумы видеокamеры детектора — темновые шумы в сенсорах (DSNU), шум считывания, шум неоднородности фотоответа сенсоров (PRNU), фотонный дробовой шум. Установлено, что математическая оценка по формуле (2) действительно дает „нижний предел“ шума — т.е. отражает то идеальное состояние, меньше которого шума быть не может. Показано, что СКО шумов, определяемое в реальных экспериментах для разных реакционных ячеек, не попадает в область, лежащую ниже математической оценки.

Проведено моделирование реальных трейсов для разных уровней фонового сигнала с разным СКО шума. Из таблицы и графиков на рис. 11–13 видно, что статистические параметры шума модельных и реальных трейсов хорошо совпадают при разных значениях интегрального сигнала яркости ВМ. Также можно отметить, что соотношение сигнал/шум наиболее высокое для импульсов, длительность которых близка ко времени экспозиции кадра. Сигналы от одиночных молекул флуорохрома ожидаемо находятся в области слабых по интенсивности.

Полученные результаты позволили оптимизировать алгоритм формирования реалистичного шума для программы-симулятора первичных данных прототипа одномолекулярного секвенатора, работа над которым ведется в ИАП РАН.

Финансирование работы

Работа выполнена в Институте аналитического приборостроения Российской академии наук (Санкт-Петербург) в рамках Федеральной научно-технической программы развития генетических технологий на 2019–2027 годы (Рег. N НИОКР 122012500127-1), соглашение №075-15-2021-1057 от 28.09.2021.

Конфликт интересов

Авторы заявляют, что у них нет конфликта интересов.

Список литературы

- [1] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, Th. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korfach, S. Turner. *Science*, **323** (5910), 133 (2009). DOI: 10.1126/science.1162986
- [2] J. Korfach, K.P. Bjornson, B.P. Chaudhuri, R.L. Cicero, B.A. Flusberg, J.J. Gray, D. Holden, R. Saxena, J. Wegener, S.W. Turner. *Methods in Enzymology*, **472**, 431 (2010). DOI: 10.1016/S0076-6879(10)72001-2
- [3] M.J. Levene, J. Korfach, S.W. Turner, M. Foquet, H.G. Craighead, W.W. Webb. *Science, New Series*, **299** (5607), 682 (2003)
- [4] B.K. Stöcker, J. Köster, S. Rahmann. *Bioinformatics*, **32** (17), 2704 (2016). DOI: 10.1093/bioinformatics/btw286
- [5] Z.-G. Wei, Sh.-W. Zhang. *Bioinformatics*, **19** (1), 177 (2018). DOI: 10.1186/s12859-018-2208-0
- [6] Yu. Ono, K. Asai, M. Hamada. *Bioinformatics*, **37** (5), 589 (2021). DOI: 10.1093/bioinformatics/btaa835
- [7] W. Zhang, B. Jia, Ch. Wei. *Bioinformatics*, **20** (1), 352 (2019). DOI: 10.1186/s12859-019-2901-7
- [8] Ch. Rohrandt, N. Kraft, P. Gießelmann, B. Brändl, B.M. Schuldt, U. Jetzek, F.-J. Müller. *Nanopore SimulatION — a raw data simulator for Nanopore Sequencing*. 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), p. 1536.
- [9] В.П. Бессмельцев, В.С. Терентьев, В.В. Вилейко, С.А. Бабин, А.М. Шалагин, А.В. Латышев, Д.А. Насимов, Л.И. Федина, Д.В. Пышный, П.Е. Воробьев, В.В. Анненков, Е.Н. Даниловцева, С.Н. Зелинский, О.Н. Верховина, М.А. Грачев, Ю.П. Галачянц. *Прикладная фотоника*, **3** (4), 388 (2016).
- [10] П.Н. Мелентьев, А.С. Калмыков, А.С. Гритченко, М.П. Шеметева, А.М. Сафонова, М.С. Марков, В.И. Балыкин, А.С. Букатин, Н.В. Ваулин, Д.А. Белов, А.А. Евстапов, Д.А. Баклыков, А.В. Андрияш, А.А. Барбашева, А.К. Кутук, В.В. Рыжков, И.А. Родионов, Д.С. Кудрявцев, В.А. Можаяева, Л.В. Сон, В.И. Цетлин, Б.Н. Хлебцов, М.С. Кобзев, Ю.О. Кузнецова, Б.Т. Шарипов, А.С. Яшкин, Я.И. Алексеев. *УФН*, **194** (11), 1130 (2024). DOI: 10.3367/UFNr.2024.07.039720

- [11] M. Mahmoud, Y. Huang, K. Garimella, P.A. Audano, W. Wan, N. Prasad, R.E. Handsaker, S. Hall, A. Pionzio, M.C. Schatz, M.E. Talkowski, E.E. Eichler, S.E. Levy, F.J. Sedlazeck. *Nat. Commun.*, **15**, 837 (2024).
DOI: 10.1038/s41467-024-44804-3
- [12] A. Rhie, S. Nurk, M. Cechova, S.J. Hoyt, D.J. Taylor, N. Altemose, P.W. Hook, S. Koren, M. Rautiainen, I.A. Alexandrov, J. Allen, M. Asri, A.V. Bzikadze, N.-Ch. Chen, Ch.-Sh. Chin, M. Diekhans, P. Flicek, G. Formenti, A. Fungtammasan, C.G. Giron, E. Garrison, A. Gershman, J.L. Gerton, P.G.S. Grady, A. Guarracino, L. Haggerty, R. Halabian, N.F. Hansen, R. Harris, G.A. Hartley, W.T. Harvey, M. Haukness, J. Heinz, Th. Hourlier, R.M. Hubley, S.E. Hunt, S. Hwang, M. Jain, R.K. Kesharwani, A.P. Lewis, H. Li, G.A. Logsdon, J.K. Lucas, W. Makalowski, Ch. Markovic, F.J. Martin, A.M. McCartney, R.C. McCoy, J. McDaniel, B.M. McNulty, P. Medvedev, A. Mikheenko, K.M. Munson, T.D. Murphy, H.E. Olsen, N.D. Olson, L.F. Paulin, D. Porubsky, T. Potapova, Fedor Ryabov, S.L. Salzberg, M.E.G. Sauria, F.J. Sedlazeck, K. Shafin, V.A. Shepelev, A. Shumate, J.M. Storer, L. Surapaneni, A.M. Taravella, F. Thibaud-Nissen, W. Timp, M. Tomaszekiewicz, M.R. Vollger, B.P. Walenz, A.C. Watwood, M.H. Weissensteiner, A.M. Wenger, M.A. Wilson, S. Zarate, Y. Zhu, J.M. Zook, E.E. Eichler, R.J. O'Neill, M.C. Schatz, K.H. Miga, K.D. Makova, A.M. Phillippy. *Nature*, **621**, 344 (2023).
DOI: 10.1038/s41586-023-06457-y
- [13] Электронный ресурс. *Camera Dhyana 400BSI V3 Tucsen Photonics Co., Ltd.* Режим доступа:
<https://www.tucsen.com/scmos-dhyana-400bsi-v3-product>