ЖЭТФ, 2018, том 153, вып. 6, стр. 879–894

КВАНТОВЫЙ ГЕНЕРАТОР СЛУЧАЙНЫХ ЧИСЕЛ, ОСНОВАННЫЙ НА ПУАССОНОВСКОЙ СТАТИСТИКЕ ФОТООТСЧЕТОВ, СО СКОРОСТЬЮ ОКОЛО 100 Мбит/с

К. А. Балыгин [a,f], ***В. И. Зайцев*** [a,f], ***А. Н. Климов*** [a,b,f], ***С. П. Кулик*** [a,f], ***С. Н. Молотков*** [c,d,e,f*]

[a] *Физический факультет, Московский государственный университет им. М. В. Ломоносова 119899, Москва, Россия*

[b] *Институт общей физики им. А. М. Прохорова Российской академии наук 119991, Москва, Россия*

[c] *Институт физики твердого тела Российской академии наук 142432, Черноголовка, Московская обл., Россия*

[d] *Академия криптографии Российской Федерации 121552, Москва, Россия*

[e] *Факультет вычислительной математики и кибернетики, Московский государственный университет им. М. В. Ломоносова 119899, Москва, Россия*

[f] *Центр квантовых технологий, Московский государственный университет им. М. В. Ломоносова 119899, Москва, Россия*

Поступила в редакцию 7 февраля 2018 г.

Представлена экспериментальная реализация квантового генератора случайных чисел. Первичным источником случайности являются последовательности фотоотсчетов от квазиоднофотонного излучения, которое регистрируется матрицей кремниевых лавинных детекторов — SiPM (Silicon Photo Multiplier). Использование SiPM позволяет надежно контролировать квантовый характер пуассоновской статистики фотоотсчетов. Специальный алгоритм неэкспоненциальной сложности позволяет извлекать из пуассоновского процесса всю случайность, содержащуюся в нем, а именно, случайную равномерную последовательность 0 и 1.

DOI: 10.7868/S0044451018060044

1. ВВЕДЕНИЕ

Случайные числа давно используются в математическом моделировании методом Монте-Карло, но наиболее широкое применение случайные числа находят в криптографии. В классической криптографии случайные числа используются в качестве секретных ключей в алгоритмах симметричного шифрования, для генерации паролей и PIN-кодов различных пластиковых карт и в системах разграничения доступа. В системах асимметричной криптографии с отрытыми ключами случайные числа используются для генерации и проверки простоты длинных чисел. Если в классических системах симметричного шифрования секретные ключи меняются на передающей и приемной сторонах при помощи оператора, то частая смена ключей, например, раз в десятки секунд, практически невозможна. Если секретные ключи меняются не столь часто, то они используются как мастер-ключи для получения производных от них сеансовых ключей, что в принципе может приводить к понижению криптостойкости системы.

В системах квантовой криптографии возможна частая смена секретных ключей, но при этом требуется большой расход случайных чисел. По сути, системы квантовой криптографии представляют

* E-mail: sergei.molotkov@gmail.com

собой распределенную систему согласования двух независимых случайных последовательностей на передающей и приемной сторонах при помощи квантовых квазиоднофотонных (в идеале однофотонных) состояний. Квантовые состояния на передающей стороне приготавливаются в соответствии со случайной последовательностью 0 и 1. На приемной стороне происходит выбор типа измерения в соответствии со случайной последовательностью 0 и 1. Если выбор измерения (0 или 1) и квантовое состояние (0 или 1) совпадают, то возможен фотоотсчет. В этом случае, если не было вторжений в канал связи, случайные биты (0 или 1) на передатчике и (0 или 1) на приемнике совпадают. Фактически происходит синхронизация двух независимых случайных последовательностей посредством посылки и измерения квантовых состояний. При несовпадении случайных битов приемника и передатчика фотоотсчета не будет, данная посылка будет отброшена после обмена через открытый канал связи номерами позиций, где были фотоотсчеты. Поскольку последовательности являются случайными и независимыми, в среднем позиции в исходных последовательностях передатчика и приемника совпадают примерно в половине случаев. Реально это совпадение происходит в существенно меньшей доле исходных последовательностей.

Эффективность однофотонных детекторов заметно меньше единицы, существуют потери в линии связи (в стандартном оптоволокне на расстоянии 100 км долетает до приемника в среднем каждый сотый фотон), нет строго однофотонных источников, вместо них используется сильно ослабленное когерентное излучение лазера с ослаблением в среднем на десятые доли фотона в импульсе. Последнее приводит к тому, что примерно лишь каждый десятый импульс излучения содержит однофотонное фоковское состояние, в остальных девяти импульсах — вакуумное состояние излучения. В итоге на генерацию общего ключа, например в 256 бит, между приемником и передатчиком требуются как минимум случайные последовательности в 10^5–10^6 бит. Данные грубые оценки показывают требуемый масштаб скорости генерации случайных чисел. Например, для генерации и смены секретного ключа в 256 бит раз в секунду требуется скорость генерации исходной случайной последовательности не менее 10^8 Мбит/с.

Данный пример показывает, что расход случайных чисел в квантовой криптографии существенно больше, чем в классической. Поэтому требуются генераторы случайных чисел с высокой скоростью генерации и доказуемым качеством случайной последовательности. Под доказуемым качеством случайной последовательности понимается следующее. Нельзя доказать, что последовательность случайная. Грубо говоря, отсутствует физический материальный эталон случайной последовательности. Можно лишь утверждать по некоторому критерию (см. ниже), что последовательность не противоречит гипотезе случайности.

Принципиально важно, что при разработке генераторов случайных чисел недостаточно того обстоятельства, что тесты на случайность по некоторому критерию проходят. Это лишь необходимое условие. Принципиально важен источник первичной случайности, который используется для получения равнораспределенной последовательности 0 и 1 и который был бы действительно источником случайности по соображениям, не зависящим от тестов (например, процесс измерений над квантовой системой, см. ниже). Многие псевдослучайные генераторы случайных чисел проходят тесты, но не являются, очевидно, истинно случайными.

Все генераторы случайных чисел можно разделить на два класса. Первый класс — физические генераторы, когда случайная последовательность извлекается из некоторого физического процесса. Второй класс — математические генераторы случайных чисел, когда случайная последовательность получается как результат математического преобразования, как правило, рекуррентного, некоторого затравочного числа. Любой математический генератор выдает псевдослучайную последовательность, которая полностью предсказуема, если известны начальные условия — затравочное число.

Для генерации ключей в системах симметричного шифрования используют исключительно физические генераторы случайных чисел. Физические генераторы основаны на измерении состояния физической системы. Если система эволюционирует по законам классической физики, т. е. эволюция описывается дифференциальными уравнениями, то случайность результата измерения связана только с неизвестностью начальных условий. В этом смысле последовательность результатов измерений является псевдослучайной — определяется, при известном законе эволюции, только неопределенностью начальных условий.

Хорошим примером, иллюстрирующим сказанное, является известная доска Гальтона [1] для демонстрации нормального распределения вероятностей как результата центральной предельной теоремы. Доска Гальтона — система с твердыми металлическими шариками, падающими через несколько

рядов тонких штырьков, расположенных в шахматном порядке. Такая система является чисто классической. Падая вниз, шарик испытывает упругое отражение от одного из штырьков в ряду, и в итоге попадает на дне в один из ящиков, расположенных по горизонтали. Бросание большого числа шариков приводит к распределению шариков по ящикам, близкому к нормальному закону.

Может ли такая система, ведущая себя по законам классической физики, привести к генерации случайности? Очевидно, нет.

Траектория каждого шарика и его финальное положение — ящик на дне, в котором он окажется — могут быть достоверно предсказаны, если известен угол, под которым он входит в первый ряд. Небольшие, но известные отклонения в начальном угле падения каждого шарика, в итоге приводят к нормальному закону распределения. Подчеркнем еще раз, что если все начальные углы известны, то все распределение по ящикам однозначно предсказуемо. «Случайность» связана только с незнанием начальных условий.

Важно (и это свойство любой классической системы), что если система приготовлена в начальный момент в одних и тех же начальных условиях и проходит одну и ту же эволюцию, то это будет приводить к одному и тому же конечному результату. В этом смысле классические физические генераторы являются псевдослучайными. На принципиальном уровне эволюция любой сложной классической системы полностью предсказуема (принципиально может быть вычислена и предсказана) при известных начальных условиях.

Этот пример также явно показывает, почему важно знать источник первичной случайности.

В данном примере исходный источник не является истинно случайным. Но если заранее не знать происхождение «случайности», т. е. то, что распределение шариков на дне произошло не из случайного (при условии известности начальных условий), а детерминированного источника, то можно принять такой источник за истинно случайный.

В физических генераторах случайных чисел, использующих, например, оцифровку шума Джонсона–Найквиста, довольно трудно (а практически контролируемым образом невозможно) отделить классический шум от квантового.

Принципиально другая ситуация с квантовыми генераторами случайных чисел, где случайность возникает как результат измерения квантовой системы. Эволюция квантовой системы самой по себе также полностью детерминирована, поскольку описывается дифференциальными уравнениями с начальными условиями.

Однако результат измерения над квантовой системой, приготовленной каждый раз при одном и том же начальном условии и испытывающей одну и ту же эволюцию, будет принципиально приводить к непредсказуемому результату. В этом смысле в микромире случайность «встроена» и возникает как результат измерения. Поэтому истинная случайность имеет место только в микромире.

Такими квантовыми процессами могут быть, например, α-распад, фотоэффект и т. д. Обзор различных реализаций квантовых генераторов случайных чисел можно найти в работе [2], см. также [3–13]. Квантовые оптические генераторы случайных чисел также можно разделить на две группы. В работах, соответствующих первой группе, в качестве исходной случайной величины используют непрерывную случайную величину. Такой величиной обычно является квадратурная компонента поля при гомодинном детектировании [3]. В последнем случае, по-видимому, приходится делать предположения о статистических свойствах распределения флуктуаций фазы. Исходной случайной величиной для преобразования в последовательность 0 и 1 является разность сигналов с двух классических детекторов, работающих не в режиме счета фотонов, а в линейном режиме. Случайная измеряемая величина возникает как разность токов двух классических фотодетекторов. Поэтому в этом случае сложно контролируемым образом выделить квантовую составляющую сигнала.

Принципиально важно знать и уметь экспериментально проверять источник первичной случайности, из которого посредством постобработки возникает случайная последовательность 0 и 1. Как пример, в работе [4] декларировалась скорость генерации в несколько ГГц, при этом постобработка состояла в простой операции XOR, для блоков исходной последовательности, сдвинутых друг относительно друга, что вряд ли можно убедительно обосновать.

Вторая группа квантовых генераторов случайных чисел в качестве первичной физической случайности использует дискретные фотоотсчеты.

Главная экспериментальная трудность при реализации высокоскоростных квантовых генераторов случайных чисел — соблюсти противоречивые требования, т. е. обеспечить квантовый характер сигнала (малое среднее число фотонов) и получить высокую скорость генерации случайных чисел. Одним из подходящих для экспериментальной реализации

квантовых генераторов является фотоэффект — поглощение фотона атомом и выбивание электрона, который может регистрироваться. Фотоэффект был открыт Столетовым еще до создания квантовой механики, поэтому не был объяснен. Последовательное объяснение фотоэффекта было дано Эйнштейном с использованием понятия квантов. Акт поглощения фотона определяется квантовомеханической вероятностью. Главная задача при реализации квантовых генераторов чисел — «дотянуться» экспериментально до квантового процесса поглощения, т. е. отделить случайность, связанную с процессом измерения, от других шумов — неконтролируемых источников «случайности».

Первая сложность состоит в обеспечении однофотонного состояния. Реально удается реализовать квазиоднофотонность состояния путем сильного ослабления лазерного излучения, которое представляет собой когерентное состояние. В когерентном состоянии задано только среднее число фотонов μ. Чем меньше среднее число фотонов, тем когерентное состояние ближе к однофотонному. Однако уменьшение среднего числа фотонов приводит к росту вакуумной компоненты в состоянии, т. е. уменьшается вероятность регистрации. Одномодовое когерентное состояние имеет вид

$$|\alpha\rangle = \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle, \quad \mu = |\alpha|^2, \tag{1}$$

где $|n\rangle$ — n-фотонное фоковское состояние. При малых $\mu \ll 1$ доля однофотонной компоненты в (1) порядка μ, а вакуумной — $|0\rangle = |\mathrm{vac}\rangle \approx e^{-\mu} \approx 1-\mu \approx 1$.

Поскольку детектор не реагирует на вакуумную компоненту поля, вероятность регистрации идеальным детектором когерентного состояния приводит к пуассоновской статистике фотоотсчетов. Вероятность обнаружить n фотонов во временном окне T есть (см., например, [14, 15])

$$P_T(n) = e^{-\mu T} \frac{(\mu T)^n}{n!}, \tag{2}$$

т. е. дается пуассоновским распределением вероятностей. Формула (1) описывает одномодовое (монохроматическое) когерентное состояние. Ниже нами будет использоваться многомодовый лазер (фактически светодиод). Каждая мода будет приводить к пуассоновской статистике фотоотсчетов со своим показателем. Однако это не критично, поскольку сумма пуассоновских процессов снова дает пуассоновский процесс с новым эффективным параметром в формуле (2), который реально и наблюдается.

Фотодетектор не различает числа фотонов, а только сам факт регистрации. Согласно [14], вероятности появления фотоотсчета $P(*)$ или его отсутствие $P(\sqcup)$ во временном окне длительностью T, даются выражениями

$$P(*) = 1 - e^{-\mu T}, \quad P(\sqcup) = e^{-\mu T}. \tag{3}$$

В итоге приходим к схеме испытаний Бернулли, независимым подбрасываниям несимметричной «монетки» с вероятностями выпадения для Head — $P(*)$ и Tail — $P(\sqcup)$.

Забегая вперед, подчеркнем, что для используемой нами процедуры извлечения случайности требуется только статистическая независимость фотоотсчетов в разных временных окнах и не требуется знание самих вероятностей $p = P(\sqcup)$ и $1 - p = P(*)$. То есть метод гарантирует извлечение истинно случайных 0 и 1 (при условии независимости испытаний) при любом значении вероятностей в (3), величины которых влияют только на скорость генерации.

Как известно, первопричина пуассоновской статистики фотоотсчетов при детектировании лазерного излучения носит принципиально квантовый характер (см. детали в [2, 14]).

Реальный лавинный фотодетектор имеет квантовую эффективность η, не равную единице, в этом случае вероятность сохраняет вид (3), но с заменой $\mu T \to \eta\mu T$. Поглощение отдельного фотона в твердотельной структуре лавинного детектора приводит к рождению электрон-дырочной пары, которая «усиливается» — рождает лавину носителей заряда, импульс тока от которой регистрируется. После рождения лавины происходит процесс рассасывания лавины, что требует определенного времени. Пока не произошло рассасывания лавины, фотодетектор не готов к новому акту регистрации, иначе это приведет к корреляции фотоотсчетов и искажению пуассоновской статистики, т. е. фотоотсчеты, особенно в близких временных окнах, перестают быть независимыми. Восстановление детектора до следующего акта регистрации обеспечивает статистическую независимость последовательных фотоотсчетов.

Время рассасывания лавины является внутренней характеристикой детектора. Это время накладывает ограничение на скорость фотодетектирования и, соответственно, на скорость генерации случайной последовательности. Типичные времена составляют от десятков до сотен наносекунд, что дает ограничение по частоте генерации даже в оптимистическом случае от 10 МГц до 100 МГц.

Первая задача состоит в контролируемом и доказуемом способе получения исходной физической случайности — пуассоновской статистики фотоотсчетов. Данная цель достигается использованием матрицы лавинных детекторов SiPM.

Вторая задача состоит в извлечении всей случайности, содержащейся в пуассоновском процессе; точнее говоря, в эффективной экстракции случайной равномерно распределенной последовательности 0 и 1. Экстракция или постобработка пуассоновской последовательности фотоотсчетов должна приводить к доказуемой истинно случайной последовательности 0 и 1 при условии, что исходная последовательность имеет пуассоновский характер.

2. ГРУППИРОВКА ПОСЛЕДОВАТЕЛЬНОСТЕЙ ФОТООТСЧЕТОВ

Любой случайный процесс содержит некоторое максимальное количество истинно случайных 0 и 1. Наша задача — извлечь максимально допустимое число истинно случайных битов из последовательностей фотоотсчетов, которое содержится в физическом процессе и не может быть превышено. Число истинно случайных битов в пересчете на такт, которое может быть получено из последовательности фотоотсчетов и пропусков $\{*, \sqcup\}^n$, зависит от способа группировки фотоотсчетов в случайные события. Более формально оно зависит от способа отображения последовательности фотоотсчетов в случайную последовательность 0 и 1: $\{*, \sqcup\}^n \to \{0,1\}^l$ $(l \le n)$.

Существуют в определенном смысле универсальные методы экстракции случайности [16–18], которые позволяют получить последовательность 0 и 1, сколь угодно близкую к истинно случайной. Мерой близости является следовое расстояние — расстояние Колмогорова между двумя распределениями вероятностей:

$$||P_X - P_U||_1 = \frac{1}{2}\sum_{x_i=0,1} |P_X(x_1, x_2, \dots, x_n) - P_U(x_1, x_2, \dots, x_n)| < \varepsilon, \quad (4)$$

$$P_U(x_1, x_2, \dots, x_n) = P_U(x_1)P_U(x_2)\dots P_U(x_n),$$

$$P_U(0) = P_U(1) = \frac{1}{2},$$

где $P_X(x)$ — функция распределения полученной последовательности 0 и 1. Вероятность $P_X(x)$ называется ε-близкой к истинно случайной. Параметр близости ε определяется требованиями по использованию случайных последовательностей. Требуемая близость к идеальной случайной последовательности достигается сжатием (хешированием) при помощи универсальных хеш-функций второго порядка, которые сами являются случайной величиной, что требует затравочной случайности. Такой метод используется, например, для случайных генераторов в [7].

Результаты последовательных актов измерений $(x_1, x_2, \dots, x_n)$ над квантовой системой в общем случае не являются статистически независимыми. При этом функция распределения не распадается на произведение,

$$P_X(x_1, x_2, \dots, x_n) \ne P(x_1)P(x_2)\dots P(x_n).$$

Число истинно случайных битов дается минимальной энтропией Реньи [18]

$$H_{min} = -\log\left(\max_{P_X(x_1,x_2,\dots,x_n)} P_X(x_1, x_2, \dots, x_n)\right), \quad (5)$$

здесь и везде ниже $\log \equiv \log_2$. Экстракторы позволяют извлечь случайность из распределения

$$P_X(x_1, x_2, \dots, x_n) \ne P_X(x_1)P_X(x_2)\dots P_X(x_n)$$

(см., например, [16–18]). Однако проблема состоит в том, что в реальной ситуации функция распределения $P(x_1, x_2, \dots, x_n)$ неизвестна и приходится строить предположения, которые трудно экспериментально проверить.

В случае, когда последовательные исходы независимы, $P_X(x_1, x_2, \dots, x_n) = P_X(x_1)P_X(x_2)\dots \dots P_X(x_n)$, минимальная энтропия в асимптотическом пределе $n \to \infty$ стремится к энтропии Шеннона (см. ниже). Более надежным является другой путь, который состоит в том, чтобы обеспечить на физическом уровне реализации такую последовательность измерений, статистические свойства которой достаточно просты для экспериментальной верификации. В нашем случае задача состоит в достижении пуассоновского характера статистики последовательности фотоотсчетов и выборе такой процедуры экстракции случайной последовательности 0 и 1, которая при пуассоновской статистике исходной последовательности будет давать истинно случайную последовательность 0 и 1, не ε-близкую (4) к истинно случайной.

2.1. Метод фон Неймана

Уместно напомнить метод извлечения истинно случайных 0 и 1 из бернуллиевской последовательности, который предложил фон Нейман еще в 1951 г. [19] и который фактически подсказывает более общий и эффективный метод экстракции случайности.

Метод состоит в следующем. Пошагово просматривается последовательность $\{*, \sqcup\}^n$. Если обнаружены две последовательные позиции $*\sqcup$, то они заменяются на $*\sqcup \to 0$, если встречается комбинация $\sqcup*$, то она заменяется на $\sqcup* \to 1$. Две другие парные комбинации $**$, $\sqcup\sqcup$ отбрасываются. Полученная последовательность 0 и 1 является равномерно распределенной случайной последовательностью. Вероятность 0 и 1 в новой последовательности строго равна 1/2; 0 и 1 равновероятны при любом значении исходных вероятностей $1-p \to *$ и $p \to \sqcup$. В данном методе, даже в самом лучшем случае, когда $p \approx 1-p$, теряется более половины исходной последовательности. В методе фон Неймана еще не использовались методы теории информации, метод не позволяет извлечь максимально возможную длину случайной последовательности, содержащейся в физическом процессе.

Данный метод, несмотря на его простоту, изящность и почти самоочевидность, не учитывает нечто существенно более общее и важное. Удобнее представить метод в виде таблицы:

$$\begin{aligned} \sqcup\sqcup &\to \text{discard}, \\ \sqcup* &\to 0, \\ *\sqcup &\to 1, \\ ** &\to \text{discard}. \end{aligned} \tag{6}$$

Как видно из (6), метод фон Неймана состоит в группировке исходных физических событий в различные классы таким образом, чтобы все представители из одного класса имели одинаковую вероятность. Как видно из (6), существуют три класса. В первом и третьем классах имеется по одному элементу. Данные классы отбрасываются. В одном классе два равновероятных элемента.

Если в классе число элементов равно степени двойки (пока считаем так для краткости рассуждений, общий случай см. ниже) (в примере выше $2^1 = 2$), то бинарное представление порядкового номера исходного блока в этом классе, начиная с нуля, и есть новый блок истинно случайных 0 и 1 (формула (6)). В более общем виде, если перенумеровать в лексикографическом порядке все последовательности в одном классе, то бинарное представление номера в классе будет блоком выходной истинно случайной последовательности. Это дает более общий метод для извлечения случайности из бернуллиевской последовательности.

Выясним теперь, как группировать последовательности фотоотсчетов в классы равновероятных последовательностей, чтобы извлечь максимум случайности.

2.2. Энтропия Шеннона, испытания Бернулли, статистика Ферми – Дирака

Для источников без памяти имеет место асимптотическое равнораспределение выходных последовательностей [20, 21]. При длинной последовательности имеется приблизительно $2^{nh(p)}$ типичных последовательностей, которые имеют равную вероятность ($h(p) = -p\log(p) - (1-p)\log(1-p)$ — бинарная энтропийная функция Шеннона [20,21]). Множество типичных последовательностей при $n \to \infty$ реализуется с вероятностью единица. Применительно к нашему случаю в асимптотическом пределе каждая типичная строка содержит приблизительно $n(1-p)$ отсчетов $*$ и, соответственно, приблизительно np пропусков $\sqcup$. Из этого факта вытекает рецепт извлечения случайной последовательности 0 и 1. Нужно перенумеровать в лексикографическом порядке все типичные последовательности и присвоить свой номер $0 \le \mathrm{Num}(x) \le [2^{nh(p)}] - 1$ ($x = x_1, x_2, \ldots, x_n$ — последовательность $*$ и $\sqcup$) каждой последовательности. Поскольку все последовательности равновероятны, бинарное представление номера может быть превращено в последовательность 0 и 1. Длина случайной последовательности равна числу двоичных разрядов, необходимых для записи максимального значения номера $\mathrm{Num}(x)$. В асимптотическом пределе из последовательности фотоотсчетов $\{\sqcup, *\}^n$ длины n можно извлечь случайную последовательность 0 и 1 длины $[nh(p)]$ битов, где $[\ldots]$ — целая часть.

На практике всегда приходится ограничиваться конечной длиной n последовательности фотоотсчетов. Для экстракции максимального числа случайных битов на такт далее будут обрабатываться все последовательности, а не только типичные. При большом n доля нетипичных последовательностей автоматически будет стремиться к 0. Поэтому метод асимптотически точный в смысле извлечения случайности. Но чтобы ничего не терять, будем обрабатывать все последовательности.

Возможны различные способы группировки последовательностей в равновероятные классы. Ранее

[22] было показано, что оптимальной группировкой — разбиением на классы равновероятных последовательностей — является группировка, при которой в один класс входят последовательности равной длины, имеющие одинаковое число фотоотсчетов и, соответственно, пропусков. Такая группировка фотоотсчетов приводит к статистике Ферми – Дирака (F–D) [23]. Последовательность разбивается на блоки длиной n, т. е. содержит n тактов (ящиков), — аналог уровней энергии. Пусть последовательность содержит $0 \le k \le n$ фотоотсчетов (фотоотсчет — аналог частицы). Все последовательности длины n с k фотоотсчетами имеют одинаковую вероятность $P(*)^k P(\sqcup)^{n-k}$ и относятся к одному классу. Число последовательностей в классе (статистический вес) равно числу способов размещения k частиц по n уровням (ящикам) так, чтобы на каждом уровне было не более одной частицы (не более одного фотоотсчета в такте). Удобно воспользоваться хорошо известной формулой для биномиальных коэффициентов:

$$(1-p+p)^n = \sum_{k=0}^{n} C_n^k (1-p)^k p^{n-k}, \qquad C_n^k = \frac{n!}{k!(n-k)!}. \tag{7}$$

Это известная формула Бернулли для подсчета вероятностей числа k успешных испытаний в серии длины n. Число последовательностей в классе есть C_n^k, и каждая из них имеет вероятность $P_n(k) = (1-p)^k p^{n-k}$. Выясним предельное число случайных битов, которые могут быть извлечены при таком способе группировки при заданной длине блока n. Количество битов (пусть нецелое, но максимальное), которое можно получить из каждой последовательности из класса k, есть $\log(C_n^k)$, соответственно, из всех последовательностей в классе — $C_n^k \log(C_n^k)$ битов. Если вероятность появления каждой последовательности в классе $P_n(k) = (1-p)^k p^{n-k}$, то среднее число случайных битов из всех классов равно

$$H_n^{F-D}(p) = \sum_{k=0}^{n} P_n(k) C_n^k \log(C_n^k). \tag{8}$$

В асимптотическом пределе ($n \to \infty$) $H_n^{F-D}(p)$ стремится к предельному значению $nh(p)$, которое дается энтропией Шеннона. Метод группировки по статистике Ферми – Дирака асимптотически точный и позволяет получить максимальное число случайных битов, содержащихся в физическом процессе. Для энтропии Шеннона имеем

$$H_n^{Sh}(p) = -\sum_{k=0}^{n} C_n^k P_n(k) \log(P_n(k)), \tag{9}$$

где $-\log(P_n(k))$ — количество истинно случайных битов, содержащихся в последовательности длины n с k отсчетами. Асимптотический предел по числу случайных битов на такт есть $H_n^{Sh}(p)/n = h(p)$, он не зависит от длины последовательности. Учитывая, что

$$\begin{aligned} H_n^{Sh}(p) &= -\sum_{k=0}^{n} C_n^k P_n(k) \log(P_n(k)) = \\ &= -\sum_{k=0}^{n} C_n^k (1-p)^k p^{n-k} \left(k\log(1-p) + (n-k)\log(p)\right) = \\ &= -\sum_{k=0}^{n} C_n^k \left(p^{n-k}\log(1-p)\frac{\partial}{\partial(1-p)}(1-p)^k + \right. \\ &\qquad \left. + (1-p)^k \log(p)\frac{\partial}{\partial p} p^{n-k} \right) = nh(p), \end{aligned} \tag{10}$$

покажем, что извлекаемое число случайных битов при группировке по Ферми – Дираку меньше теоретического предела при конечных n, но выходит на предельное значение с ростом n, т. е. $H_n^{F-D}(p) \to H_n^{Sh}(p)$ при $n \to \infty$. Пусть для краткости $N_n(K) = C_n^K$, так как

$$\sum_{k=0}^{n} N_n(k) P_n(k) = 1, \quad N_n(k) P_n(k) \le 1, \quad \log(N_n(k)) \le -\log(P_n(k)), \tag{11}$$

откуда следует, что

$$\begin{aligned} H_n^{F-D}(p) &= \sum_{k=0}^{n} P_n(k) N_n(k) \log(N_n(k)) \le \\ &\le H_n^{Sh}(p) = -\sum_{k=0}^{n} P_n(k) N_n(k) \log(P_n(k)). \end{aligned} \tag{12}$$

При $n \to \infty$ число фотоотсчетов $*$ в последовательности длины n равно $k \approx (1-p)n$, соответственно число пропусков $\sqcup$ равно $n-k \approx np$. При этом

$$\log(N_n(k)) = \log(C_n^k) = \log\left(\frac{n!}{(n-k)!k!}\right).$$

С учетом формулы Стирлинга $n! \approx \sqrt{2\pi n}\,(n/e)^n$ находим в главном приближении

$$\log(C_n^K) \approx n \log\left(\frac{1}{p(1-p)}\right) = nh(p),$$

подставляя в (12), получаем $H_n^{F-D}(p) \to H_n^{Sh}(p) = nh(p)$.

Интересно отметить связь (9) с энтропией Больцмана в статистической механике [23]. Одна позиция,

имеющая истинно случайные биты — равновероятное появление 0 или 1 — несет один бит информации. Энтропия Больцмана с точностью до константы Больцмана равна $S = \log(W)$, где W — статистический вес состояния физической системы, т. е. число возможных микросостояний (способов), с помощью которых можно реализовать данное макроскопическое состояние системы. Применительно к нашему случаю $W = C_n^k$ — это число способов размещения k частиц по n ящикам, оно же есть число равновероятных последовательностей в классе и определяет информацию — число истинно случайных битов, которые требуются для нумерации всех состояний системы.

Распределение Бернулли, статистика Ферми – Дирака, энтропия Шеннона и энтропия Больцмана оказываются тесно связанными друг с другом через количество информации в битах, которое нужно для полного описания состояния системы — нумерации всех состояний. Фактически каждое микросостояние системы нумеруется некоторым номером, для бинарного представления которого требуется не более $[S] = [\log(W)] + 1$ двоичных разрядов.

3. НУМЕРАЦИЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ФОТООТСЧЕТОВ ВНУТРИ КЛАССА РАВНОВЕРОЯТНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ПО ТРЕУГОЛЬНИКУ ПАСКАЛЯ

После разбиения на классы равновероятных последовательностей необходимо перенумеровать все последовательности в данном классе в произвольном порядке. Чем больше длина обрабатываемого блока, тем больше случайности можно извлечь, тем ближе к асимптотическому пределу. Поэтому желательно иметь обрабатываемые блоки как можно длиннее.

Умозрительно можно задать $0 \le k \le n$ и для каждого k сгенерировать все последовательности, записать их в память и присвоить каждой номер (прямая адресация). Например, при длине блока в $n = 64$ такта масштаб числа последовательностей фотоотсчетов $2^{64} \approx 10^{19}$, что требует объема памяти в миллион Терабит (примерно 10 миллиардов флешек, каждая емкостью 1 Гбит — по одной с каждого жителя Земли), т. е. из-за громадного объема памяти прямой способ нумерации технически невозможен. Поэтому требуется эффективный метод нумерации «на ходу» — по мере появления фотоотсчетов в последовательности.

Удобно воспользоваться замечательным методом из теории арифметического кодирования, открытым В. Ф. Бабкиным в 1971 г. [24], который позволяет нумеровать бернуллиевские последовательности с полиномиальными, а не экспоненциальными ресурсами по памяти и времени. В дальнейшем этот метод переоткрывался разными авторами.

Номер каждой последовательности вычисляется по мере ее возникновения по тактам. Пусть последовательные позиции фотоотсчетов $*$ обозначены как $i_1, i_2, \ldots, i_n$. После просмотра всей последовательности известно полное число k фотоотсчетов в ней.

Обозначим номер позиции индексом i_m. Тогда имеется взаимно однозначное соответствие между последовательностями фотоотсчетов $(i_1, i_2, \ldots, i_n)$ и номерами последовательностей $\mathrm{Num}(i_1, i_2, \ldots, i_n)$ $(0 \le \mathrm{Num}(i_1, i_2, \ldots, i_n) \le C_n^k - 1)$. Имеем

$$\mathrm{Num}(i_1, i_2, \ldots, i_n) = C^1_{i_1-1} + \\ + C^2_{i_2-1} + \ldots + C^k_{i_k-1}, \quad C^l_j = 0, \quad j < l. \quad (13)$$

Верхний индекс биномиального коэффициента отвечает за порядковый номер по мере появления фотоотсчета $(*)$. Нижний индекс — это номер позиции текущего фотоотсчета минус единица. Как видно из (13), такая процедура делается «на ходу» по мере появления последовательности по тактам.

Биномиальные коэффициенты вычисляются один раз для всех классов последовательностей и помещаются в таблицу размером $n \times n$.

При появлении первого отсчета в позиции i_1 выбирается число (биномиальный коэффициент) на пересечении $(i_1 - 1)$-й строки и первого столбца матрицы. При появлении второго отсчета берется коэффициент в матрице на пересечении $(i_2 - 1)$-й строки и второго столбца и т. д. По мере поступления фотоотсчетов соответствующие биномиальные коэффициенты из таблицы складываются. В итоге получается номер последовательности $\mathrm{Num}(i_1, i_2, \ldots, i_n)$.

Используемые ресурсы по памяти: таблица размером $n \times n$, в каждой клетке таблицы биномиальный коэффициент — целое число, для бинарного представления которого нужно n бит. В итоге требуемый ресурс по памяти $\mathrm{Memory} = n^3$ бит. В нашем случае (см. детали ниже) $n = 64$, $\mathrm{Memory} = 262144$ бит — ничтожный объем памяти по сравнению с памятью, требуемой при прямой адресации. Такой объем памяти составляет сотые доли процента от памяти стандартной флешки.

Процесс нумерации представляет собой движение по треугольнику Паскаля. Каждой последовательности фотоотсчетов отвечает своя траекто-

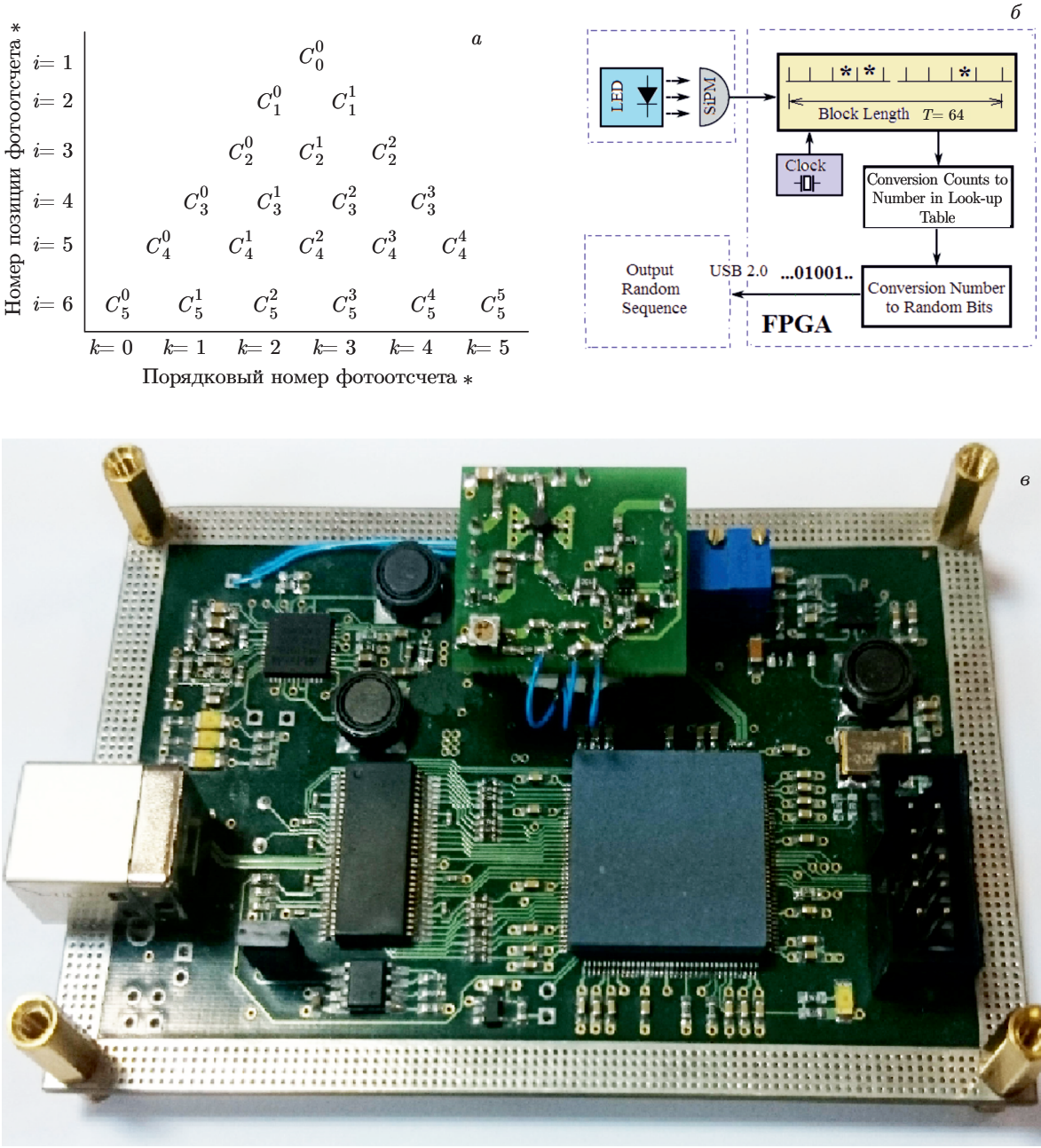

Рис. 1. *а*) Таблица для нумерации последовательностей фотоотсчетов, которая представляет собой известный треугольник Паскаля (см. пример в тексте). *б*) Функциональная схема генератора случайных чисел. *в*) Внешний вид генератора

рия по треугольнику Паскаля. Сказанное поясняется рис. 1*а*.

Пусть для примера $n = 6$. Пусть последовательность фотоотсчетов $\{(i_1 = 1, k = 1), (i_2 = 4, k = 2), (i_5, k = 3), (i_6, k = 4)\} = (100111)$. Номер последовательности вычисляется по таблице, в которой приведены только отличные от нуля коэффициенты: $\mathrm{Num}((100111)) = C_0^1 + C_3^2 + C_4^3 + C_5^4 = 0+3+4+5 = 12$.

4. ПОЛУЧЕНИЕ БЛОКА СЛУЧАЙНЫХ 0 И 1 ИЗ НОМЕРОВ ПОСЛЕДОВАТЕЛЬНОСТЕЙ ФОТООТСЧЕТОВ

Выше для иллюстрации мы использовали простой случай, когда число последовательностей в классе есть степень двойки (2^m). В этом случае бинарное представление номера последовательности сразу дает случайный блок 0 и 1 длиной m бит. В

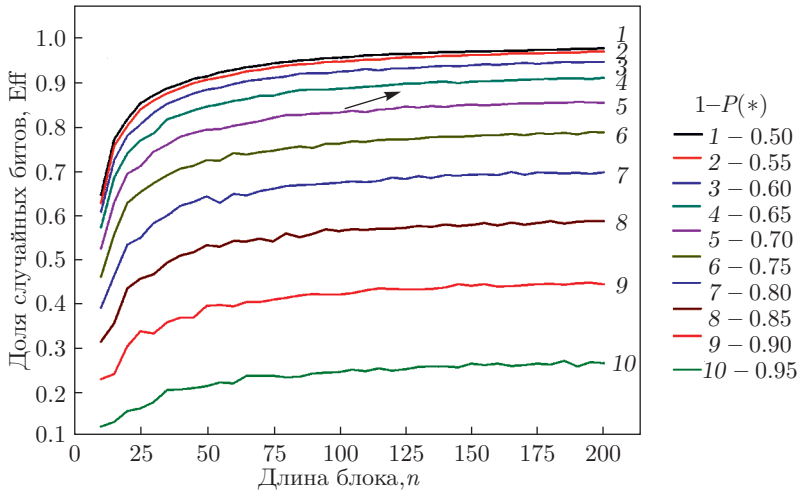

Рис. 2. (В цвете онлайн) Зависимости доли числа случайных битов в зависимости от длины обрабатываемого блока для различных значений вероятности фотоотсчетов за такт $P(*)$

общем случае число последовательностей в классе может быть любым, не обязательно степень двойки. Тогда длина в битах случайного блока 0 и 1, выдаваемого на выход, зависит от того, в каком диапазоне лежит номер текущей последовательности.

Пусть последовательность принадлежит к классу последовательностей с k фотоотсчетами, а полное число последовательностей в классе — N_k. Номера последовательностей находятся в диапазоне $0 \le \mathrm{Num} \le N_k - 1$. Пусть номер текущей последовательности равен Num. Пусть бинарное представление числа последовательностей в классе $N_k = \sum_{i=0}^{i_{max}} 2^{k_i}$. Процедура выполняется рекурсивно. Если номер текущей последовательности Num находится в интервале $2^{k_0} + 2^{k_1} + \ldots + 2^{k_{i-1}} < \mathrm{Num} \le 2^{k_0} + 2^{k_1} + \ldots + 2^{k_{i-1}} + 2^{k_i} - 1$ $(i \le i_{max})$, то выходной случайной последовательностью будет k_i младших разрядов бинарного представления Num. Число номеров последовательностей в этом диапазоне равно 2^{k_i}. Для примера, приведенного в предыдущем разделе имеем. Число элементов в классе $N_4(6) = C_6^4 = 15$, порядковый номер $\mathrm{Num} = 12 = (1100)$ (младшие разряды справа). Номер лежит в диапазоне $2^0 + 2^1 + 2^2 = 7 < \mathrm{Num} \le 2^0 + 2^1 + 2^2 + 2^3 - 1 = 14$, число разрядов $k_3 = 3$, случайным блоком 0 и 1 являются три младших позиции бинарного представления номера $\mathrm{Num} = 12$, а именно, (100).

Нужно знать, при каких длинах блоков последовательностей фотоотсчетов достигается выход на асимптотический режим по извлечению случайности при различных вероятностях фотоотсчета в один такт $P(*)$. Это необходимо знать при выборе экспериментальных длин обрабатываемых последовательностей. Для определения рабочей длины обрабатываемого блока были проведены тестовые вычисления для различных значений вероятностей $*$ и $\sqcup$. В качестве генератора случайности использовался программный генератор псевдослучайных чисел.

Число случайных битов в пересчете на один такт в зависимости от длины блока при разных значениях вероятностей $P(*)$ приведено на рис. 2. Как будет видно ниже, значение $P(*)$ составляет 0.34, поэтому при длине блока в 64 такта практически достигается выход на асимптотику.

5. РЕАЛИЗАЦИЯ КВАНТОВОГО ГЕНЕРАТОРА СЛУЧАЙНЫХ ЧИСЕЛ НА ОСНОВЕ SiPM

5.1. Выбор физической реализации генератора

Основная проблема при фотодетектировании квазиоднофотонных сигналов лавинными детекторами состоит в учете мертвого времени последних, что ограничивает тактовую частоту и темп фотоотсчетов.

Тактовая частота не может превышать обратное время рассасывания лавины. В твердотельных лавинных детекторах имеют место эффекты остаточной пульсации (afterpulsing) после регистрации реального фотона, которые приводят к паразитным отсчетам после регистрации фотона. Данные эффекты нарушают идеальную пуассоновскую статистику фотоотсчетов [25–27].

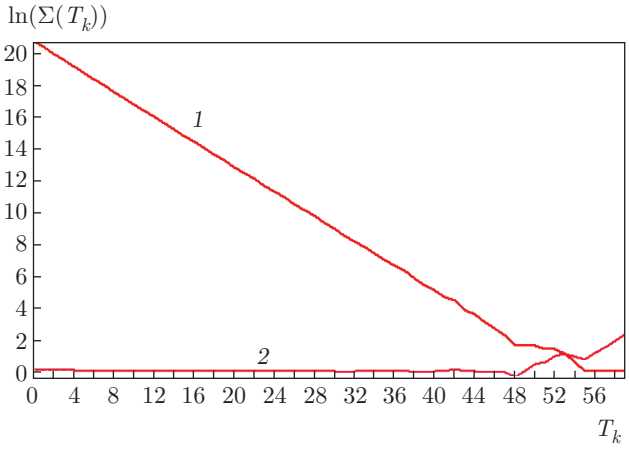

Рис. 3. Зависимость *1* — логарифм количества временных интервалов $\ln(\Sigma(T_k))$ в зависимости от длины интервала T_k, k — число тактов. Метки по оси абсцисс отвечают числу тактов на частоте 120 МГц. Зависимость *2* — логарифм количества временных интервалов с вычитанием наклона. Наблюдаемая вероятность отсчета за один такт $P(*) = 0.34$. Гистограмма строилась с накоплением, полное число событий в гистограмме $\sum_{k=1}^{128} \Sigma(T_k) = 11565488883$. Число отсчетов в первом ящике гистограммы 1013656942, соответственно $\ln(1013656942) = 20.737$. При верхней границе по длительности интервалов в 128 тактов вероятность отсчетов практически равна нулю. Флуктуации при временном интервале между отсчетами при числе тактов больше 64 связаны с редкими событиями из-за экспоненциальной зависимости функции распределения временных интервалов. Длина обрабатываемых блоков для извлечения случайных блоков была выбрана 64 такта

Отметим, что исследование возможного паразитного влияния перекрестных помех (cross-talk) между соседними пикселями на статистику фотоотсчетов проводилось ранее, например, в работе [28]. При используемом среднем числе фотонов искажения статистики выявлено не было (см. рис. 3).

Для устранения паразитных отсчетов можно использовать не одиночные лавинные фотодетекторы, а матрицу лавинных детекторов, содержащую тысячу лавинных детекторов, которая обычно называется кремниевый фотоумножитель (SiPM) [26, 27]. Среднее число фотонов во временно́м окне, определяемом тактовой частотой, составляет не более одной тысячной фотона на пиксел — отдельный детектор в SiPM, поэтому после регистрации фотона конкретным отдельным фотодетектором вероятность того, что следующий фотон попадет в тот же самый детектор, крайне мала. В этом случае мертвое время отдельного фотодетектора не влияет на регистрацию фотонов другими детекторами в матрице, что позволяет увеличить тактовую частоту. Фактически в каждом временно́м окне имеет место только один акт регистрации SiPM, что позволяет достичь контролируемым образом пуассоновской статистики фотоотсчетов, которую можно надежно экспериментально проверить.

Функциональная схема квантового генератора случайных чисел представлена на рис. 1*б*, на рис. 1*в* показан внешний вид генератора. В схеме используется минимальное число элементов. В качестве SiPM использовалась матрица детекторов, технология которой разработана в МИФИ–Пульсар (Москва, Россия) и которая изготовлена в Технологическом центре МИЭТ (Зеленоград, Россия). SiPM имел чувствительную область приблизительно 1×1 мм^2, матрица состояла из $N_{pix} = 1156$ пикселей с активной площадкой $32 \times 32\mu^2$. Рабочее напряжение (несколько вольт выше пробоя) составляет 40 В [27]. Температура детектора стабилизировалась на уровне 25 °C. В качестве источника излучения использовался лазерный светодиод (SLD3143VL) фирмы Sony с рабочей длиной волны 405 нм. Для постобработки при реализации математических алгоритмов использовалась программируемая логическая интегральная схема (ПЛИС – FPGA Intel FPGA (Altera)) с тактовой частотой 120 МГц. В качестве внешнего интерфейса использовался интерфейс USB 2.0 для питания и вывода результирующей случайной последовательности в непрерывном режиме. Преимущество SiPM, который использовался для регистрации фотонов, состоит в том, что он имеет большое сопротивление R_q нагрузочных резисторов, которое превышает 1 МОм. В этой связи SiPM обладает довольно низкой вероятностью остаточной пульсации (afterpulsing), поскольку мертвое время каждого пиксела обратно пропорционально R_q. Во время процесса восстановления пиксела, которое определяется мертвым временем, большая часть носителей заряда освобождается из ловушек, которые были «заселены» в момент рождения лавины носителей, что не позволяет инициировать вторичную зарядку ловушек в данном пикселе. Еще одна очень важная особенность такого SiPM — довольно короткий по времени сигнал с пиксела, который составляет примерно 1 нс.

Следует отметить, что требования к параметрам SiPM, диктуемые применением в криптографических приложениях, отличаются от требований при стандартном использовании SiPM, как, например, в медицинских приложениях в позитронно-эмиссионной томографии.

5.2. Измерение среднего числа фотонов на пиксел в SiPM и проверка статистики фотоотсчетов

Для того чтобы быть уверенным, что генератор действительно работает в квантовом режиме, требуется оценка среднего числа фотонов, падающих на отдельный пиксел SiPM. Эта оценка принципиально важна для того, чтобы быть уверенным, что генератор действительно работает как квантовый и что фотоотсчеты действительно происходят от квазиоднофотонного излучения. Реально наблюдаемой величиной является вероятность отсчета за один такт $P(*) = 0.34$ (см. рис. 3). Данная вероятность равна $P(*) = \mu\eta N_{pic}$, μ — среднее число фотонов на один пиксел в SiPM за один такт, $\eta \approx \approx 0.1$ — квантовая эффективность пиксела, $N_{pic} = = 1156$ — число пикселей в матрице. В итоге получаем $\mu = P(*)/\eta N_{pic} \approx 2.94 \cdot 10^{-3}$ фотонов/(в такт на пиксел), т. е. приходятся примерно тысячные «доли фотона» на пиксел. Для когерентного состояния с пуассоновской статистикой вероятность появления одного фотона равна $P(n=1) = e^{-\mu}\mu \approx \mu$ ($\mu \ll 1$), соответственно, вероятность появления двух фотонов $P(n=2) = e^{-\mu}\mu^2/2 \approx 4.3 \cdot 10^{-6}$. Таким образом, реализован практически однофотонный режим.

Использование SiPM позволяет достичь практически идеальной пуассоновской статистики фотоотсчетов. При пуассоновской статистике интервалы между последовательными отсчетами представляют собой случайную величину, подчиняющуюся геометрическому распределению:

$$P(T_k) = (1 - P(*))^{k-1} P(*), \tag{14}$$

где T_k — число тактов между последовательными моментами регистрации. При идеальной пуассоновской статистике логарифм вероятности $\ln(P(T_k))$ (k — число тактов) должен представлять собой линейную зависимость от k. На рис. 3 показана экспериментальная гистограмма. Как следует из рис. 3, зависимость демонстрирует пуассоновскую статистику. При большом расстоянии между фотоотсчетами вероятность событий крайне мала, поэтому наблюдаются отклонения от прямой на рис. 3. Последнее означает, что длины обрабатываемых последовательностей должны лежать в области участка прямой линии. Нами была выбрана длина в $n = 64$ такта, где можно гарантировать пуассоновский характер статистики фотоотсчетов.

Достигнутая скорость генерации результирующей случайной последовательности 0 и 1 составляла 99.82 Мбит/с.

Интересно сравнить экспериментальную скорость генерации случайных чисел с данными теоретических расчетов (см. рис. 2). При значении $P(*) = 0.34$ число случайных битов на один такт (кривая *4* на рис. 2, отмечена стрелкой) — эффективность экстракции — равно Eff ≈ 0.832 бит. При рабочей тактовой частоте $f = 120$ МГц скорость генерации случайных битов есть $0.832 \cdot 120$ Мбит/с $= = 99.84$ Мбит/с, что с хорошей точностью совпадает с экспериментально наблюдаемым значением.

В асимптотическом пределе, когда длина обрабатываемого блока $n \to \infty$, теоретический предел по скорости генерации случайной последовательности есть $h(P(*))f = 110.978$ Мбит/с ($h(p) = -p\log(p) - -(1-p)\log(1-p)$ — бинарная энтропийная функция Шеннона, log берется по основанию 2). Таким образом, достигнутая скорость по генерации случайной последовательности лишь примерно на 10 % не доходит до теоретического предела, т. е. извлекается практически вся случайность, которая содержится в исходном физическом процессе.

6. СТАТИСТИЧЕСКИЕ ТЕСТЫ СЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

В открытой печати имеется несколько наборов тестов на случайность [29–32]. Для проверки статистических свойств случайных последовательностей нами был выбран стандартный набор тестов NIST [32]. Данный набор тестов является минимально необходимым и является основанием для исследования последовательностей другими наборами специальных тестов. Кратко остановимся на идеологии проверки последовательности на случайность. Как уже упоминалось выше, нельзя доказать, что данная последовательность произошла из источника истинной случайности, можно лишь утверждать, что она не противоречит гипотезе случайности по некоторому статистическому критерию.

Проверяется гипотеза H_0 о том, что последовательность является истинно случайной. При этом предположении различные статистики S (группировки 0 и 1) также являются случайными величинами, распределения вероятностей различных статистик при длине последовательности $n \to \infty$ должны стремиться к некоторым эталонным распределениям для случайной последовательности. Фиксируется некоторый уровень значимости α и пороговое значение для каждой статистики. Если вероятность уклонения статистики превышает пороговое значение, то гипотеза о случайности отклоняется. Но это еще не

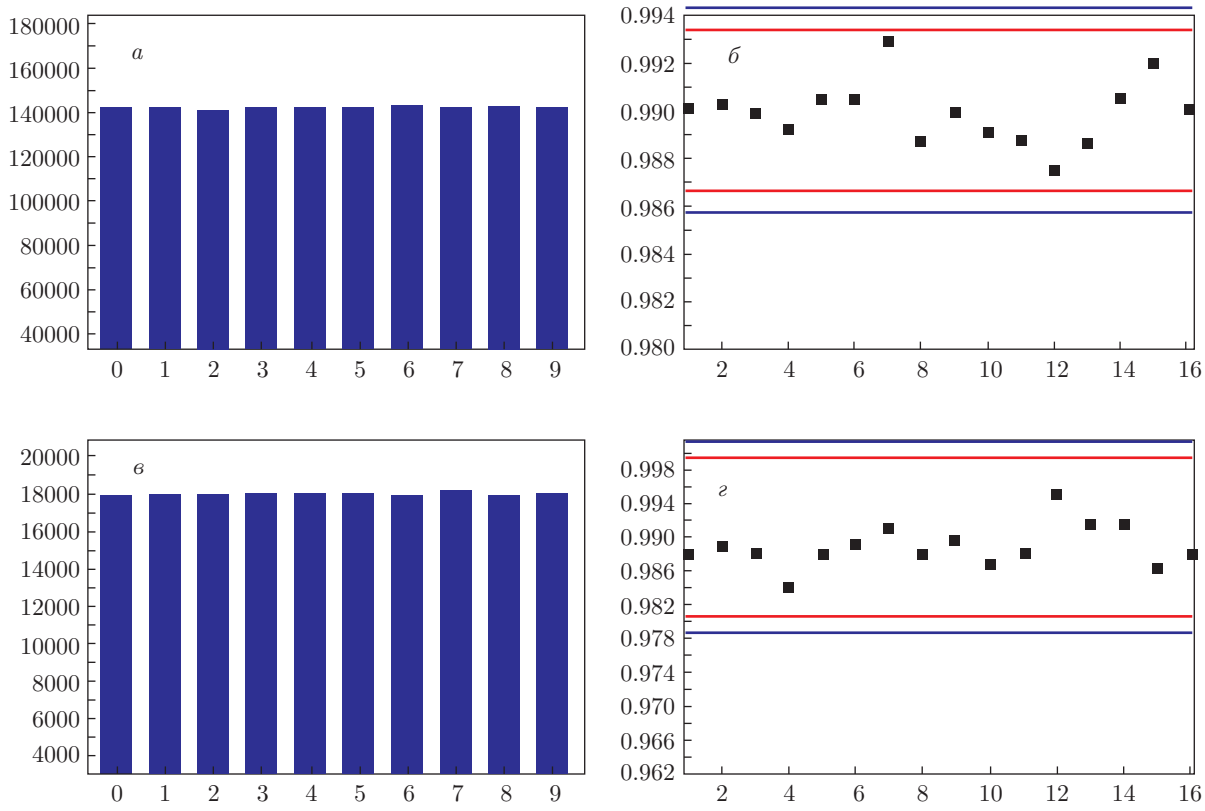

Рис. 4. (В цвете онлайн) Гистограмма значений ν_j (P-value), по всем тестам попадающих в 10 интервалов $j = 0, \ldots, 9$. Интервалу $[0.0, 0.1]$ отвечает значение $j = 0$ по оси абсцисс, интервалу $[0.1, 0.2]$ — значение $j = 1$ по оси абсцисс и т. д. По оси ординат — доля последовательностей, которые прошли тесты по значениям P-value. По оси абсцисс номера отвечают различным тестам по номенклатуре NIST [32]. Красными горизонтальными линиями показаны верхняя и нижняя границы («три сигма») соответственно для всех тестов, кроме № 13 (Random Excursions Test) и № 14 (Random Excursions Variant Test) (см. таблицу). Для данных двух тестов число тестируемых последовательностей N не определено заранее, а определяется в процессе тестирования. Верхняя и нижняя границы («три сигма») для данных двух тестов показаны синими горизонтальными линиями. Для остальных тестов число тестируемых последовательностей $M = 8000$ каждая длиной $1 \cdot 10^6$ бит (рис. *а,б*), $M = 1000$ каждая длиной $2 \cdot 10^6$ бит (*в,г*)

означает, что последовательность действительно не случайная. Это означает лишь то, что даже генератор идеальных случайных последовательностей может выдать последовательность, которая имеет такое уклонение.

Далее, подсчитывается вероятность P-value — вероятность того, что даже идеальный случайный источник может выдать последовательность с таким уклонением статистики. Если $P > \alpha$, то гипотеза H_0 принимается, при $P < \alpha$ гипотеза отклоняется, последовательность считается не случайной.

6.1. *P*-values для различных тестов

Приведем интерпретацию значения P-value. При заданном уровне значимости α значения P-value имеется вероятность того, что даже идеальный генератор может сгенерировать с такой вероятностью последовательность, которая будет выглядеть как не случайная для данного теста. Чем меньше P-value, тем с меньшей вероятностью идеальный генератор «имеет право» сгенерировать такую последовательность. Если вычисленное P-value больше α, то тест считается пройденным. Стандартное значение уровня значимости для α: $\alpha \in [0.001, 0.01]$ [32]. Было использовано значение $\alpha = 0.01$. Были проведены тесты с разными длинами и числом тестируемых последовательностей:

1) число тестируемых последовательностей $M = 8000$, длина каждой последовательности равнялась $L = 1 \cdot 10^6$ бит (рис. 4*а,б*);

2) число тестируемых последовательностей $M = 1000$, длина каждой последовательности равнялась $L = 2 \cdot 10^6$ бит (рис. 4*в,г*).

Таблица. Доля последовательностей, прошедших различные тесты

№	Название теста	Доля послед. $M = 8000$, $L = 1 \cdot 10^6$	Доля послед. $M = 1000$, $L = 2 \cdot 10^6$
1	Frequency Test	0.9901	0.9880
2	Block Frequency	0.9902	0.9886
3	Cumulative Sums	0.9899	0.9880
4	Cumulative Sums Reverse	0.9892	0.9840
5	Runs	0.9905	0.9880
6	Longest Runs	0.9905	0.9886
7	Rank	0.9929	0.9910
8	FFT Fast Fourier Transform	0.9888	0.9879
9	Non Overlapping Template	0.9899	0.9893
10	Overlapping Template	0.9891	0.9867
11	Universal	0.9987	0.9880
12	Approximate Entropy	0.9874	0.9950
13	Random Excursions	0.9883	0.9914
14	Random Excursions Variant	0.9904	0.9915
15	Serial	0.9921	0.9860
16	Linear Complexity	0.9901	0.9880

Доля последовательностей, прошедших тесты, сама является случайной величиной. Допустимый диапазон флуктуаций определяется дисперсией P-value. P-value являются случайными величинами с распределением Бернулли с двумя исходами: один исход — тест пройден, второй исход — тест не пройден. Допустимый разброс P-value должен укладываться в интервал «три сигма». Величина дисперсии для P-value есть $\sqrt{P(1-P)/M}$ (M — число тестируемых последовательностей). Согласно [32], при уровне значимости $\alpha = 0.01$ все P-value должны попадать в интервал «три сигма»:

$$1 - P \pm 3\sqrt{\frac{P(1-P)}{M}} = 0.99 \pm 3\sqrt{\frac{0.99 \cdot 0.01}{M}}.$$

1. Интервал «три сигма» оказывается равным $[0.987, 0.993]$ ($M = 8000$).

2. Интервал «три сигма» оказывается равным $[0.981, 0.999]$ ($M = 1000$).

Как видно из таблицы и рис. 4, доля последовательностей, прошедших тесты, укладывается в доверительный интервал «три сигма» с хорошим запасом. По данному критерию гипотеза о происхождении последовательностей из случайного источника является справедливой.

6.2. Однородность значений P-values

Напомним, что P-value само является случайной величиной. Поэтому частота появлений значений P-values для различных тестов при большом объеме выборки распределена по нормальному закону. Если величины распределены по нормальному закону, статистика, которая приводится ниже, имеет распределение Пирсона (см., например, [33]).

Приведем тест на однородность значений P-value. При большом числе тестируемых последовательностей суммарное значение P-value по всем тестам есть сумма одинаково распределенных случайных величин, которая распределена по гауссовому нормальному закону [32]. Определим статистику

$$X_N^2 = \sum_{j=1}^{N} \frac{(\nu_j - Np_j)^2}{Np_j},$$

где ν_j — доля значений P-value, попадающих в j-й интервал [0,1], p_j — истинная вероятность попадания в j-й интервал. При большом числе тестируемых последовательностей распределение вероятностей данной статистики X_N^2 не зависит от распределения p_j входящих в нее величин и стремится к

распределению Пирсона $\chi^2(N-1)$ с $(N-1)$-й степенью свободы [33]. Рекомендуемое тестами NIST число интервалов равно $N = 10$ [32]. Пороговое значение при нулевой гипотезе (H_0-последовательность случайна) допустимого разброса получается при заданном уровне значимости α из соотношения $\Pr\{X_N^2 > t_\alpha | H_0\} = \alpha$, где $t_\alpha = \chi^2_{1-\alpha,N-1}$ — $(1-\alpha)$-я квантиль распределения χ^2 с $(N-1)$-й степенью свободы. В нашем случае $N = 10$ — число интервалов для значений P-value. Иными словами, если уклонение статистики $X_N^2 > \chi^2_{1-\alpha,N-1}$, то нулевая гипотеза H_0 отвергается, последовательность считается не случайной, поскольку уклонение статистики от допустимой нормы (при заданной вероятности — уровне значимости) превышено. Если P-value от P-value

$$\hat{P} = \frac{\int\limits_{X_K^2}^{\infty} dx\, e^{-x/2} x^{K/2-1}}{2^{K/2}\Gamma(K/2)}, \quad K = N-1,$$

не менее $\hat{P} > 0.0001$, то тест на однородность P-value считается пройденным и принимается гипотеза H_0. На рис. 4 приведены результаты теста на равномерность P-value. Значение P-value от P-value равно 1) P-value $= 0.01333485$ для $L = 8000$; 2) P-value $= 0.88110029$ для $L = 1000$. Значения P-value от P-value должны быть больше критического $\hat{P}_c > 0.0001$.

Тест на равномерность P-value с запасом пройден (рис. 4). Доля последовательностей, прошедших тест, показана на рис. 4*б,в* для всех тестов (номера тестов NIST [32] приведены по горизонтали). Для всех тестов доля последовательностей, прошедших тесты по величинам P-value, лежит в пределах «три сигма», что означает успешное прохождение тестов, и гипотеза о случайности последовательности принимается.

Один из авторов (С. Н. М.) выражает благодарность коллегам из Академии криптографии Российской Федерации за обсуждения и поддержку. Выражаем благодарность Елене Поповой и Сергею Виноградову за любезно предоставленные образцы SiPM и обсуждения. Авторы также благодарят И. М. Арбекова за обсуждения.

Работа выполнена при финансовой поддержке Российского научного фонда (грант №16-12-00015).

ЛИТЕРАТУРА

1. F. Galton, *Natural Inheritance*, Macmillan (1894).

2. M. Herrero-Collantes and J. Carlos Garcia-Escartin, Rev. Mod. Phys. **89**, 015004 (2017).

3. C. Gabriel, C. Wittmann, D. Sych, R. Dong, W. Mauerer, U. L. Andersen, C. Marquardt, and G. Leuchs, Nature Photonics Lett. **4**, 771 (2010).

4. Jie Yang, Jinlu Liu, Qi Su, Zhengyu Li, Fan Fan, Bingjie Xu, and Hong Guo, Optic Express **24**, 27475 (2016).

5. X. Ma, F. Xu, H. Xu, X. Tan, B. Qi, and H.-K. Lo, Phys. Rev. A **87**, 062327 (2013).

6. D. Frauchiger, R. Renner, and M. Troyer, arXiv:1311.4547, 2013.

7. D. Stucki, S. Burri, E. Charbon, C. Chunnilall, A. Meneghetti, and F. Regazzoni, Proc. SPIE **8899** (2013).

8. Y. Okawachi, M. Yu, K. Luke, D. O. Carvalho, M. Lipson, and A. L. Gaeta, Opt. Lett. **41**, 4194 (2016).

9. F.-X. Wang, C. Wang, W. Chen, S. Wang, F.-S. Lv, D.-Y. He, Z.-Q. Yin, H.-W. Li, G.-C. Guo, and Z.-F. Han, J. Light Wave Techn. **33**, 3319 (2015).

10. Q. Yan, B. Zhao, Z. Hua, Q. Liao, and H. Yang, Rev. Sci. Instr. **86**, 073113 (2015).

11. Y.-Q. Nie, H.-F. Zhang, Z. Zhang, J. Wang, X. Ma, J. Zhang, and J.-W. Pan, Appl. Phys. Lett. **104**, 051110 (2014).

12. K. S. Kravtsov, I. V. Radchenko, S. P. Kulik, and S. N. Molotkov, J. Opt. Soc. Amer. **32**, 1743 (2015); arXiv:1507.02059.

13. И. В. Радченко, Дисс.... канд. физ.-матем. наук, Институт общей физики им. А. М. Прохорова РАН, Москва (2015).

14. Д. Н. Клышко, *Фотоны и нелинейная оптика*, Наука, Москва (1980).

15. Д. Н. Клышко, А. В. Масалов, УФН **165**, 1249 (1995).

16. L. Trevisan, J. ACM **48**, 860 (2001).

17. W. Mauerer, C. Portmann, and V. B. Scholz, arXiv: 1212.0520.

18. R. König, R. Renner, and C. Schaffner, IEEE Trans. Information Theory **55**, 4337 (2009).

19. J. von Neumann, Appl. Math. Ser. **12**, 36 (1951), Washington: U. S. National Bureau of Standarts.

20. C. E. Shannon, Bell System Techn. J. **XXVII**, 379 (1948).

21. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley (1991).

22. С. Н. Молотков, Письма в ЖЭТФ **105**, 374 (2017).

23. Л. Д. Ландау, Е. М. Лифшиц, *Статистическая физика*, т. V, ч. I, Наука, Москва (1995).

24. В. Ф. Бабкин, Проблемы передачи информации **7**, 13 (1971).

25. S. Vinogradov, Nucl. Instr. Meth. Phys. Res. A **695**, 247 (2012).

26. M. Danilov, Nucl. Instr. Meth. A **582**, 451 (2007).

27. P. Buzhan, B. Dolgoshein, L. Filatov, A. Ilyin, V. Kaplin, A. Karakash, S. Klemin, R. Mirzoyan, A. N. Ottec, E. Popova, V. Sosnovtsev, and M. Teshimac, Nucl. Instr. Meth. Phys. Res. A **567**, 78 (2006).

28. D. A. Kalashnikov, Si-Hui Tan, and L. A. Krivitsky, Opt. Express **20**, 5044 (2012).

29. D. E. Knuth, *The Art of Computer Programming*, Addison Wesley, Cambridge (1981), Vol. 2.

30. G. Marsaglia, http://stat.fsu.edu/pub/diehard, *Die Hard*: *A Battery of Tests for Random Number Generators*.

31. P. L'Ecuyer and R. Simard, http://www.iro.umontreal.ca/lecuyer, *TestU01*: *A Software Library in ANSI C for Empirical Testing of Random Number Generators* (2002).

32. http://csrc.nist.gov/rng/SP800-22b.pdf, *A Statistical Test Suite for the Validation of Random Number Generators and Pseudo Random Number Generators for Cryptographic Applications*.

33. H. Cramer, *Mathematical Methods of Statistics*, Asia Publ. House (1946).