Июнь 2021 г. *Том 191, № 6*

УСПЕХИ ФИЗИЧЕСКИХ НАУК

МЕТОДИЧЕСКИЕ ЗАМЕТКИ

Об экстракции квантовой случайности

И.М. Арбеков, С.Н. Молотков

Обсуждаются вопросы о природе случайности, конструктивных и доказуемых способах её получения (экстракции) из наблюдений над физическими системами. Истинная случайность существует только в микромире — при квантово-механическом описании физических систем и измерений над ними — и является фундаментальным свойством квантовых систем. При классическом описании физических систем случайность отсутствует и фактически вводится "вручную" через неопределённость — неизвестность начальных условий. Обсуждается, как можно реально "дотянуться" до квантовой случайности на примере квантового устройства — генератора случайных чисел. Рассматриваются также вопросы, связанные с "доказательством" случайности — тестированием числовых последовательностей, проводится анализ логических построений, лежащих в основе такого тестирования. При этом используется необходимый математический аппарат, который не требует специальных предварительных знаний. Для понимания достаточно стандартных сведений из университетских курсов по квантовой механике и теории вероятностей. Цель, которую ставили перед собой авторы, — провести единую логическую линию от происхождения случайности в квантовой области до её экстракции, физической реализации и тестирования.

Ключевые слова: квантовые генераторы случайных чисел, экстракция случайности

PACS numbers: 03.67.Dd, 42.50.Ex DOI: https://doi.org/10.3367/UFNr.2020.11.038890

Содержание

И.М. Арбеков [1], **С.Н. Молотков** [1, 2, 3, 4, a]

[1] Академия криптографии Российской Федерации, ул. Ярцевская 30, 121552 Москва, Российская Федерация

[2] Институт физики твёрдого тела РАН, ул. Академика Осипьяна 2, 142432 Черноголовка, Московская обл., Российская Федерация

[3] Московский государственный университет им. М.В. Ломоносова, факультет вычислительной математики и кибернетики, Ленинские горы 1, стр. 52, 119991 Москва, Российская Федерация

[4] Московский государственный университет им. М.В. Ломоносова, Центр квантовых технологий, Ленинские горы 1, стр. 35, 119991, Москва, Российская Федерация

E-mail: [a] molotkov@issp.ac.ru

Статья поступила 15 мая 2020 г., после доработки 28 октября 2020 г.

Самое удивительное в этом мире то, что он познаваем.
А. Эйнштейн

1. Введение. О природе случайности и основах построения квантовых генераторов случайных чисел

Случайные числа широко используются в различных областях науки и техники, например, при моделировании физических процессов методом Монте-Карло. Со случайными числами в повседневной жизни сталкивается любой человек — это компьютерные пароли доступа, PIN-коды смарт-карт и других электронных устройств.

Особенно широкое применение случайные числа находят в криптографии. Генератор случайных чисел (ГСЧ) является составной частью систем криптографической защиты информации, его качество во многом определяет криптостойкость таких систем.

Шифрование больших массивов информации требует частой смены секретных ключей, вырабатываемых с использованием генераторов случайных чисел. Если ключ меняется часто, то объёмы информации, которые шифруются на отдельных ключах, будут невелики, это делает секретную связь надёжной.

В *классических системах* симметричного шифрования секретные ключи меняются на передающей и приёмной сторонах с помощью устройств ввода с цифровых носителей ограниченного объёма. Это в свою очередь ограничивает скорость смены ключей, поскольку требует регулярной замены носителей ключевой информации. Частая смена ключей в течение *всего срока эксплуатации криптосистемы* становится практически невозможной. Поэтому в *классических системах* секретные ключи, как правило, используются в виде мастер-ключей для получения производных от них *сеансовых* ключей, что в целом не обеспечивает высокой криптостойкости. Частую смену секретных ключей способны обеспечить современные системы *квантовой криптографии*, но при этом выработка каждого секретного ключа требует большого объёма случайных чисел.

Насколько большого?

Система квантовой криптографии представляет собой распределённую систему согласования и хэширования в секретный финальный ключ случайных битовых последовательностей на передающей и приёмной сторонах, образованных с помощью передачи по оптическому каналу квантовых квазиоднофотонных (в идеале однофотонных) состояний [1].

Для иллюстрации оценим объём случайных чисел — случайных бит, требуемых для получения секретного ключа при длине оптоволоконной линии, например, 100 км. В стандартном оптическом волокне с удельными потерями $\approx 0{,}2$ дБ км$^{-1}$ расстояние в 100 км до приёмника способен преодолеть в среднем один из 10^2 фотонов.

Поскольку не существует строго однофотонных источников, вместо них используется сильно ослабленное когерентное излучение лазера. Когерентное состояние является суперпозицией фоковских состояний с числом фотонов $k = 0, 1, \ldots$, с соответствующими весами, в состоянии задано лишь среднее число фотонов μ. Ослабление когерентного состояния до значений среднего числа фотонов в импульсе на уровне $\mu \approx 0{,}1$ приводит к тому, что примерно лишь один из 10 импульсов излучения содержит однофотонное фоковское состояние, в остальных девяти импульсах присутствует вакуумное состояние поля.

Регистрация квазиоднофотонных состояний происходит с использованием лавинных фотодетекторов, эффективность которых значительно меньше единицы. Типичные значения квантовой эффективности однофотонных лавинных фотодетекторов составляют величину $\eta \approx 0{,}1$, что ещё примерно в 10 раз снижает скорость создания секретного ключа.

Обеспечение криптографической *секретности* общего финального ключа длиной в 256 бит требует обработки и сжатия (хэширования) случайных битовых последовательностей на передающей и приёмной сторонах длиной примерно в 10^4 бит.

В итоге на генерацию одного секретного ключа в 256 бит требуется случайная последовательность с выхода ГСЧ длиной не менее чем

$$10^2(\text{потери в линии}) \times 10^1(\mu) \times 10^1(\eta) \times \\ \times\, 10^4(\text{хэширование}) = 10^8 \text{ бит}.$$

Данный пример показывает, что для реализации систем квантовой криптографии требуются генераторы случайных чисел с высокой скоростью генерации и доказуемой "случайностью" выходной последовательности.

Случайность на интуитивном уровне как некоторый процесс, в котором каждый следующий шаг непредсказуем, представляется достаточно понятной, однако при более детальном рассмотрении оказывается, что понятие *случайности* оказывается далеко не тривиальным. Откуда взять "хорошую" случайность, точнее *истинную* случайность, и как определить, что случайность является хорошей и, как максимум, можно ли получить *истинную* случайность?

Принципиально важно, что при разработке генераторов случайных чисел недостаточно того обстоятельства, что вырабатываемые ими последовательности проходят тестирование на случайность по некоторому критерию. Это лишь необходимое условие. Принципиально важен источник первичной случайности, который используется для получения равнораспределённой последовательности из 0 *и* 1 *и который был бы действительно источником случайности по соображениям, не зависящим от рекомендуемых тестов* (*например, источник случайности как процесс измерений над квантовой системой* (*см. ниже формулы* (3)–(8))). *Многие псевдослучайные генераторы случайных чисел, типичными представителями которых являются регистры сдвига с обратной связью, проходят тестирование, но не являются истинно случайными.*

Необходимые сведения об "идеологии" и методах тестирования конечных битовых последовательностей на случайность изложены в разделах 11 и 12, которые являются достаточно самостоятельными и могут быть рекомендованы для прочтения мало знакомому с предметом читателю независимо от других разделов.

Обобщая вышесказанное, мы приходим к выводу, что, строго говоря, само понятие *случайности* требует дополнительного математического определения. Обсудим сначала ситуацию на качественном уровне.

Случайные числа возникают как результат работы ГСЧ. Генераторы [1] случайных чисел можно разделить на два класса: *математические* и *физические*. Сразу отметим, что для генерации ключей в системах симметричного шифрования, претендующих на *высокую* криптографическую стойкость, используются исключительно *физические* ГСЧ.

Математические генераторы представляют собой преобразование, обычно рекурсивное:

$$x_i = \mathcal{F}(x_{i-1}) = \mathcal{F}\big(\mathcal{F}\big(\mathcal{F}(\ldots\mathcal{F}(x_0))\big)\big)\,, \tag{1}$$

где $\mathcal{F}$ — некоторая функция, x_0 — начальное значение (затравка), которое выбирается "вручную".

Является ли последовательность чисел $\{x_i\}$ случайной? Очевидно — нет, поскольку, если известна затравка и само преобразование $\mathcal{F}$, то известна вся последовательность. Именно поэтому такие генераторы называются псевдослучайными, так как они полностью зависят от начальных условий: вся "случайность" сосредоточена в неизвестном "затравочном" значении x_0 и отчасти в преобразовании $\mathcal{F}$, возможно, скрываемом от нелигитимной стороны. Таким образом, математическое преобразование не может дать *истинной* случайности.

Физические генераторы основываются на измерении состояния физической системы, и их также можно разделить на два типа: *классические* и *квантовые*.

[1] В русскоязычной литературе по криптографии и различных документах чаще вместо "генератор" используется словосочетание "датчик случайных чисел".

Первый тип генераторов — *классические*.

Для классических генераторов выходная битовая последовательность представляет собой функцию от реально наблюдаемых физических величин, порождаемых недетерминированной физической системой. С позиций аксиоматического построения теории вероятностей для строгого доказательства случайности (независимости и равновероятности) выходной последовательности требуется задание исходного вероятностного пространства, на котором реально наблюдаемые физические величины были бы случайными величинами (функциями от элементов пространства) и посредством соответствующих преобразований генератора случайных чисел порождали бы или доказуемо независимую и равновероятную битовую последовательность, или близкую к ней в пределах установленных теоретико-вероятностных требований [2].

Если теперь отвлечься от аксиоматических теоретико-вероятностных предположений и в свою очередь предположить, что система эволюционирует по законам классической физики, т.е. эволюция описывается дифференциальными уравнениями, то *случайность* результата измерения будет связана только с неизвестностью начальных условий. Как и выше, можно сказать, что последовательность результатов измерений в этом случае является псевдослучайной, поскольку она определяется при известном законе эволюции только неопределённостью начальных условий.

Часто утверждается в пользу классических физических систем, что их эволюция является сложной, а траектории, отвечающие близким начальным условиям, экспоненциально быстро расходятся в фазовом пространстве. Тем не менее это всё равно траектории, и если начальные условия известны, то траектории предсказуемы точно.

Хорошим примером, иллюстрирующим "детерминизм" классических систем, является "доска Гальтона" [2] для демонстрации закона нормального распределения вероятностей как результата применения центральной предельной теоремы [3].

Доска Гальтона — система с твёрдыми металлическими шариками, падающими из центра верхней части доски через большое количество тонких штырьков, расположенных ниже в шахматном порядке (рис. 1). Система чисто классическая. Падая вниз, шарик испытывает упругие отражения (отклонения) в ту или другую сторону от встретившихся ему на пути штырьков и в итоге попадает в один из ящиков, расположенных внизу по горизонтали. Итоговое смещение по горизонтали трактуется как сумма отдельных случайных отклонений (как сумма большого числа случайных величин), вероятностное распределение которой, по центральной предельной теореме, должно быть нормальным. Это подтверждается визуальным сходством итоговой картинки распределения шариков по ящикам (гистограммы) с плотностью нормального распределения.

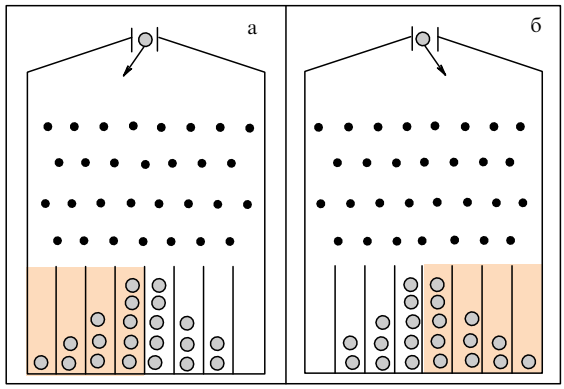


Рис. 1. Доска Гальтона — пример классической физической системы, иллюстрирующий появление "случайности" при неизвестных начальных условиях. Умозрительный пример экстракции "случайности": (а) присваивается "случайный" бит 0, если после бросаний шаров их больше в левой половине ячеек, (б) присваивается "случайный" бит 1, если после бросаний шаров их больше в правой половине ячеек. Если такой генератор "случайных" 0 и 1 работает как чёрный ящик (на выходе доступно только распределение шаров по ящикам, но не начальные условия), то злоумышленник может контролировать первичную случайность по своему усмотрению (задавать начальное положение и скорость шаров) и последовательность 0 и 1 будет скомпрометирована, а ключ будет целиком известен — не секретен.

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

Траектория каждого шарика и его финальное положение, а именно ящик, в котором он окажется, могут быть достоверно предсказаны, если известны угол и скорость, под которыми шарик входит в первый ряд. Небольшие, но известные отклонения в начальном угле падения и начальной скорости каждого шарика приводят к разбеганию траекторий и, в конечном итоге, к распределению по ящикам, похожему на нормальное.

Подчеркнём ещё раз, что если все начальные углы и скорости известны, то всё распределение по ящикам однозначно предсказуемо. Видимая "случайность" связана только с недостоверностью начальных условий.

Важно, и это свойство любой классической системы, что если она приготовлена в начальный момент в одних и тех 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 (может быть вычислена и предсказана) при известных начальных условиях.

В данном примере "выход" доски Гальтона не является *истинно* случайным. Но если заранее неизвестно происхождение "случайности", т.е. то, что распределение шариков по ящикам задаётся *известными* начальными условиями, то можно принять такой источник за истинно случайный. Этот пример также показывает, почему важно знать и контролировать источник первичной случайности.

[2] Такой, на наш взгляд, вполне естественный подход к обоснованию случайности выходной последовательности физических генераторов случайных чисел, видимо, из-за сложности поставленной задачи, не нашёл отражения в научных, включая обзорные, статьях по данной тематике (см., например, [2]). Авторы публикаций ограничиваются, как правило, вопросами практической реализации, используя непредсказуемость того или иного физического процесса (например, явления джиттера (от англ. jitter — дрожание) [3], метастабильности [4]) как основу построения ГСЧ.

[3] Хотя изначально доска Гальтона использовалась для иллюстрации в задачах наследования генетических признаков [5].

Особое внимание к происхождению первичной случайности, извлекаемой из физических генераторов, уделяется в системах симметричной криптографии.

В условиях, когда не найдено криптографических методов снижения стойкости, использующих алгоритмические слабости шифрующего преобразования, стойкость криптосистемы определяется исключительно *секретностью* ключа, т.е. тем, насколько выбор ключа из ключевого множества близок к случайному и равновероятному выбору.

Умозрительно можно представить ситуацию получения равновероятного битового ключа, например, с использованием редукции выборочных значений (финальных отклонений) доски Гальтона: если выборочное значение больше среднего значения, то полагается, что бит равен 1, если меньше, то бит равен 0. Но если генератор работает как чёрный ящик, т.е. выдаёт на выход только распределение шариков без контроля начальных данных, а злоумышленник способен управлять ими, то, очевидно, полученный таким образом ключ может быть скомпрометирован (см. также пояснения к рис. 1).

Таким образом, выход любого генератора случайных чисел, использующего в качестве первичной случайности измерения над классической системой, нельзя считать истинно случайным.

Само понятие вероятности отсутствует в классической физике в том смысле, что математическая теория вероятностей представляет собой отдельную математическую дисциплину, никоим образом не "привязанную" к классической физике. Точнее говоря, теория вероятностей представляет собой внешнюю теорию по отношению к классической физике; теорию, которая привносится в классическую физику "вручную" явно или неявно через неопределённость начальных условий. Законы классической физики верны в условиях "макрообъектов" и при уменьшении размеров системы становятся несправедливыми.

Поясним это на примере доски Гальтона. Конечно, положения вбитых штырьков испытывают нулевые флуктуации, что при строгом учёте должно приводить к отклонению от траектории, рассчитанной "классическим" образом. Будет ли такой вклад в "классическую" траекторию существенным? Неточность скорости по сравнению с классическими оценками после столкновения со штырьком из-за его положения Δx можно оценить из соотношения неопределённостей

$$\Delta v \approx \frac{\hbar}{\Delta x} \approx 10^{-27} \text{ см с}^{-1} \qquad (2)$$

при массе штырька в 1 г. Такой вклад квантовых эффектов, разумеется, нельзя заметить. При уменьшении размеров системы пренебрегать этим вкладом становится всё труднее.

Перейдём к рассмотрению физических генераторов второго типа — квантовых (см., например, обзор [6]). Извлечение случайности в таком генераторе основано на измерении квантовой системы.

В отличие от измерений в классической физике, измерения над квантовой системой, каждый раз приготовленной в определённом и одном и том же состоянии, дают случайный результат, что является фундаментальным законом природы в микромире. Поэтому истинно случайными могут быть только квантовые генераторы случайных чисел.

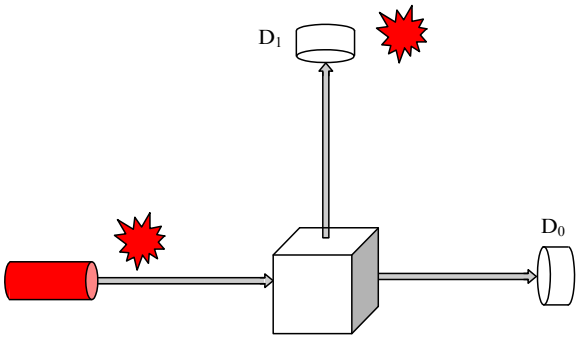


Рис. 2. Наглядный пример квантовой системы, иллюстрирующий принципиальную непредсказуемость результатов измерений над квантовой системой.

Вообще говоря, эволюция квантовой системы также описывается дифференциальными уравнениями и зависит от начальных условий. Однако даже при известных начальных условиях принципиально невозможно предсказать исход измерений над квантовой системой. При одних и тех же начальных условиях, одной и той же эволюции квантовой системы результат наблюдения или измерения принципиально непредсказуем. Это является главным отличием квантовых систем от классических. *В классической физике нет фундаментальных запретов на то, чтобы измерять состояние классической системы без её возмущения.*

Обычно в качестве умозрительного примера, иллюстрирующего принципиальную непредсказуемость результата измерения над квантовой системой, приводят следующий пример (рис. 2).

Пусть источник каждый раз испускает одинаковый однофотонный пакет, который попадает на симметричный светоделитель 50/50, за которым расположены два детектора, D_0 и D_1. Возможен отсчёт *только одного* детектора, причём при одних и тех же начальных условиях — приготовлении однофотонного пакета — и его эволюции принципиально невозможно предсказать, какой из детекторов сработает. Истинная случайность имеет место только в квантовой области, в ней вероятность *встроена* в аппарат квантовой механики, в отличие от вероятности в классической физике, в которую она привносится извне.

Уточним, что имеется в виду.

Результат измерений над квантовой системой, находящейся в состоянии $|\psi\rangle$, сводится к проецированию состояния системы на одно из состояний базиса измерений $\{|\phi_i\rangle\}_{i=1}^N$, где $\{|\phi_i\rangle\}_{i=1}^N$ — ортонормированные состояния, N — число исходов измерения (ограничимся только ортогональными измерениями фон Неймана). Квадрат модуля скалярного произведения

$$P_\psi(i) = \left|\langle\phi_i|\psi\rangle\right|^2 \qquad (3)$$

интерпретируется как вероятность. Фактически это отражает борновскую интерпретацию квадрата модуля вектора состояний — волновой функции. Сумма вероятностей по всем исходам измерений равна единице:

$$\sum_{i=1}^N P_\psi(i) = \sum_{i=1}^N \left|\langle\phi_i|\psi\rangle\right|^2 = \sum_{i=1}^N \langle\psi|\phi_i\rangle\langle\phi_i|\psi\rangle = \\ = \langle\psi|\left(\sum_{i=1}^N |\phi_i\rangle\langle\phi_i|\right)|\psi\rangle = \langle\psi|(I)|\psi\rangle = 1\,, \qquad (4)$$

где I — единичный оператор. В этом смысле вероятность встроена в аппарат квантовой механики.

Таким образом, измерение квантовых систем даёт вероятность, встроенную в сам процесс измерений, причём результат измерений нельзя предсказать принципиально, в отличие от результата измерений классических систем.

Важный шаг в построении квантового генератора случайных чисел — это найти подходящую квантовую систему и способ измерения над ней, чтобы в максимально "чистом" виде извлечь квантовую случайность. Такими квантовыми процессами могут быть, например, α-распад, фотоэффект и т.д. Следует отметить, что квантовые эффекты для создания генераторов случайных чисел в криптографии использовались и ранее, например, на основе источников радиоактивного излучения и т.п. Получившиеся генераторы были технически несовершенными и медленными. Трудно было совместить противоречивые требования: сохранить квантовость процесса и обеспечить высокую скорость генерации. На современном технологическом уровне это оказывается уже возможным. Рассмотрим ситуацию более подробно.

Фотоэффект был открыт А.Г. Столетовым в Московском университете ещё до появления понятия квантов, поэтому не был объяснён. Последовательное объяснение фотоэффекта на основе квантовой теории дано А. Эйнштейном [7][4]. Первопричина случайности (пуассоновской статистики) фотоотсчётов при детектировании лазерного излучения носит принципиально квантовый характер, и она обусловлена поглощением фотонов атомами (см. детали в [8, 9]).

В реальной ситуации при создании квантовых генераторов случайных чисел приходится ослаблять лазерное излучение до квазиоднофотонного уровня, для того чтобы фотоотсчёты не были слишком частыми. Это требование, с одной стороны, связано со стремлением получить для измерений истинно квантовый (по возможности, однофотонный) процесс, а с другой — с конечным временем восстановления фотодетектора после регистрации. Восстановление фотодетектора до следующего акта регистрации обеспечивает статистическую независимость последовательных фотоотсчётов.

Рассмотрим сначала идеальный вариант. В качестве подходящего квантового процесса выбираем фотодетектирование ослабленного когерентного состояния. Когерентное состояние представляет собой суперпозицию с разным фоковским числом фотонов:

$$|\alpha\rangle = \exp\left(-\frac{|\alpha|^2}{2}\right)\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}|n\rangle\,, \tag{5}$$

где $|\alpha|^2 = \mu$ — среднее число фотонов в когерентном состоянии.

Формально процесс детектирования сводится к проецированию на подпространство с фоковским числом фотонов $k \geqslant 1$. Реальные лавинные фотодетекторы не различают число фотонов и регистрируют только либо наличие, либо отсутствие фотоотсчёта. Такое измерение, согласно проекционному постулату измерений, даётся разложением единицы:

$$I = |0\rangle\langle 0| + \sum_{k=1}^{\infty}|k\rangle\langle k|\,. \tag{6}$$

[4] Хотя в работе [7] само слово "квант" в явном виде не употреблялось, использовалась дискретность энергии излучения.

Каждому элементарному событию отвечает свой проекционный оператор, тогда как измерение (6) имеет два исхода: отсутствие и наличие фотоотсчёта.

Отсутствие фотоотсчёта — это проектор $\mathcal{P}_0 = |0\rangle\langle 0|$ на подпространство с нулевым фоковским числом фотонов (вакуумную компоненту состояния), наличие фотоотсчёта — проектор $\mathcal{P}_{k\geqslant 1} = \sum_{k=1}^{\infty}|k\rangle\langle k|$ на подпространство с фоковским числом фотонов $k \geqslant 1$.

Вероятность фотоотсчёта $(*)$ выражается как (см., например, [8])

$$P(*) = \mathrm{Tr}\left\{|\alpha\rangle\langle\alpha|\mathcal{P}_{k\geqslant 1}\right\} = 1 - \exp(-\mu)\,, \tag{7}$$

соответственно вероятность отсутствия фотоотсчёта $(\sqcup)$

$$P(\sqcup) = \mathrm{Tr}\left\{|\alpha\rangle\langle\alpha|\mathcal{P}_0\right\} = \exp(-\mu)\,,\quad P(\sqcup) + P(*) = 1\,. \tag{8}$$

Строго говоря, состояние (5) является монохроматическим (формально бесконечно протяжённым). Если измерения проводятся в конечном временно́м окне T, то в (7), (8) надо сделать замену $\mu \to \mu_T$ (см. детали в [8]). Такая замена интуитивно понятна, поскольку μ_T — доля от среднего числа фотонов, которая набирается во временно́м окне[5] T. Конечно, должно быть $\mu_T \ll 1$.

События $*$ и $\sqcup$ — наличие и отсутствие фотоотсчёта — мы положим в основу процесса извлечения случайности.

Как будет видно в дальнейшем, в процедуре извлечения случайности потребуется только статистическая независимость фотоотсчётов во временных окнах T, тогда как сами вероятности $P(*)$ и $P(\sqcup)$ могут быть произвольными. При выполнении требования независимости процедура фотодетектирования приводит к бернуллиевской схеме испытаний, основанной на квантовых явлениях.

Принципиально важно, что вероятность исходов имеет квантовую природу — исходы принципиально непредсказуемы, статистически независимы, истинно случайны.

Любая классическая схема, полученная каким-либо способом из классической физической системы (например, при редукции итогового отклонения шарика для доски Гальтона), не является истинно случайной — исход точно предсказуем, если известны начальные условия и эволюция классической системы.

Главная экспериментальная трудность при реализации высокоскоростного квантового генератора случайных чисел, основанного на детектировании фотоотсчётов, состоит в выполнении противоречивых требований: с одной стороны, необходимо обеспечить квантовый характер сигнала с малым средним числом фотонов в импульсе, а с другой — получить высокую скорость генерации случайных чисел. Реальный лавинный фотодетектор имеет квантовую эффективность $\eta < 1$, в этом случае вероятности (7), (8) сохраняют вид, но с заменой $\mu_T \to \eta\mu_T$.

Поглощение отдельного фотона в твердотельной структуре лавинного фотодетектора приводит к рожде-

[5] Величины μ и μ_T — безразмерные. Физический смысл μ — число фотонов (при малых $\mu \ll 1$), которое можно обнаружить, если целиком доступно для измерения когерентное состояние протяжённостью $L \approx c/(\delta\omega)$ (c — скорость света, $\delta\omega$ — ширина спектра состояния). Для монохроматического состояния $\delta\omega \to 0$ протяжённость $L \to \infty$. Если из всей протяжённости доступна только длина ΔL (соответственно, интервал времени $T = \Delta L/c$), то в этом окне будет зарегистрирована доля $\mu_T = \mu\Delta L/L$.

нию электрон-дырочной пары, которая "усиливается" — рождает лавину носителей заряда, импульс тока от которой регистрируется. После рождения лавины происходит процесс рассасывания лавины, что требует определённого времени. До тех пор пока не произошло рассасывания лавины, фотодетектор не готов к новому акту регистрации, иначе это приведёт к корреляции фотоотсчётов и искажению пуассоновской статистики, т.е. фотоотсчёты, особенно в близких временны́х окнах, перестанут быть независимыми.

Восстановление фотодетектора до следующего акта регистрации является первым из двух условий, при которых обеспечивается статистическая независимость последовательных фотоотсчётов. Второе условие состоит в стабильности интенсивности оптического поля лазера. В этом случае распределение $(P(*), P(\sqcup))$ будет стационарным, а последовательные фотодетектирования (регистрации фотонов) будут строго независимыми [10].

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

Первая задача состоит в контролируемом и доказуемом способе получения первичной "квантовой" случайности — пуассоновского случайного процесса. На физическом уровне к решению этой задачи можно наиболее близко подойти, если использовать для регистрации матрицу фотодетекторов SiPM (Silicon Photo Multiplier) (см. разделы 9, 10).

Вторая задача состоит в эффективном извлечении (экстракции) из пуассоновского случайного процесса случайной равномерно распределённой последовательности из 0 и 1.

Вторая задача разбивается на два этапа. Первый этап — реализация физического устройства, дающего "квантовую" случайность в виде бернуллиевской последовательности из событий $*$ и $\sqcup$ — наличия или отсутствия фотоотсчёта. Второй этап — извлечение из бернуллиевской последовательности случайной и равновероятной последовательности из 0 и 1.

2. Связь количества информации и количества случайности

Интуитивно понятно, что из всякой конечной случайной выборки можно извлечь некоторое количество истинно случайных, равновероятных бит. При измерениях физического процесса важно знать верхнюю границу этой истинной случайности, для того чтобы установить, насколько эффективен конкретный способ.

Зададим дискретную случайную величину A с распределением $P_A(a)$, $a \in \{a_1, \ldots, a_m\}$. Измерения над физической системой представим в виде n-кратной выборки

$$L_n = (a_{i_1}, \ldots, a_{i_n}),\ a_{i_j} \in \{a_1, \ldots, a_m\},\ j = \overline{1,n},$$

из распределения случайной величины A.

Пусть $\nu_1, \ldots, \nu_m$ — частоты появления исходов $\{a_1, \ldots, a_m\}$ в последовательности L_n. При больших n в соответствии с законом больших чисел частоты становятся близкими (нестрого) к вероятностям, т.е. $\nu_1 \approx \approx nP_A(a_1), \ldots, \nu_m \approx nP_A(a_m)$. Тогда вероятность

$$P(L_n) = \prod_{k=1}^{m} P_A^{\nu_k}(a_k) \approx \prod_{k=1}^{m} \big(P_A(a_k)\big)^{nP(a_k)} = 2^{-nH(A)},$$

где $H(A)$ — энтропия Шеннона распределения случайной величины A,

$$H(A) = -\sum_{k=1}^{m} P_A(a_k) \log P_A(a_k),$$

все логарифмы здесь и далее берутся по основанию 2.

На качественном уровне это означает, что практически все возможные последовательности $(a_{i_1}, \ldots, a_{i_n})$, которые могут получиться при измерении физического процесса, равновероятны, а их число равно $2^{n'}$, $n' = = nH(A)$ (будем считать его целым). Это те последовательности, назовём их типичными, в которых число мест, занимаемых исходами $\{a_1, \ldots, a_m\}$, практически равно $\{nP_A(a_1), \ldots, nP_A(a_m)\}$, $\sum_{k=1}^{m} nP_A(a_k) = n$.

Упорядочим (пронумеруем) типичные последовательности как $(a_{i_1}^{(j)}, \ldots, a_{i_n}^{(j)})$, $j = \overline{0, 2^{n'} - 1}$. Поставим в соответствие каждой типичной последовательности $(a_{i_1}^{(j)}, \ldots, a_{i_n}^{(j)})$ двоичную последовательность $(\varepsilon_1^{(j)} \ldots \varepsilon_{n'}^{(j)})$ — двоичное разложение номера j, заполняя тем самым всё множество битовых последовательностей:

$$\left\{ \begin{array}{ccc} a_{i_1}^{(0)}, \ldots, a_{i_N}^{(0)} & \to & 0 \ldots 0 \\ \ldots & \to & \ldots \\ a_{i_1}^{(j)}, \ldots, a_{i_n}^{(j)} & \to & \varepsilon_1^{(j)} \ldots \varepsilon_{n'}^{(j)} \\ \ldots & \to & \ldots \\ a_{i_1}^{(2^{n'}-1)}, \ldots, a_{i_n}^{(2^{n'}-1)} & \to & 1 \ldots 1 \end{array} \right\}. \qquad (9)$$

Тогда при измерении физического процесса после получения $(a_{i_1}^{(j)}, \ldots, a_{i_n}^{(j)})$ и выбора соответствующей $(\varepsilon_1^{(j)} \ldots \varepsilon_{n'}^{(j)})$ получаем *равновероятный* выбор битовых последовательностей длиной n', т.е. извлекаем *истинно случайную* последовательность.

Пусть при реализации некоторого реального алгоритма извлечения случайности мы извлекаем n' случайных равновероятных бит из случайной выборки размером n. Целесообразно принять относительную величину $\lambda = n'/n$ за "количество случайности", измеряемое в битах, приходящихся на одно измерение физического процесса при извлечении (экстракции) двоичной равновероятной последовательности.

Описанный выше алгоритм извлечения случайности является оптимальным, для него "количество случайности" является максимально возможным:

$$\lambda_{\max} = \frac{n'}{n} = H(A),$$

но он не может быть эффективно реализован на практике, поскольку требует огромной памяти для построения и хранения таблицы (9). Величина $\lambda_{\max} = H(A)$ является верхней границей, по которой можно судить об эффективности того или иного реального алгоритма.

Результат о равнораспределённости и соответствующей мощности множества типичных последовательностей, на качественном уровне сформулированный выше, является одним из фундаментальных результатов теории информации, представленным в теоремах Шеннона для дискретного источника сообщений без памяти [11–14]. Энтропия $H(A)$ известна также как количество информации, содержащейся в вероятностной схеме A.

В терминологии теории информации исходы $\{a_1, \ldots, a_m\}$ — это буквы (алфавита), типичные последовательности $(a_{i_1}^{(j)}, \ldots, a_{i_n}^{(j)})$ — случайные сообщения дискретного источника, двоичное разложение $(\varepsilon_1^{(j)} \ldots \varepsilon_{n'}^{(j)})$ — кодированные сообщения. Величина $H(A)$ — энтропия Шеннона — количество информации, измеряемое в битах, приходящихся на одну букву сообщения, характеризующее минимальную длину двоичного представления сообщений для их передачи по каналу связи без ошибок.

Таким образом, энтропия Шеннона на букву для дискретного источника случайных сообщений и есть мера максимального количества истинной случайности.

Для алфавита $\{\sqcup, *\}$ с распределением вероятности $\{P(\sqcup), P(*)\}$ количество случайности (в асимптотическом пределе)

$$h\big(P(*)\big) = h\big(P(\sqcup)\big) < 1\,, \tag{10}$$

где учтено, что $P(\sqcup) + P(*) = 1$, $h(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ — бинарная энтропийная функция Шеннона.

В реальной ситуации длина обрабатываемого блока n всегда конечна. В такой ситуации кроме типичных последовательностей будут появляться нетипичные последовательности, естественно, с меньшей, но не с исчезающе малой вероятностью. Такие последовательности также содержат некоторое истинно случайное число, 0 и 1, которое не хотелось бы терять.

Рассмотренный выше асимптотический случай даёт общий способ экстракции случайных 0 и 1. Основная идея извлечения случайности основывается на том, что типичные последовательности являются равновероятными, поскольку содержат (асимптотически) один и тот же набор частот $\nu_1 = nP_A(a_1), \ldots, \nu_m = nP_A(a_m)$ появления символов $a_1, \ldots, a_m$, а различаются только их перестановкой.

Для конечного n все последовательности можно разбить на несколько классов, различающихся набором частот $\nu_1, \ldots, \nu_m$: типичные последовательности образуют один из этих классов. Внутри каждого класса последовательности различаются перестановкой символов и являются равновероятными. Для каждого класса можно провести нумерацию последовательностей и из равновероятных номеров постараться извлечь истинно случайные биты.

Следующий вопрос, на который следует ответить: как найти эффективный способ нумерации последовательностей внутри каждого класса? Число последовательностей огромно (экспоненциально по n) даже при достаточно небольших длинах блоков n. Прямая нумерация "в лоб" является экспоненциально сложной, поэтому практически не реализуема. Однако, как мы покажем в разделах 9, 10, существует эффективный способ извлечения истинно случайных бит, который имеет полиномиальную сложность по длине блока n при условии независимости исходных символов. Этот способ осуществлён в реальных устройствах [15–18], пример которых приведён в разделе 9.

3. Метод фон Неймана

Прежде чем привести полиномиальный алгоритм экстракции случайных 0 и 1, который гарантирует истинную случайность и извлекает её из всех, а не только типичных последовательностей, удобно привести частный метод извлечения случайных 0 и 1 из бернуллиевской последовательности, предложенный фон Нейманом[6] ещё в 1951 г.

Опишем метод фон Неймана на примере алфавита $\{\sqcup, *\}$ с распределением вероятностей $\{P(\sqcup), P(*)\}$.

Последовательность из событий $*$, $\sqcup$ длиной n разбивается на последовательные пары без зацепления, которые "на ходу" просматриваются. Встретившаяся парная комбинация $(*, \sqcup)$ заменяется нулём, комбинация $(\sqcup, *)$ заменяется единицей. Две другие парные комбинации, $(*, *)$ и $(\sqcup, \sqcup)$, отбрасываются. Полученная последовательность 0 и 1 является *равновероятно* распределённой случайной последовательностью, поскольку вероятность нуля $P(0) = P(*)P(\sqcup)$ равна вероятности единицы $P(1) = P(\sqcup)P(*)$. Таким образом, в оставшейся части имеют место равенства $P(0) = P(1) = 1/2$ независимо от значений $P(*)$ и $P(\sqcup)$.

Важно подчеркнуть, что этот метод применим при любых исходных вероятностях $P(*)$ и $P(\sqcup)$.

Нетрудно увидеть, что в методе фон Неймана даже в самом благоприятном случае, когда $P(*)$, $P(\sqcup)$ близки к $1/2$, количество извлекаемой случайности составляет не более $1/4$ бита на один знак последовательности событий $*, \sqcup$, тогда как максимальное количество случайности, которое в принципе может быть извлечено из последовательности событий $*, \sqcup$, равно $\lambda_{\max} = h(P(*)) = h(\approx 1/2)$, т.е. весьма близко к 1.

Из данного метода, очевидного, простого и изящного, можно извлечь нечто существенно более общее и важное.

Метод удобно представить в виде таблицы:

$$\begin{aligned} (\sqcup, \sqcup) &\;\to\; \text{отбрасывается}\,, \\ (\sqcup, *) &\;\to\; 0\,, \\ (*, \sqcup) &\;\to\; 1\,, \\ (*, *) &\;\to\; \text{отбрасывается}\,. \end{aligned} \tag{11}$$

Из представления (11) метода фон Неймана можно вывести следующие шаги алгоритма извлечения случайности.

1. Первый шаг — выбор длины n обрабатываемого блока исходной последовательности, в данном случае $n = 2$.

2. Второй шаг — разбиение всех блоков длиной n на различные классы так, чтобы все представители (блоки) из одного класса имели одинаковое число $\sqcup$ и $*$, а следовательно, одинаковую вероятность.

3. Третий шаг — классы, состоящие из одного блока, отбрасываются — это класс $\sqcup\sqcup$ и класс $**$.

4. Четвёртый шаг — все равновероятные блоки внутри класса нумеруются и сводятся в таблицу соответствия: $(\sqcup, *) \to 0$, $(*, \sqcup) \to 1$, здесь $0, 1$ — номера блоков.

5. Пятый шаг — полученный в опыте блок длиной $n = 2$ сравнивается с таблицей, определяется номер блока, на выход выдаётся двоичное представление номера блока, в данном случае 0 или 1.

Принципиальный шаг метода фон Неймана, который мы будем использовать в дальнейшем, — это разбиение всех возможных блоков на классы равновероятных блоков, содержащих одинаковое число $*$ и $\sqcup$ и различаю-

[6] Фон Нейман также первый предложил программный генератор псевдослучайных чисел [19].

щихся только перестановкой элементов. Как видно из (11), в методе фон Неймана получаются три класса. В первом и третьем классах имеется по одному элементу. Эти классы отбрасываются. Во втором классе — два равновероятных элемента: $(\sqcup, *)$ и $(*, \sqcup)$. Число элементов во втором классе равно степени двух, а именно 2^1, число извлекаемых истинно случайных бит $\log 2^1 = 1$.

В разделе 2 показано, что для извлечения всей случайности, близкой к асимптотическому пределу, требуется нумеровать блоки большой длины. Попытки решить задачу "в лоб", используя таблицу вида (9), оказываются неосуществимыми, а именно экспоненциально сложными по длине n обрабатываемого блока.

Пусть, например, $n = 64$. Тогда для записи таблицы требуется объём памяти $n \times 2^n = 2^{37}$ Гбайт. Поиск нужного номера по составу наблюдаемого блока и, соответственно, двоичного разложения его номера потребует просмотра всей таблицы и будет состоять в среднем из $2^n \approx 10^{19}$ шагов.

В разделах 5–7 мы рассмотрим эффективный полиномиальный способ нумерации и экстракции истинной случайности "на ходу". Метод требует объёма памяти $n^3 = 2^5$ кбайт.

4. Предельное число равновероятных бит. Точные утверждения

Сформулируем точное утверждение о предельном числе равновероятных бит, которые можно извлечь из неравновероятной последовательности, состоящей из событий $*$ и $\sqcup$, длиной n при $n \to \infty$.

Все последовательности длиной n можно разбить на непересекающиеся классы $\mathcal{R}_n(k)$, $k = \overline{0,n}$, внутри которых последовательности содержат k событий $*$ и соответственно $n-k$ событий $\sqcup$. Число последовательностей в классе

$$\left|\mathcal{R}_n(k)\right| = C_n^k = \frac{n!}{k!(n-k)!}\,, \tag{12}$$

все последовательности из данного класса имеют одинаковую вероятность:

$$P_n(k) = (1-p)^{n-k}p^k, \quad p = P(*), \quad 1-p = P(\sqcup)\,. \tag{13}$$

Обозначим

$$\ell_n(k) = \left[\log\left(C_n^k\right)\right], \tag{14}$$

где $[x]$ — целая часть числа x.

Поставим в соответствие каждой последовательности из класса $\mathcal{R}_n(k)$ двоичную последовательность из множества $\{0,1\}^{\ell_n(k)}$. Если $\log C_n^k$ не целое, то "лишние" последовательности из класса $\mathcal{R}_n(k)$ отбросим, т.е. вычеркнем из реальной выборки.

При условии, что последовательность длиной n выбирается из класса $\mathcal{R}_n(k)$ (с учётом вычеркнутых последовательностей), её выбор является равновероятным, следовательно, становится равновероятным и выбор двоичных последовательностей длиной $\ell_n(k)$.

Среднее число случайных равновероятных бит определим как

$$\mathcal{L}_n = \sum_{k=0}^{n} \ell_n(k) C_n^k (1-p)^{n-k} p^k = \sum_{k=0}^{n} \ell_n(k) C_n^k P_n(k)\,. \tag{15}$$

Утверждение 1. Предел

$$\lim_{n\to\infty} \frac{1}{n}\,\mathcal{L}_n = h(p)\,, \tag{16}$$

где $h(p)$ — бинарная энтропийная функция Шеннона,

$$h(p) = -p\log(p) - (1-p)\log(1-p)\,.$$

Доказательство. Воспользуемся формулой Стирлинга

$$\begin{aligned} n! &= \sqrt{2\pi n}\, n^n \exp(-n)\big(1+o(1)\big)\,,\\ \log(n!) &= n\log n\big(1+o(1)\big)\,, \end{aligned} \tag{17}$$

для оценки биномиальных коэффициентов.

Имеет место также асимптотическое представление для суммы вероятностей распределения Бернулли за пределами отклонения $\ln n\sqrt{n}$ от np-математического ожидания, $n \to \infty$ [20]:

$$\sum_{k\notin(np-\ln n\sqrt{n},\, np+\ln n\sqrt{n})} C_n^k P_n(k) = o(1)\,. \tag{18}$$

Представим также

$$\ell_n(k) = \log C_n^k - \varepsilon_n(k)\,, \tag{19}$$

где $0 \leqslant \varepsilon_n(k) \leqslant 1$. Кроме того, поскольку

$$\sum_{k=0}^{n} C_n^k = 2^n\,, \tag{20}$$

для любого $k = \overline{0,n}$ имеет место неравенство

$$\log C_n^k \leqslant n\,. \tag{21}$$

Тогда нетрудно проследить следующую цепочку соотношений:

$$\begin{aligned} \frac{1}{n}\mathcal{L}_n &= \frac{1}{n}\sum_{k=0}^{n} \ell_n(k) C_n^k P_n(k) = \frac{1}{n}\sum_{k=0}^{n} \big(\log C_n^k\big) C_n^k P_n(k) + \\ &+ O\left(\frac{1}{n}\right) = \frac{1}{n} \sum_{k\in(np-\ln n\sqrt{n},\, np+\ln n\sqrt{n})} \big(\log C_n^k\big) C_n^k P_n(k) + \\ &+ o(1) = \frac{1}{n}\big(n\log n - \big(np\log np + n(1-p)\log n(1-p)\big)\big) \times \\ &\times \sum_{k\in(np-\ln n\sqrt{n},\, np+\ln n\sqrt{n})} C_n^k P_n(k)\big(1+o(1)\big) = h(p)\big(1+o(1)\big)\,. \end{aligned} \tag{22}$$

Следовательно, предел (22) равен

$$\lim_{n\to\infty} \frac{1}{n}\,\mathcal{L}_n = h(p)\,. \tag{23}$$

Доказанное *утверждение* 1 тесно связано с теоремами Шеннона [11–13] для дискретного источника сообщений без памяти, которые являются фундаментальными результатами классической теории информации.

Из теорем Шеннона следует, что для источника случайных сообщений без памяти, т.е. при независимом выборе букв (у нас это $*$ и $\sqcup$), практически все сообщения имеют (асимптотически) одну и ту же вероятность $2^{-nh(p)}$ (1-я теорема), а их число равно $2^{nh(p)}$ (2-я теорема). Это так называемые типичные последовательности, составляющие множество последовательностей, вероятность которого очень близка к 1. Очевидно, что длина двоич-

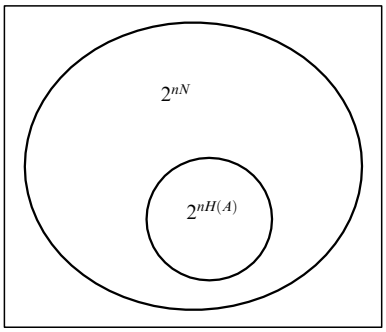


Рис. 3. Иллюстрация полного множества и множества типичных последовательностей, генерируемых дискретным источником без памяти, — каждая точка в множестве — это последовательность. Оба множества имеют экспоненциально большой размер по длине последовательностей, тем не менее вероятность попасть в множество типичных последовательностей при большой серии использования источника стремится к единице, соответственно, вероятность того, что последовательность попадёт в экспоненциально большое множество нетипичных последовательностей, стремится к нулю.

ного кодирования или двоичная нумерация всех типичных последовательностей требует $\log 2^{nh(p)} = nh(p)$ двоичных разрядов, или $h(p)$ двоичных разрядов на одну букву сообщения. Это соответствует предельному среднему значению в (23) (см. качественную иллюстрацию множества всех последовательностей и множества типичных последовательностей на рис. 3). Очевидно, что при равновероятном выборе типичных последовательностей получаем равновероятный выбор двоичных векторов, отвечающих двоичному представлению номеров типичных последовательностей.

Энтропия источника $h(p) < 1$ — предельное (максимальное) количество истинно случайных бит, приходящихся на один символ первичной бернуллиевской последовательности событий $*$ и $\sqcup$.

5. Практическая нумерация

Утверждения, сделанные в разделе 4, справедливы в асимптотическом пределе $n \to \infty$. Для конструирования квантовых генераторов случайных чисел асимптотических результатов оказывается недостаточно по двум причинам.

Первая причина. Теорема кодирования Шеннона для источника является, по сути, теоремой существования, поскольку не даёт конструктивного, алгоритмически эффективного — полиномиального — способа нумерации.

Вторая причина, точнее вопрос: а что делать с последовательностями, которые не являются типичными? Напомним, что типичные последовательности (асимптотически) — это последовательности, у которых числа событий $*$ и $\sqcup$ очень близки к средним значениям $nP(*)$ и $nP(\sqcup)$. Число событий $*$ и $\sqcup$ при большой, но конечной длине n "дышит" — испытывает заметные флуктуации относительно их математических ожиданий. Такие последовательности, конечно, также содержат некое количество истинно случайных бит, которые не хотелось бы терять. Нетипичные последовательности появляются с гораздо меньшей суммарной вероятностью, чем типичные, но, тем не менее, не с нулевой.

По этой причине для извлечения всей случайности, которая содержится во всех бернуллиевских последовательностях (блоках) конечной длины, хотелось бы получить доказуемый метод экстракции истинно случайных 0 и 1 из всех последовательностей — с любым числом $*$ и $\sqcup$, а не только из типичных.

Для решения поставленной задачи мы будем использовать метод двоичного кодирования, открытый В.Ф. Бабкиным [21] в 1971 г., — метод нумерации бернуллиевских последовательностей с полиномиальными ресурсами по времени и памяти. Метод возник в теории арифметического кодирования (другое название — кодирование без потерь), и, на наш взгляд, он должен был давно привлечь внимание в разработке генераторов случайных чисел. Данный метод, несомненно, является одной из жемчужин теории кодирования.

Перейдём к описанию метода. Рассмотрим блок длиной n, в котором k событий $*$ произошли на местах $i_1, i_2, \ldots, i_k$, $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$.

Присвоим блоку номер

$$\mathrm{Num}\,(i_1, i_2, \ldots, i_k) = C^1_{i_1-1} + C^2_{i_2-1} + \ldots + C^{k-1}_{i_{k-1}-1} + C^k_{i_k-1}\,, \tag{24}$$

где, как обычно, полагаем $C^i_j = 0$, если $j < i$. Данное равенство и даёт способ нумерации методом В.Ф. Бабкина.

Ввиду исключительной важности для решения задач по созданию высокоскоростного ГСЧ и с целью пропаганды научного наследия В.Ф. Бабкина приведём здесь два *утверждения*, устанавливающих взаимно однозначное соответствие состава $(i_1, i_2, \ldots, i_k)$ обрабатываемого блока его номеру $\mathrm{Num}\,(i_1, i_2, \ldots, i_k)$, подсчитанному в виде суммы биномиальных коэффициентов (24) [21].

Утверждение 2. Имеют место соотношения

$$\begin{aligned} &\min_{i_1, i_2, \ldots, i_k} \mathrm{Num}\,(i_1, i_2, \ldots, i_k) = 0\,, \\ &\max_{i_1, i_2, \ldots, i_k} \mathrm{Num}\,(i_1, i_2, \ldots, i_k) = C^k_n - 1\,. \end{aligned} \tag{25}$$

Доказательство. Действительно, несложно убедиться, что

$$\min_{i_1, i_2, \ldots, i_k} \mathrm{Num}\,(i_1, i_2, \ldots, i_k) = \mathrm{Num}(1, 2, \ldots, k) = 0\,. \tag{26}$$

Далее,

$$\begin{aligned} &\max_{i_1, i_2, \ldots, i_k} \mathrm{Num}\,(i_1, i_2, \ldots, i_k) = \\ &\qquad = \max_{i_1, i_2, \ldots, i_k} \left(C^1_{i_1-1} + C^2_{i_2-1} + \ldots + C^k_{i_k-1}\right) = \\ &\qquad = \mathrm{Num}\,(n-k, n-k+1, \ldots, n) = \\ &\qquad = C^1_{n-k-1} + C^2_{n-k} + \ldots + C^k_{n-1} = C^k_n - 1. \end{aligned} \tag{27}$$

Последнее равенство нетрудно получить из последовательного применения известного соотношения

$$C^k_n = C^{k-1}_{n-1} + C^k_{n-1}\,. \tag{28}$$

Утверждение 3. Соотношение между блоками с k событиями $*$ на местах $i_1, i_2, \ldots, i_k$ и номерами $\mathrm{Num}\,(i_1, i_2, \ldots, i_k)$ взаимно однозначное.

Доказательство. Рассмотрим два блока с номерами

$$\begin{aligned} &\mathrm{Num}\,(i^{(1)}_1, \ldots, i^{(1)}_s, i_{s+1}, \ldots, i_k)\,, \\ &\mathrm{Num}\,(i^{(2)}_1, \ldots, i^{(2)}_s, i_{s+1}, \ldots, i_k)\,, \end{aligned} \tag{29}$$

и пусть $i^{(1)}_s > i^{(2)}_s$ — это первый номер (начиная справа), при котором блоки "разошлись" по позициям события $*$.

С использованием *утверждения* 2 для доказательства взаимно однозначного соответствия достаточно показать, что номера двух разных блоков в (29) не совпадают ни при каких значениях позиций события $*$. Это будет так, если разница номеров в (29) положительна. Имеем

$$\begin{aligned}&\mathrm{Num}\,(i_1^{(1)},\ldots,i_s^{(1)},i_{s+1},\ldots,i_k)-\\&-\mathrm{Num}\,(i_1^{(2)},\ldots,i_s^{(2)},i_{s+1},\ldots,i_k)\geqslant\\&\geqslant\min_{i_1^{(1)},\ldots,i_{s-1}^{(1)}}\mathrm{Num}\,(i_1^{(1)},\ldots,i_{s-1}^{(1)},i_s^{(1)},i_{s+1},\ldots,i_k)-\\&-\max_{i_1^{(2)},\ldots,i_{s-1}^{(2)}}\mathrm{Num}\,(i_1^{(2)},\ldots,i_{s-1}^{(2)},i_s^{(2)},i_{s+1},\ldots,i_k)\,.\end{aligned}\tag{30}$$

Найдём

$$\begin{aligned}&\min_{i_1^{(1)},\ldots,i_{s-1}^{(1)}}\mathrm{Num}\,\bigl(i_1^{(1)},\ldots,i_{s-1}^{(1)},i_s^{(1)},i_{s+1},\ldots,i_k\bigr)=\\&=\mathrm{Num}\,\bigl(1,\ldots,s-1,i_s^{(1)},i_{s+1},\ldots,i_k\bigr)=\\&=0+C^s_{i_s^{(1)}-1}+\bigl(C^{s+1}_{i_{s+1}-1}+\ldots+C^{k-1}_{i_{k-1}-1}+C^k_{i_k-1}\bigr)\,.\end{aligned}\tag{31}$$

Воспользуемся формулой (27) при $n=i_s^{(2)}$, $k=s$ и учтём, что $i_{s-1}^{(2)}<i_s^{(2)}$. Тогда получим

$$\begin{aligned}&\max_{i_1^{(2)},\ldots,i_{s-1}^{(2)}}\mathrm{Num}\,\bigl(i_1^{(2)},\ldots,i_{s-1}^{(2)},i_s^{(2)},i_{s+1},\ldots,i_k\bigr)=\\&=\max_{i_1^{(2)},\ldots,i_{s-1}^{(2)}}\left\{C^1_{i_1^{(2)}-1}+C^2_{i_2^{(2)}-1}+\ldots+C^{s-1}_{i_{s-1}^{(2)}-1}\right\}+\\&+\bigl(C^{s+1}_{i_{s+1}-1}+\ldots+C^{k-1}_{i_{k-1}-1}+C^k_{i_k-1}\bigr)=\\&=C^1_{i_s^{(2)}-s-1}+C^2_{i_s^{(2)}-s}+\ldots+C^{s-1}_{i_s^{(2)}-2}+C^s_{i_s^{(2)}-1}+\\&+\bigl(C^{s+1}_{i_{s+1}-1}+\ldots+C^{k-1}_{i_{k-1}-1}+C^k_{i_k-1}\bigr)=\\&=\bigl(C^s_{i_s^{(2)}}-1\bigr)+\bigl(C^{s+1}_{i_{s+1}-1}+\ldots+C^{k-1}_{i_{k-1}-1}+C^k_{i_k-1}\bigr)\,.\end{aligned}\tag{32}$$

Следовательно, с учётом равенств (31), (32) и соотношения $i_s^{(1)}>i_s^{(2)}$ имеем

$$\begin{aligned}&\mathrm{Num}\,\bigl(i_1^{(1)},\ldots,i_s^{(1)},i_{s+1},\ldots,i_k\bigr)-\\&-\mathrm{Num}\,\bigl(i_1^{(2)},\ldots,i_s^{(2)},i_{s+1},\ldots,i_k\bigr)\geqslant\\&\geqslant\min_{i_1^{(1)},\ldots,i_{s-1}^{(1)}}\mathrm{Num}\,\bigl(i_1^{(1)},\ldots,i_{s-1}^{(1)},i_s^{(1)},i_{s+1},\ldots,i_k\bigr)-\\&-\max_{i_1^{(2)},\ldots,i_{s-1}^{(2)}}\mathrm{Num}\bigl(i_1^{(2)},\ldots,i_{s-1}^{(2)},i_s^{(2)},i_{s+1},\ldots,i_k\bigr)=\\&=C^s_{i_s^{(1)}-1}-C^s_{i_s^{(2)}}+1\geqslant 1\,.\end{aligned}\tag{33}$$

Этот факт устанавливает взаимно однозначное соответствие между блоками из множества $\mathcal{R}_n(k)$ и их номерами [21].

Дадим ещё простое эвристическое доказательство *утверждения* 2, основанное на рекурсии.

Пусть номер последовательности с k событиями $*$ в позициях $i_1,i_2,\ldots,i_k$ есть $\mathrm{Num}\,(i_1,i_2,\ldots,i_k)$. Далее, пусть номер последовательности с $k-1$ событиями $*$ в позициях $i_1,i_2,\ldots,i_{k-1}$ есть $\mathrm{Num}\,(i_1,i_2,\ldots,i_{k-1})$.

Номера последовательностей $(i_1,i_2,\ldots,i_k)$ и $(i_1,i_2,\ldots,i_{k-1})$ различаются на некоторое число последовательностей. Подсчитаем это число последовательностей: оно равно числу способов размещения последовательностей с k $*$, у которых k-й фотоотсчёт $*$ находится в позиции i_k-1 — предыдущей позиции по сравнению с последовательностью, у которой событие $*$ в позиции i_k. Число таких последовательностей равно числу способов размещения k фотоотсчётов по i_k-1 ящикам, т.е. $C^k_{i_k-1}$.

Таким образом, получаем "нисходящую" со старших номеров рекуррентную формулу

$$\begin{aligned}&\mathrm{Num}\,(i_1,i_2,\ldots,i_k)=\mathrm{Num}\,(i_1,i_2,\ldots,i_{k-1})+C^k_{i_k-1}=\\&=\mathrm{Num}\,(i_1,i_2,\ldots,i_{k-2})+C^{k-1}_{i_{k-1}-1}+C^k_{i_k-1}=\ldots=\\&=C^1_{i_1-1}+C^2_{i_2-1}+\ldots+C^{k-1}_{i_{k-1}-1}+C^k_{i_k-1}\,.\end{aligned}\tag{34}$$

6. Сложность нумерации. Треугольник Паскаля

Нумерация блоков выполняется последовательно, "на ходу", по мере поступления событий $*$ и $\sqcup$. Задаётся размер блока n, вычисляется один раз таблица биномиальных коэффициентов (табл. 1) размером $(n-1)\times n$. Значение k заранее не фиксируется, случаи $k=0$ и $k=n$ исключаются из рассмотрения как маловероятные. Для хранения каждого биномиального коэффициента требуется не более n двоичных разрядов, что непосредственно следует из соотношений $\log C^k_n\leqslant\log 2^n=n$. Таким образом, полный объём памяти для хранения табл. 1, которая представляет собой "треугольник Паскаля", точнее половину "треугольника", требует не более n^3 двоичных разрядов.

Таблица 1. "Треугольник Паскаля"

	1	2	3	4	5	…	$n-1$	n
i_1	0	1	C^1_2	C^1_3	C^1_4	…	C^1_{n-2}	C^1_{n-1}
i_2	0	0	1	C^2_3	C^2_4	…	C^2_{n-2}	C^2_{n-1}
i_3	0	0	0	1	C^3_4	…	C^3_{n-2}	C^3_{n-1}
…	…	…	…	…	…	…	…	…
i_{n-1}	0	0	0	0	0	…	0	1

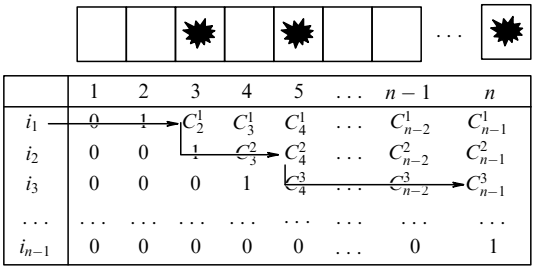
Нумерация последовательностей сводится к движению по некоторой траектории на треугольнике Паскаля с последовательным суммированием биномиальных коэффициентов (см. пример на рис. 4).

Если в первый раз событие $*$ встретилось на месте m_1, то берётся значение биномиального коэффициента на пересечении строки с номером i_1 (первое событие $*$) со столбцом с номером m_1.

Если второй раз событие $*$ встретилось на месте m_2 ($m_2>m_1$), то берётся значение биномиального коэффициента на пересечении строки с номером i_2 со столбцом с номером m_2 и прибавляется к предыдущему значению биномиального коэффициента.

Если в s-й раз событие встретилось на месте m_s ($m_s>m_{s-1}$), то берётся значение биномиального коэффициента на пересечении строки с номером i_s со столбцом с номером m_s и прибавляется к предыдущей сумме биномиальных коэффициентов.

Процесс останавливается, когда просмотрен весь блок размером n. В соответствии с разделом 5 получается номер блока с событиями $*$ и $\sqcup$ в виде двоичного представления — это ещё не случайные биты.



	1	2	3	4	5	...	$n-1$	n
i_1	0	1	C_2^1	C_3^1	C_4^1	...	C_{n-2}^1	C_{n-1}^1
i_2	0	0	1	C_3^2	C_4^2	...	C_{n-2}^2	C_{n-1}^2
i_3	0	0	0	1	C_4^3	...	C_{n-2}^3	C_{n-1}^3
...	...	...	...	...	...	...	...	...
i_{n-1}	0	0	0	0	0	...	0	1

Рис. 4. Пример "треугольника Паскаля", иллюстрирующий вычисление номера последовательности фотоотсчётов "на ходу", по мере появления фотоотсчётов. Пример последовательности с тремя фотоотсчётами $*$. Нумерация сводится к движению по траектории на "треугольнике Паскаля" и последовательному суммированию биномиальных коэффициентов в таблице. В примере номер позиции первого фотоотсчёта $i_1 = 3$, второго $i_2 = 5$, третьего $i_3 = n$.

После того как номер конкретной последовательности из $\sqcup$ и $*$ получен, из его двоичного представления извлекается блок истинно случайных 0 и 1.

7. Извлечение случайности

Мощность множества блоков с k событиями $*$ и $n-k$ событиями $\sqcup$ есть $|\mathcal{R}_n(k)| = C_n^k$. Согласно (24) нумерация блоков $\mathcal{R}_n(k) = C_n^k$ происходит начиная с 0 до $C_n^k - 1$.

Пусть n — чётное, что удобно при компьютерной реализации. Описанный ниже метод работает при любых n. Рассмотрим представление $|\mathcal{R}_n(k)|$ в виде суммы:

$$|\mathcal{R}_n(k)| = 2^{r_m} + \ldots + 2^{r_1} + 2^{r_0}\,, \quad r_m > r_{m-1} > \ldots r_1 > r_0\,. \tag{35}$$

Пусть теперь реализовался блок, имеющий состав $(i_1, i_2, \ldots, i_k)$ событий $*$. Номер блока имеет двоичное разложение вида

$$\mathrm{Num}\,(i_1, i_2, \ldots, i_k) = \varepsilon_{r_m+1} 2^{r_m+1} + \varepsilon_{r_m} 2^{r_m} + \\ + \varepsilon_{r_m-1} 2^{r_m-1} + \ldots + \varepsilon_1 2^1 + \varepsilon_0 2^0\,, \quad \varepsilon_r \in \{0,1\}\,, \tag{36}$$

и соответствующее двоичное представление $(\varepsilon_{r_m+1}, \varepsilon_{r_m}, \varepsilon_{r_m-1}, \ldots, \varepsilon_1, \varepsilon_0)$.

Извлечение блока $\{\varepsilon\}$ случайных 0 и 1 производится из двоичного представления $(\varepsilon_{r_m+1}, \varepsilon_{r_m}, \varepsilon_{r_m-1}, \ldots, \varepsilon_1, \varepsilon_0)$, причём по-разному, в зависимости от того, в каком диапазоне чисел между 0 и $C_n^k - 1$ лежит номер $\mathrm{Num}\,(i_1, i_2, \ldots, i_k)$ текущего блока. А именно:

Номер	Блок $\{\varepsilon\}$ случайных 0 и 1
$0 \leqslant \mathrm{Num}\,(i_1, i_2, \ldots, i_k) \leqslant 2^{r_0} - 1\,,$	$\varepsilon_{r_0-1}, \ldots, \varepsilon_0\,,$
$2^{r_0} \leqslant \mathrm{Num}\,(i_1, i_2, \ldots, i_k) \leqslant 2^{r_0} + 2^{r_1} - 1\,,$	$\varepsilon_{r_1-1}, \ldots, \varepsilon_0\,,$
$2^{r_0} + 2^{r_1} \leqslant \mathrm{Num}\,(i_1, i_2, \ldots, i_k) \leqslant 2^{r_0} + 2^{r_1} + 2^{r_2} - 1\,,$	$\varepsilon_{r_2-1}, \ldots, \varepsilon_0\,,$
...	...
$2^{r_0} + \ldots + 2^{r_m} \leqslant \mathrm{Num}\,(i_1, i_2, \ldots, i_k) \leqslant 2^{r_0} + \ldots + 2^{r_m} - 1\,,$	$\varepsilon_{r_m-1}, \ldots, \varepsilon_0\,.$

(37)

Пронумеруем строки (неравенства) как $0, \ldots, j, \ldots, m$. Тогда в j-й строке — подклассе — содержится 2^{r_j} различных равновероятных номеров $\mathrm{Num}\,(i_1, i_2, \ldots, i_k)$, которым однозначно соответствуют двоичные векторы из пространства $\{0,1\}^{r_j}$. Тогда по каждому текущему номеру $\mathrm{Num}\,(i_1, i_2, \ldots, i_k)$ на выход выдаётся соответствующий ему блок $\{\varepsilon\}$ случайных 0 и 1 (табл. 2).

Таблица 2. Пример экстракции истинно случайных блоков

Позиции $*$ и $\sqcup$ (i_1, i_2)	Номер $N(i_1, i_2)$	Двоичное представление	Случайный блок $\{\varepsilon\} = \varepsilon_{r_{j-1}}, \ldots, \varepsilon_0$
$**\sqcup\sqcup\sqcup\sqcup\sqcup$	0	00000	00
	1	00001	01
$j = 0$	2	00010	10
	$3 = 2^{r_0} - 1$	00011	11
	4	00100	100
	5	00101	101
	6	00110	110
	7	00111	111
$j = 1$	8	01000	000
	9	01001	001
	10	01010	010
	$11 = 2^{r_1} + 2^{r_0} - 1$	01011	011
	12	01100	1100
	13	01101	1101
	14	01110	1110
	15	01111	1111
	16	10000	0000
	17	10001	0001
$j = 2$	18	10010	0010
	19	10011	0011
	20	10100	0100
	21	10101	0101
	22	10110	0110
	23	10111	0111
	24	11000	1000
	25	11001	1001
	26	11010	1010
$\sqcup\sqcup\sqcup\sqcup\sqcup\sqcup** $	$27 = 2^{r_2} + 2^{r_1} + 2^{r_0} - 1$	11011	1011

Рассмотрим пример, иллюстрирующий общий метод, для $n = 8$, $k = 2$. В этом случае

$$\left|\mathcal{R}_n(k)\right| = \frac{8!}{2!6!} = 28 = 2^4 + 2^3 + 2^2,$$
$$m = 2, \quad r_m = 4, \quad r_1 = 3, \quad r_0 = 2\,. \tag{38}$$

8. Истинная случайность

Ниже мы покажем, что при реализации метода нумерации В.Ф. Бабкина и извлечении из него как его продукта блоков $\{\varepsilon\}$ случайных 0 и 1 любая выходная двоичная последовательность любой длины L будет равновероятной, т.е. истинно случайной. Ещё раз отметим, что это возможно только при исходных предположениях о том, что последовательность событий из $*$ и $\sqcup$ (на физическом уровне — фотоотсчётов) является бернуллиевской, т.е. независимой.

Рассмотрим случайную независимую последовательность событий из $*$ и $\sqcup$, которую разобьём на последовательность блоков длиной n. Тогда мы получим независимую последовательность пар чисел (k_i, j_i), $i = 1, 2, \ldots$, где k_i — число событий $*$ в i-м блоке, j_i — номер подкласса, в который попадает номер блока, с совместным распределением $P(k, j)$, точный вид которого, как мы увидим ниже, неважен. *Важно только то, что последовательность пар* (k_s, j_s), $s = 1, 2, \ldots$ *статистически независима.*

Зафиксируем k — число событий $*$ в блоке длиной n. Тогда номер блока $\mathrm{Num}\,(i_1, i_2, \ldots, i_k)$ случайно попадает в один из подклассов (см. (37) и табл. 2), которые

нумеруются (см. раздел 7) числами $0,\ldots,j,\ldots,m$. Обозначим этот подкласс $\mathcal{R}_n(k,j)$, его размер (мощность) по построению (см. табл. 2) равен 2^{r_j} (r_j, вообще говоря, зависит от k). Подмножества номеров $\mathcal{R}_n(k,j)$ не пересекаются и являются разбиением всего множества $\mathcal{R}_n(k)$ номеров $\mathrm{Num}\,(i_1,i_2,\ldots,i_k)$: $\mathcal{R}_n(k)=\bigcup_{j=0}^{m}\mathcal{R}_n(k,j)$.

Итак, вся случайность у нас задана на исходной бернуллиевской последовательности событий из $*$ и $\sqcup$.

Пусть реализовалось событие — блок размером n с фиксированной парой (k,j). Зададимся вопросом: какова вероятность того, что будет выбран конкретный номер $\mathrm{Num}\,(i_1,i_2,\ldots,i_k)$ из j-го подкласса номеров $\mathcal{R}_n(k,j)$, при условии, что он туда попадёт?

Интересующая нас вероятность — это условная вероятность, которая имеет вид

$$P\left(\mathrm{Num}\,(i_1,i_2,\ldots,i_k)|\mathrm{Num}\,(i_1,i_2,\ldots,i_k)\in\mathcal{R}_n(k,j)\right)=$$

$$=\frac{P\left(\mathrm{Num}\,(i_1,i_2,\ldots,i_k)\right)}{\sum_{\mathrm{Num}^*(i_1,i_2,\ldots,i_k)\in\mathcal{R}_n(k,j)}P\left(\mathrm{Num}^*(i_1,i_2,\ldots,i_k)\right)}=$$

$$=\frac{P^k(*)P^{n-k}(\sqcup)}{2^{r_j}P^k(*)P^{n-k}(\sqcup)}=2^{-r_j}\,. \tag{39}$$

Поскольку, по построению, каждому номеру $\mathrm{Num}\,(i_1,i_2,\ldots,i_k)$ из j-го подкласса отвечает соответствующий двоичный блок $(\varepsilon_{r_j-1},\ldots,\varepsilon_0)$, то из (39) следует, что при фиксированной паре (k,j) реализуется равновероятная схема выбора двоичных векторов $(\varepsilon_{r_j-1},\ldots,\varepsilon_0)$ из пространства $\{0,1\}^{r_j}$. Нетрудно также увидеть, что при фиксированной паре (k,j) биты любой части отрезка $(\varepsilon_{r_j-1},\ldots,\varepsilon_0)$ будут появляться также случайно и равновероятно.

Теперь мы покажем, что для любой двоичной последовательности $(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_L)$, извлекаемой из исходной последовательности событий из $*$ и $\sqcup$, её вероятность будет равна

$$P\,(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_L)=\frac{1}{2^L}\,.$$

По построению (см. раздел 7), при фиксированном размере блока n конкретная двоичная последовательность $(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_L)$ может получиться из любой последовательности событий из $*$ и $\sqcup$, длина которых ограничена величиной nM для некоторого максимального M, n — размер блока. Каждая такая последовательность порождает последовательность пар $(k_1,j_1),(k_2,j_2),\ldots,(k_M,j_M)$ — реализаций возможных значений k_s — числа событий $*$ в блоках, и j_s — номеров подклассов, $s=\overline{1,M}$; из пар (k_s,j_s) получаются отдельные двоичные отрезки последовательности $(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_L)$.

Отсюда нетрудно увидеть, что конкретная двоичная последовательность $(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_L)$ может получиться из любой последовательности пар $(k_1,j_1),(k_2,j_2),\ldots,(k_M,j_M)$.

Пусть пара (k_1,j_1) порождает битовый отрезок $\boldsymbol{\varepsilon}_1=(\varepsilon_1,\ldots,\varepsilon_{m_1})$ длиной m_1, пара (k_2,j_2) порождает битовый отрезок $\boldsymbol{\varepsilon}_2=(\varepsilon_{m_1+1},\ldots,\varepsilon_{m_1+m_2})$ длиной m_2, и т.д., пара (k_{M^*},j_{M^*}) порождает битовый отрезок $\boldsymbol{\varepsilon}_{M^*}=(\varepsilon_{m_1+1+\ldots+m_{M^*-1}+1},\ldots,\varepsilon_{m_1+\ldots+m_{M^*}})$ длиной m_{M^*} так, что

$$m_1+m_2+\ldots+m_{M^*}=L,\ M^*\leqslant M\,. \tag{40}$$

Обозначим через KJ множество всех последовательностей $(k_1,j_1),(k_2,j_2),\ldots,(k_M,j_M)$, $P\,(KJ)=1$. Тогда имеем представление

$$P\,(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_L)=\sum_{\left((k_1,j_1),(k_2,j_2),\ldots,(k_M,j_M)\right)\in KJ}\times$$

$$\times\,P\left(\boldsymbol{\varepsilon}_1,\boldsymbol{\varepsilon}_2,\ldots,\boldsymbol{\varepsilon}_{M^*},(k_1,j_1),(k_2,j_2),\ldots,(k_M,j_M)\right). \tag{41}$$

Вследствие независимости исходной последовательности случайные пары

$$(k_1,j_1),(k_2,j_2),\ldots,(k_{M^*},j_{M^*}) \tag{42}$$

будут независимы, а значит, и независимы "тройки"

$$\left(\boldsymbol{\varepsilon}_1,(k_1,j_1)\right),\left(\boldsymbol{\varepsilon}_2,(k_2,j_2)\right),\ldots,\left(\boldsymbol{\varepsilon}_{M^*},(k_{M^*},j_{M^*})\right). \tag{43}$$

Полученная в (39) условная равновероятность выбора двоичных векторов означает, что

$$P(\boldsymbol{\varepsilon}_s|k_s,j_s)=2^{-m_s},\ s=\overline{1,M^*}\,.$$

Отсюда с использованием независимости "троек" (43) получаем:

$$P\,(\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_L)=$$

$$=\sum_{\left((k_1,j_1),(k_2,j_2),\ldots,(k_M,j_M)\right)\in KJ}\prod_{s=1}^{M^*}P\left(\boldsymbol{\varepsilon}_s,(k_s,j_s)\right)\prod_{s=M^*+1}^{M}P\,(k_s,j_s)=$$

$$=\sum_{\left((k_1,j_1),(k_2,j_2),\ldots,(k_M,j_M)\right)\in KJ}\prod_{s=1}^{M^*}P\,(\boldsymbol{\varepsilon}_s|k_s,j_s)P\,(k_s,j_s)\prod_{s=M^*+1}^{M}P\,(k_s\,,j_s)=$$

$$=\sum_{\left((k_1,j_1),(k_2,j_2),\ldots,(k_M,j_M)\right)\in KJ}\prod_{s=1}^{M^*}2^{-m_s}P\,(k_s,j_s)\prod_{s=M^*+1}^{M}P\,(k_s,j_s)=$$

$$=2^{-L}\sum_{\left((k_1,j_1),(k_2,j_2),\ldots,(k_M,j_M)\right)\in KJ}\prod_{s=1}^{M}P\,(k_s,j_s)=2^{-L}\,. \tag{44}$$

Таким образом, мы показали, что любая выходная двоичная последовательность любой длины, извлекаемая из последовательных блоков, является равновероятной, т.е. истинно случайной. Равновероятность двоичной последовательности является следствием бернуллиевости (независимости) исходной физической последовательности из событий $*$ и $\sqcup$, что обеспечивается на физическом уровне реализации (см. разделы 9–12).

Для оценки эффективности экстракции 0 и 1 в зависимости от величин $P\,(*)$, $P\,(\sqcup)$ и длины блока n было проведено компьютерное моделирование, результаты которого представлены на рис. 5.

9. Физическая реализация квантового генератора случайных чисел

В этом разделе для иллюстрации идей, изложенных в разделах 2–8, мы приведём пример квантового генератора случайных чисел.

При создании высокоскоростного квантового генератора случайных чисел с контролируемой квантовым характером последовательностью фотоотсчётов приходится находить компромисс между взаимно противоречивыми требованиями.

Для того, чтобы источник первичной случайности имел квантовый характер, когерентное состояние долж-

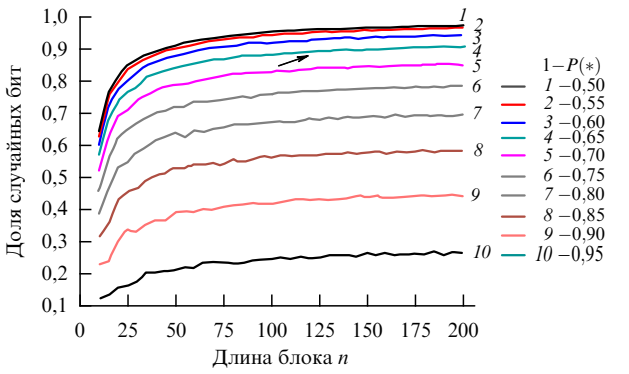


Рис. 5. (В цвете онлайн.) Эффективность экстракции случайных бит из бернуллиевской последовательности ∗ и ⊔ при различных значениях вероятностей $P(*)$ и $P(\sqcup)$ как функция длины обрабатываемого блока n. Генерация последовательностей фотоотсчётов ∗ и ⊔ проводилась с помощью математического генератора псевдослучайных чисел. Данные зависимости требуются для предварительной оценки длины блока n и оценки выхода на асимптотический режим $n \to \infty$ при заданных экспериментальных значениях $P(*)$ и $P(\sqcup)$.

но быть квазиоднофотонным, т.е. среднее число фотонов должно быть $\mu_T \ll 1$. Малое среднее число фотонов приводит к малой вероятности детектирования во временно́м окне T. Вероятность детектирования будет пропорциональна $\eta\mu_T$, $\eta < 1$ — квантовая эффективность фотодетектора.

Рассмотрим, опуская технические подробности, процесс фотодетектирования. Генерация случайных чисел при фотодетектировании фактически представляет собой достаточно тонкий физический эксперимент в том смысле, что, как и в любом физическом эксперименте при проверке какого-либо теоретического предположения, требуется исключить влияние факторов, которые вносят нежелательные искажения. Применительно к нашей ситуации основная проблема состоит в реализации квантовых измерений — фотодетектирования квазиоднофотонных состояний излучения — таким образом, чтобы вероятность действительно сводилась к проецированию (см. (3)–(8)). Единственными приемлемыми устройствами для этой цели являются лавинные фотодетекторы.

Факт регистрации фотона (рис. 6) выглядит как импульс тока (или напряжения) — "клик" от лавины носителей на выходе фотодетектора, порождённых поглощением фотона.

Пусть на входе фотодетектора имеется однофотонное состояние, поглощение фотона происходит каким-то одним конкретным атомом внутри полупроводниковой структуры детектора, что приводит к появлению электрон-дырочной пары. Однако зарегистрировать импульс тока от отдельной электрон-дырочной пары из-за его малой величины практически невозможно. Поэтому исходная электрон-дырочная пара ускоряется в полупроводниковой структуре и порождает лавину неравновесных носителей, или всплеск тока, которая уже регистрируется.

Это обстоятельство приводит к появлению мёртвого времени — времени рассасывания лавины. До тех пор пока рассасывания не произошло, детектор не готов к регистрации следующего фотона. По этой причине периодичность (тактовая частота) опроса детектора не может быть меньше мёртвого времени. Кроме того, в твердотельных лавинных детекторах имеют место так называемые послеимпульсные (afterpulsing) эффекты — ложные срабатывания после регистрации реального фотона. Неравновесные носители могут "залипать" на

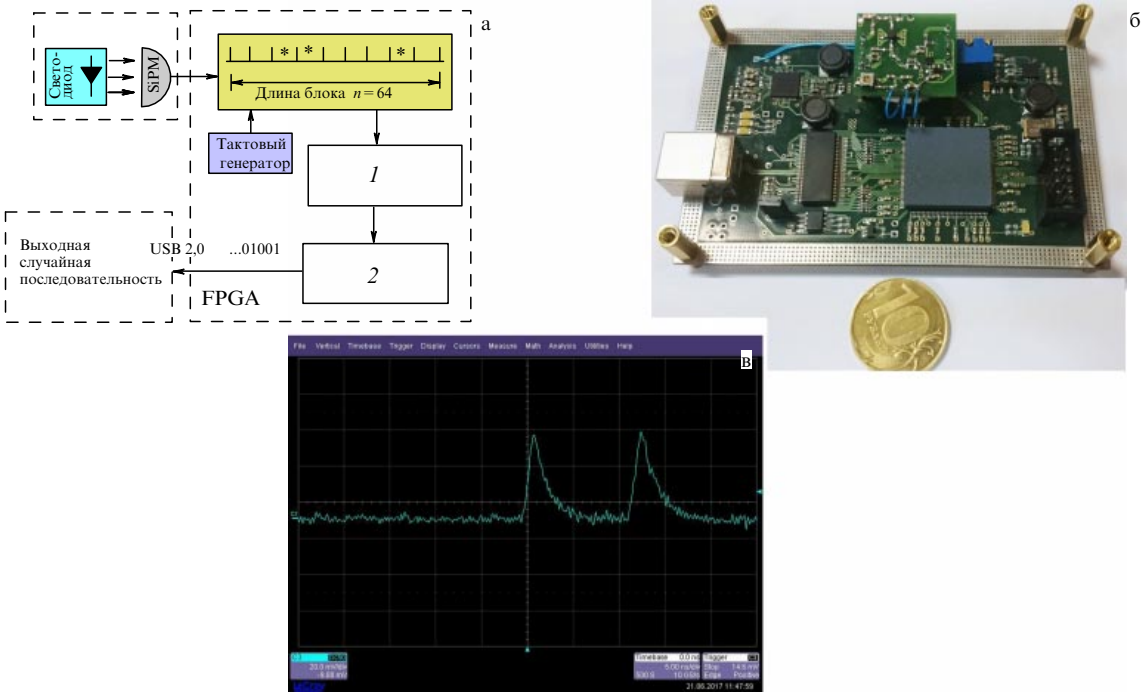


Рис. 6. (а) Функциональная схема квантового генератора случайных чисел: *1* — преобразование отсчётов в номер последовательности через таблицу с треугольником Паскаля, *2* — преобразование номера блока в случайные биты, FPGA — программируемая интегральная логическая схема. (б) Внешний вид устройства. (в) Пример импульсов тока, отражающих процесс фотодетектирования.

дефектах структуры, а затем рекомбинировать и испускать ложный фотон, который регистрируется, т.е. получаются паразитные отсчёты после регистрации реального фотона.

Таким образом, скорость генерации последовательности фотоотсчётов лимитируется как малым средним числом фотонов, так и мёртвым временем детектора.

Для устранения паразитных отсчётов мы использовали не одиночные лавинные фотодетекторы, а матрицу — кремниевый фотоумножитель (SiPM) [22], содержащую более тысячи лавинных детекторов. Среднее число фотонов во временно́м окне T, определяемом тактовой частотой, составляет не более одной тысячной фотона на пиксел, т.е. на отдельный детектор в SiPM, поэтому после регистрации конкретным отдельным фотодетектором фотона вероятность того, что следующий фотон попадёт в тот же самый детектор, крайне мала. В этом случае мёртвое 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 свойства бернуллиевости или независимости последовательности фотоотсчётов, которое можно надёжно экспериментально проверить.

В принципе возможно влияние регистрации в одном пикселе на регистрацию в другом пикселе за счёт электрических наводок в цепи SiPM (так называемые перекрёстные помехи (crosstalk)), что может приводить к искажению статистики фотоотсчётов. Отметим, что исследование возможного паразитного эффекта перекрёстных помех между соседними пикселами на статистику фотоотсчётов проводилось ранее [23] и искажения статистики выявлено не было.

Принципиальная функциональная схема и внешний вид квантового генератора случайных чисел показаны на рис. 6.

В рассматриваемом генераторе получена вероятность детектирования фотона $P(*) = 0{,}3$ на временно́м интервале, отвечающем тактовой частоте $f = 200$ МГц электроники. В асимптотическом пределе, когда длина обрабатываемого блока $n \to \infty$, теоретический предел по скорости генерации случайной последовательности

$$h\big(P(*)\big) f = 0{,}88 \times 200 \approx 176 \text{ Мбит с}^{-1}.$$

В качестве SiPM использовалась матрица детекторов, технология которой разработана в МИФИ-Пульсар (Москва, Россия). Матрица изготовлена в Технологическом центре Национального исследовательского университета "Московский институт электронной техники" (МИЭТ) (Зеленоград, Россия). Матрица SiPM, имевшая чувствительную область приблизительно 1×1 мм, состояла из $N_{\text{pix}} = 1156$ пикселов с активной площадкой 32×32 мкм. Рабочее напряжение (несколько вольт выше пробоя) 40 В [22]. 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 — Field Programmable Gate Array, Intel FPGA (Altera)) с тактовой частотой 200 МГц. В качестве внешнего интерфейса использовался USB 2.0 для питания и вывода результирующей двоичной случайной последовательности в непрерывном режиме. Дополнительное преимущество матрицы SiPM состоит в том, что она имеет большое сопротивление R_{q} нагрузочных резисторов, которое превышает 1 МОм, что приводит к быстрому рассасыванию лавины и низкой вероятности послеимпульсных эффектов. Ещё одна очень важная особенность такого SiPM — довольно короткий по времени сигнал с пиксела, длительностью примерно 1 нс.

10. Статистика фотоотсчётов, оценка среднего числа фотонов на пиксел

Для того чтобы быть уверенным, что генератор действительно работает в квантовом режиме, требуется оценка среднего числа фотонов, падающих на отдельный пиксел SiPM за один такт. При чисто квантовом режиме и пуассоновской статистике числа фотонов в импульсе последовательные регистрации (фотоотсчёты) представляют собой бернуллиевскую последовательнось. Целочисленные интервалы в числе тактов k между последовательными регистрациями представляют собой случайную величину $\xi \in \{0, 1, \ldots\}$, подчиняющуюся геометрическому распределению:

$$P(\xi = k) = \big(1 - P(*)\big)^k P(*). \tag{45}$$

При пуассоновской статистике логарифм вероятности $\ln P(\xi = k)$ (k — число тактов) должен представлять собой линейную зависимость от k. На рисунке 7 показана экспериментальная гистограмма (зависимость $\ln(N(k))$, $N(k)$ — число отсчётов в k-м "ящике" гистограммы), полученная на "эффективной" длине выборки в 6×10^9 тактов.

Линейность графика на рис. 7 демонстрирует пуассоновскую статистику. При большом расстоянии между фотоотсчётами вероятность $P(\xi = k)$ крайне мала, что даёт заметную на графике погрешность для значений $k > 30$.

Для извлечения случайности была выбрана длина блока $n = 64$ такта, что удобно при практической реализации на архитектуре FPGA.

Оценим μ_T — среднее число фотонов на такт из экспериментальных данных.

Экспериментальная гистограмма приведённая на рис. 7, представляет собой линейную зависимость. Полный объём выборки $N_{\text{tot}} = 6 \times 10^9$, число отсчётов в нулевом ящике гистограммы $N(0) = 1897992414$. Отсюда можно получить, учитывая (45), оценку

$$P(*) = P(\xi = 0) \approx \frac{N(0)}{N_{\text{tot}}} = \frac{1897992414}{6 \times 10^9} \approx 0{,}3\,. \tag{46}$$

Данную вероятность можно представить в виде

$$P(*) = \mu_T \eta N_{\text{pic}}\,,$$

где μ_T — среднее число фотонов на один пиксел в SiPM за один такт, $\eta = 0{,}1$ — квантовая эффективность пиксела, $N_{\text{pic}} = 1156$ — число пикселов в матрице.

В итоге получаем $\mu_T = P(*)/(\eta N_{\text{pic}}) \approx 2{,}6 \times 10^{-3}$ фотонов за один такт на 1 пиксел, т.е. на один пиксел за один такт приходится по порядку величины несколько тысячных "долей фотона". Для когерентного состояния с пуас-

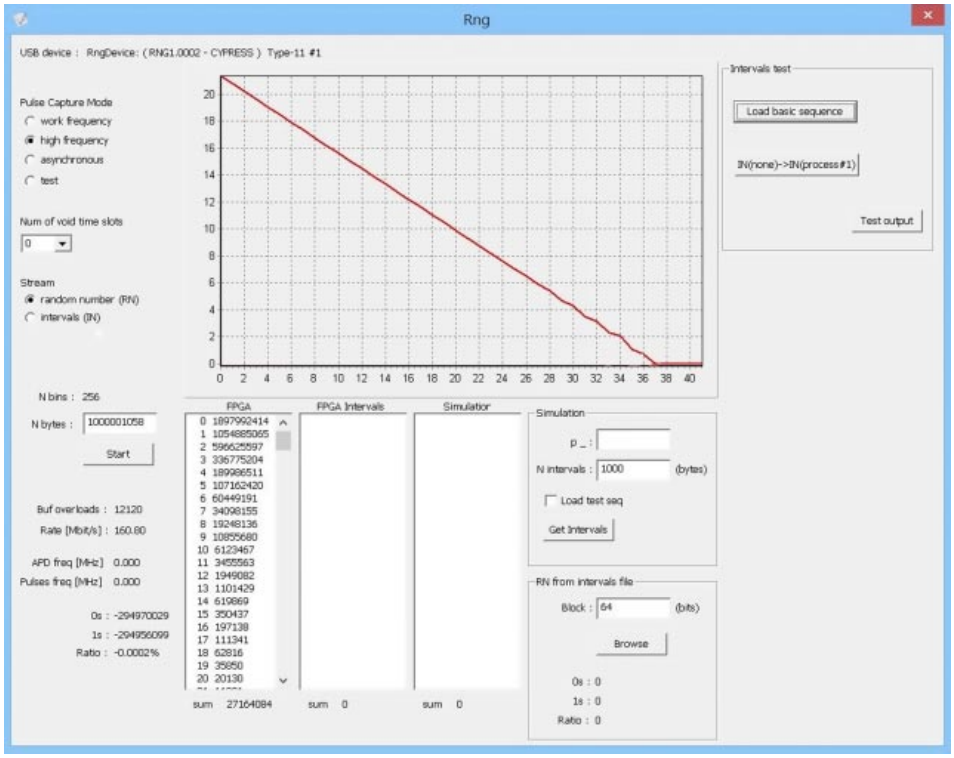


Рис. 7. Пример "живых" экспериментальных данных по проверке бернуллиевского характера последовательностей фотоотсчётов. В колонке FPGA показано число фотоотсчётов в гистограмме в зависимости от числа "пустых" импульсов между последовательными фотоотсчётами. Гистограмма соответствует логарифму геометрического распределения. Скорость генерации случайных бит из последовательности фотоотсчётов 160,80 Мбит с^{-1}. Длина обрабатываемого блока $n = 64$.

соновским распределением с параметром μ_T вероятность появления одного фотона $P_1 = \exp(-\mu_T)\mu_T \approx \mu_T \approx \approx 2{,}6 \times 10^{-3}$, вероятность появления двух фотонов $P_2 = = \exp(-\mu_T)\mu_T^2/2 \approx 3{,}4 \times 10^{-6}$. Таким образом, можно утверждать, что в квантовой части генератора реализован практически однофотонный режим.

Оценим эффективность экстракции случайных 0 и 1 по сравнению с теоретическим асимптотическим пределом $h(P(*))f \approx 0{,}88 \times 200 \approx 176$ Мбит с^{-1}. При длине блока обрабатываемой последовательности $n = 64$ и тактовой частоте 200 МГц скорость генерации случайных 0 и 1 при вероятности фотоотсчёта $P(*) \approx 0{,}3$ (см. рис. 5, кривая *5*, отвечающая $1 - P(*) \approx 0{,}7$) получаем ≈ 160 Мбит с^{-1}, что близко к теоретическому асимптотическому предельному значению 176 Мбит с^{-1}.

11. Как проверять случайность? Статистические тесты случайных последовательностей

Отсутствие доказательств не является доказательством отсутствия.
Из криптографического жаргона [7]

Приведённый в качестве эпиграфа к этому разделу "афоризм", часто упоминаемый при обсуждении тестов на случайность, в полной мере отражает фундаментальную причину — отсутствие "линейки", которой можно измерить случайность. Принципиально невозможно доказать, что данная последовательность из 0 и 1 является истинно случайной, т.е. то, что события, отвечающие появлению 0 и 1, строго равновероятны и независимы, — можно лишь доказать, что данная последовательность не противоречит гипотезе случайности по некоторому статистическому критерию. В этом смысле утверждение о том, что последовательность случайна, начинает зависеть от выбора критерия случайности.

При исследовании мы имеем дело с некоторой конечной последовательностью или выборкой, состоящей из 0 и 1. Общая идеология тестирования последовательности на случайность сводится к следующему.

Допустим, что исследуемая последовательность из 0 и 1 произошла из источника истинной случайности, где вероятности $P(0) = P(1) = 1/2$, выбор 0 и 1 является независимым.

Последнее предположение является аксиоматическим и означает только то, что вероятность любой двоичной последовательности $(\varepsilon_1, \ldots, \varepsilon_n)$, $\varepsilon_i \in \{0, 1\}$, полагается равной

$$P(\varepsilon_1, \ldots, \varepsilon_n) = \prod_{i=1}^{n} P(\varepsilon_i) = 2^{-n}. \qquad (47)$$

Это принципиальный момент. Все остальные математические результаты, так или иначе связанные с разработкой системы тестирования на случайность, следуют из двух равенств (47): первое означает независимость, второе — равновероятность.

[7] Хотя в разговорной лексике также используется изречение: "*Отсутствие доказательств вины не есть доказательство невиновности*" — в том смысле, что презумпции случайности для тестируемой последовательности не существует.

В чём же состоит система тестирования?

Имея "на руках" двоичную последовательность, вообще говоря, очень большого размера и только, мы ничего не можем о ней сказать. Желательно получить некоторый обозримый набор значений, тесно связанных с двоичной последовательностью, который бы в концентрированном виде её представлял и позволил бы разработать разумный критерий, "хороша" она или нет.

Например, первое значение, которое является наиболее естественным, это число единиц в последовательности S_1. По интуитивному пониманию равновероятности, для "хорошей" последовательности величина S_1 должна быть близка к $n/2$.

Двоичные последовательности, очевидно, имеют разный состав нулей и единиц, поэтому при подсчёте величины S_1 будут иметь место отклонения от $n/2$. В связи с этим мы говорим о величине S_1 как о статистике, т.е. как о переменной величине, заданной на множестве наблюдений — равновероятных двоичных последовательностях. Ясно, что точно получить $n/2$ не удастся, поэтому следующий вопрос, который возникает при разработке критерия, — это принять решение, какие отклонения статистики можно считать естественными при "хорошей" двоичной последовательности, а какие — нет.

Этот вопрос решается вероятностными методами. В условиях независимости и равновероятности (47) находится асимптотическое, при $n \to \infty$, распределение вероятностей статистики S_1, в данном случае нормальное распределение. Асимптотическое распределение "хорошо" по двум причинам:

— его аналитический вид известен, предельных распределений в теории вероятностей вообще не так много;

— оно "работает" при произвольных, но, конечно, достаточно больших n.

Для статистики S_1 асимптотическое нормальное распределение позволяет получить вероятность отклонения в виде

$$P\left(\left|S_1 - \frac{n}{2}\right| > t\frac{\sqrt{n}}{2}\right) = 2\bigl(1 - \Phi(t)\bigr)\,,\ t \geqslant 0\,, \tag{48}$$

где $\Phi(t)$ — функция стандартного нормального распределения,

$$\Phi(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}t\,.$$

Пусть вероятность (48) мала. Это означает, что отклонение статистики S_1 от величины $n/2$ более чем на $t\sqrt{n}/2$ маловероятно, т.е. такое отклонение недопустимо, если тестируемая последовательность — "хорошая".

Зададим малое значение вероятности α и найдём значение t_α из уравнения $2(1 - \Phi(t)) = \alpha$.

Теперь у нас всё готово для формулировки критерия согласия с гипотезой $\mathcal{H}$ о независимости и равновероятности двоичной последовательности:

— если $|S_1 - n/2| \leqslant t_\alpha\sqrt{n}/2$, то гипотеза $\mathcal{H}$ принимается, последовательность — "хорошая";

— если $|S_1 - n/2| > t_\alpha\sqrt{n}/2$, то гипотеза $\mathcal{H}$ отклоняется, последовательность — "плохая".

Казалось бы, это всё, задаём конкретное числовое значение α — уровень значимости критерия (обычно $\alpha = 0{,}001 - 0{,}1$) — и отправляем критерий в работу для тестирования последовательностей, снимаемых с генератора случайных чисел.

Но здесь появляется одно обстоятельство, которое несколько портит нашу стройную картину.

Последовательность считается "плохой" хотя и с малой, но не нулевой вероятностью α (гипотеза $\mathcal{H}$ отклоняется), когда на самом деле она "хорошая". Таким образом, проводя тестирование, например, N раз, мы получим выход за границы критерия примерно в αN случаях при условии, что последовательность "хорошая". Что делать с этими последовательностями? Исключать из рассмотрения? А со всей объединённой выборкой размера nN? Оставить её для "употребления", например, если полученное относительное число выходов действительно близко к α? Насколько "допустимо" близко? Мы рискуем пойти по кругу.

Есть ещё один тонкий момент. Что следует предпринять, если во всех N тестированиях статистика ни разу не вышла за границы критерия, т.е. согласие "слишком хорошее"? Например, для регулярной последовательности 0101010101... во всех тестированиях отклонение будет нулевым...

В методических рекомендациях Национального института стандартов и технологий США (National Institute of Standards and Technology, NIST) [24], являющихся в настоящий момент общепринятой практикой, представлена следующая концепция тестирования. Мы рассмотрим её на примере статистики ζ отклонения от среднего числа единиц вида

$$\zeta = \frac{|S_1 - n/2|}{\sqrt{n/4}}\,,$$

имеющей функцию распределения $F_\zeta(t) = 2\Phi(t) - 1$. Рекомендации NIST основаны на том математическом факте, что распределение случайной величины $\xi = 1 - F_\zeta(\zeta)$ является равномерным на отрезке $[0, 1]$. Этот результат остаётся верным для любой случайной величины ζ со своей функцией распределения $F_\zeta(t)$.

Пусть тестируются N двоичных последовательностей длиной n каждая, в результате чего вычисляются N отклонений $\zeta^{(i)} = |S_1^{(i)} - n/2|/\sqrt{n/4}$, $i = \overline{1, N}$, и соответственно N значений $\xi^{(i)} = 1 - F_\zeta(\zeta^{(i)})$, которые называются *p-value*.

Среди отклонений $\zeta^{(i)}$ есть, конечно, значения $\zeta^{(i')}$, которые велики или даже выходят за границу t_α критерия. Нетрудно увидеть, что для них значения *p-value* $\xi^{(i')} = 1 - F_\zeta(\zeta^{(i')})$ "тяготеют" к нулю. Если $\zeta^{(i'')}$ малы, т.е. согласие "очень хорошее", то значения $\xi^{(i'')} = 1 - F_\zeta(\zeta^{(i'')})$ "тяготеют" к единице.

Никакие отдельные отрезки двоичной последовательности, даже "плохие", не выбрасываются, значения *p-value* не исключаются.

Строится гистограмма частот попадания значений *p-value* $\xi^{(i)} = 1 - F_\zeta(\zeta^{(i)})$, $i = \overline{1, N}$ в интервалы разбиения единичного отрезка на 10 равных частей $[0,0,1)$, $[0,1,0,2), \ldots, [0,9,1)$:

$$v_1, v_2, \ldots, v_{10},\ \sum_{k=1}^{10} v_k = N\,.$$

Если согласие "плохое", то гистограмма имеет перекос влево, если "хорошее" — то вправо.

Теперь уже эта гистограмма тестируется на равномерность по критерию согласия χ-квадрат.

Вычисляется статистика

$$\chi = \sum_{k=1}^{10} \frac{(v_k - Np_k)^2}{Np_k}\,, \quad p_k = \frac{1}{10}\,,$$

которая имеет асимптотическое распределение χ-квадрат с $m = 9$ степенями свободы и функцией распределения

$$F_\chi(z) = \frac{1}{2^{m/2}\Gamma(m/2)} \int_0^z \exp\left(-\frac{x}{2}\right) x^{m/2-1}\,\mathrm{d}x\,,$$

где $\Gamma(y)$ — гамма-функция. Задаётся значение $\alpha = 0{,}0001$ и вычисляется z_α из уравнения $1 - F_\chi(z) = \alpha$. Вероятность выхода статистики χ за границу z_α равна α.

Если

$$\chi = \sum_{k=1}^{10} \frac{(v_k - Np_k)^2}{Np_k} > z_\alpha\,, \tag{49}$$

то совокупная двоичная последовательность размером nN считается не прошедшей критерий, основанный на статистике числа единиц S_1.

В методических рекомендациях NIST [24] подобная процедура предлагается для целого ряда приведённых там статистик, каждая из которых нацелена на "обнаружение" определённого типа отклонения распределения двоичной последовательности от гипотезы $\mathcal{H}$, рекомендованные значения: $n = 10^6$, $N = 10^2$.

Как итог исследования, перечисляются все критерии согласия, отмечаются те, в которых статистика χ вышла за границу z_α, т.е. где критерий не пройдён. Делать вывод в целом о пригодности генератора случайных чисел к работе предоставляется экспериментатору.

В настоящее время сформировалось несколько рекомендованных наборов тестов (критериев согласия, статистик) на случайность [24–26]. Набор тестов NIST [24] является минимально необходимым и служит основанием для исследования последовательностей другими наборами специальных тестов.

С точки зрения прохождения двоичной последовательностью критериев на случайность нельзя исключить ситуацию, в которой последовательность может маскироваться под случайную — ограниченный набор статистик ведёт себя так же, как набор статистик для истинно случайной последовательности.

В частности, хорошо известно, что выходные двоичные последовательности всех современных средств криптографической защиты информации проходят любые критерии на случайность, но не являются таковыми в истинном смысле. При длине ключа в 256 бит они отражают случайность (равновероятность) выбора из множества в 2^{256} двоичных векторов, но никак не из 2^n векторов, где n — длина выходной последовательности.

Истинную случайность может гарантировать только квантовый генератор случайных чисел при надлежайшей настройке его технических параметров.

12. Проверка результатов различных тестов (статистик) на однородность. Экспериментальные результаты

Уровень значимости α критерия (теста, статистики) можно понимать как то, с какой вероятностью идеальный генератор может генерировать последовательности, которые будут выглядеть для нас как неслучайные. Уровень α мы задаём сами, обычно $\alpha \in [0{,}001, 0{,}1]$ [24].

Каждый тест пытается найти свою "улику на неслучайность", например на неравновероятность, наличие корреляций, скрытых периодов, марковской зависимости, на повышенную вероятность появления некоторого двоичного "шаблона" и т.п. Можно сказать, что каждый тест ориентирован на обнаружение своего, специфического, отклонения от истинно случайной последовательности. Такое отклонение называют ещё конкурирующей гипотезой.

Если последовательность истинно случайная, а уровень значимости один и тот же для всех тестов, то она должна "однородно" проходить разные тесты. С более формальной точки зрения это означает, что при большом числе тестирований доля последовательностей, прошедших каждый тест при заданном α, должна быть примерно одинаковой, а именно примерно равной $1 - \alpha$, и не зависеть от теста.

Поэтому дополнительной проверкой на случайность может быть проверка на "однородность" (рис. 8) числа последовательностей, прошедших тестирование для совокупности критериев. По этой причине такую вторичную проверку (в русскоязычной литературе часто называемую вторичной маркировкой) иногда называют "тест тестов". Данная процедура не имеет строгого логического обоснования и основывается, по сути, на качественных соображениях.

Пусть каждый тест применяется к M различным двоичным последовательностям. Относительная доля последовательностей, прошедших тестирование, должна укладываться по каждому тесту в интервал "три сигма" [27]:

$$(1 - \alpha) \pm 3\sqrt{\frac{\alpha(1-\alpha)}{M}}\,.$$

Пример результатов подсчёта относительной доли последовательностей, прошедших тестирование для совокупности критериев, приведён в табл. 3. Полная длина составляла 8×10^9 бит (точнее значение 8000000464) (см. рис. 1), число блоков последовательностей $M = 2000$, каждый длиной $L = 4 \times 10^6$ бит, $\alpha = 0{,}01$. Интервал "три сигма" равен $[0{,}983, 0{,}997]$.

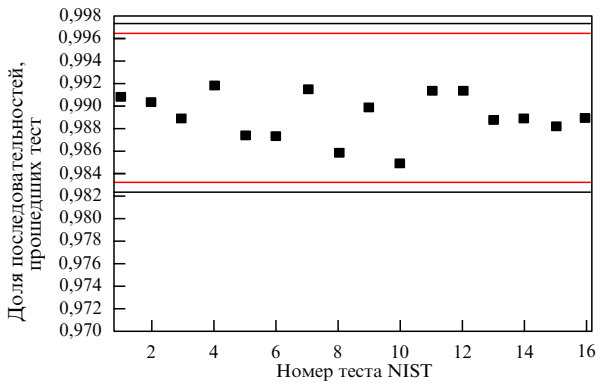


Рис. 8. (В цвете онлайн.) Пример теста на однородность *p-value* для 16 тестов NIST — гистограмма *p-value* для каждого из 16 тестов. Тестировалось $M = 2000$ последовательностей, каждая длиной $L = 4 \times 10^6$ бит. Красными горизонтальными линиями показан интервал "три сигма". Как видно, тест на однородность значений проходится с запасом.

Таблица 3. Доля последовательностей, прошедших различные тесты

№	Название теста	Доля последовательностей $M = 2000$ $L = 4000000$
1	Frequency Test	0,9910
2	Block Frequency	0,9905
3	Cumulative Sums	0,9890
4	Cumulative Sums Reverse	0,9920
5	Runs	0,9875
6	Longest Runs	0,9875
7	Rank	0,9915
8	Fast Fourier Transform (FFT)	0,9860
9	Non Overlapping Template	0,9910
10	Overlapping Template	0,9850
11	Universal	0,9915
12	Approximate Entropy	0,9915
13	Random Excursions	0,9899
14	Random Excursions Variant	0,9893
15	Serial	0,9890
16	Linear Complexity	0,9890

13. Заключение

В недалёком прошлом алгоритмами получения случайных двоичных последовательностей и методов их тестирования занимались в основном специалисты, работающие в области криптографии. Физическое научное общество также использовало случайность для своих целей, например для моделирования процессов в задачах ядерной и статистической физики. Ввиду объективной изолированности криптографов и физиков в каждом научном обществе разрабатывались свои алгоритмы и формировалось своё понимание того, что такое случайность. Появление квантовой информатики, в частности квантовой криптографии, инициировало в последнее время неизбежное взаимное проникновение данных научных обществ и разрабатываемых ими методов.

В настоящей статье мы пытались показать, что получение квантовой случайности и проверка её свойств, по существу, ничем не отличаются от любого физического эксперимента по проверке того или иного теоретического закона. Любой эксперимент такого рода сводится к выделению в "чистом виде" тех факторов, которые мы хотим проверить, при устранении нежелательных внешних воздействий, искажающих результаты эксперимента. Любой физический эксперимент повторяется конечное число раз в одних и тех же условиях, после чего делается вывод о подтверждении или опровержении исследуемого закона. Однако логически ниоткуда не следует, что повторённый ещё раз эксперимент приведёт к тем же результатам.

Для пояснений имеет смысл провести параллель между "тестом тестов" при проверке случайности и физическим экспериментом. Пусть выполнена большая серия экспериментов по проверке физического закона, которая подтверждает его правильность. Но любой физический эксперимент имеет некоторую погрешность. Согласие эксперимента с теорией принимается, если результаты эксперимента укладываются в погрешность. Выбор допустимой погрешности осуществляется, по сути, "вручную" и не имеет математического обоснования. Пусть каждая серия экспериментов повторяется многократно. Если результат какой-то серии вышел за пределы погрешности, то следует ли считать, что физический закон несправедлив или это нужно списать на "нечистоту" конкретной серии? Если доля таких неудачных серий мала по какому-то критерию, то на эти выбросы не следует обращать внимания. Критерий малости опять выбирается "вручную". В остальных удачных сериях уклонения от идеальной (теоретической модели) должны быть однородными по сериям. Здесь просматривается явная аналогия с "тестом тестов" при проверке случайности.

В полной мере это относится к экспериментам для получения случайности только, возможно, в ещё более отчётливом и концентрированном виде.

Пусть генерируется последовательность 0 и 1 длиной 10^9 бит. Всего таких последовательностей 2^{10^9}. Напомним, что число атомов в видимой части Вселенной оценивается как $2^{256} \approx 10^{77}$. С умозрительной точки зрения для проверки равновероятности всех 2^{10^9} последовательностей их надо, как минимум, в таком количестве сгенерировать и посмотреть на частоту, с которой они встречаются. Очевидно, что это невозможно. Проводя тесты только над одной последовательностью, мы фактически пытаемся сделать умозаключение о свойствах экспоненциально большой "совокупности".

Источник (*лазер*) *приготавливает квантовое* (*квазиоднофотонное*) *состояние*[8]*, которое подвергается измерению. Если проекционный постулат* (*формула* (3)) *справедлив и эксперимент проведён "чисто", то первичная последовательность фотоотсчётов является бернуллиевской. Если последовательность бернуллиевская, то из неё* (*и это строго доказуемый математический факт*) *извлекается истинно случайная последовательность* 0 *и* 1. *По этой логике проверка последовательностей на случайность, по сути, означает проверку бенуллиевского характера последовательности фотоотсчётов, который является следствием проекционного постулата* (*формула* (3)) — *одного из фундаментальных постулатов квантовой механики.*

В этом смысле проверка на случайность ничем не отличается от интерпретации любого физического эксперимента. Тем не менее (см. эпиграф в начале заметок), не имея возможности даже записать такие огромные последовательности, мы можем выносить суждения о их свойствах.

Исследования в этой области были инициированы вполне практическими целями, приведённый выше пример квантового генератора *на идейном уровне является простым физическим устройством*, принципы работы которого основаны на фундаментальных законах квантовой физики и могут быть поняты без специальных знаний. В связи с этим авторы не могут отказать себе в удовольствии привести следующий афоризм: "*Всё нужное просто, что сложно — то не нужно*" (М.Т. Калашников).

[8] В идеале для проверки проекционного постулата хотелось бы использовать строго однофотонное фоковское состояние. Однако, несмотря на многочисленные эксперименты, строго однофотонный источник на сегодняшний день отсутствует. Для строго однофотонного источника корреляционная функция второго порядка должна иметь провал строго до нуля [8, 10]. В существующих экспериментах провал корреляционной функции строго до нуля не продемонстрирован, что означает присутствие неоднофотонных фоковских компонент в излучении. По этой причине использование сильно ослабленного когерентного состояния является предпочтительным и более надёжным.

Благодарности. Выражаем благодарность нашим коллегам по Академии криптографии Российской Федерации за многочисленные обсуждения и поддержку. Авторы также выражают благодарность коллегам по Центру квантовых технологий Московского государственного университета им. М.В. Ломоносова: К.А. Балыгину, И.Б. Боброву, В.И. Зайцеву, В.А. Кирюхину, А.Н. Климову, С.П. Кулику, И.В. Синильщикову — за приятное и активное сотрудничество. Выражаем благодарность Елене Поповой и Сергею Виноградову за любезно предоставленные образцы SiPM и обсуждения.

Список литературы

1. Bennett C H, Brassard G, in *Proc. of the IEEE Intern. Conf. on Computers, Systems, and Signal Processing, Bangalore, 10–12 December 1984* (Piscataway, NJ: IEEE, 1984) p. 175
2. Koç Ç K (Ed.) *Cryptographic Engineering* (New York: Springer, 2009)
3. Василенко В В *Информационные войны* (3) 23 (2012)
4. Srinivasan S et al., in *2010 IEEE Symp. on VLSI Circuits, 16–18 June 2010, Honolulu, HI, USA* (Piscataway, NJ: IEEE, 1984) p. 203, https://doi.org/10.1109/VLSIC.2010.5560296
5. Galton F *Natural Inheritance* (London: Macmillan, 1894)
6. Herrero-Collantes M, Garcia-Escartin J C *Rev. Mod. Phys.* **89** 015004 (2017)
7. Einstein A *Ann. Physik* **17** 132 (1905)
8. Клышко Д Н *Фотоны и нелинейная оптика* (М.: Наука, 1980); Пер. на англ. яз.: Klyshko D N *Photons and Nonlinear Optics* (New York: Gordon and Breach, 1988)
9. Клышко Д Н, Масалов А В *УФН* **165** 1249 (1995); Klyshko D N, Masalov A V *Phys. Usp.* **38** 1203 (1995)
10. Mandel L, Wolf E *Optical Coherence and Quantum Optics* (Cambridge: Cambridge Univ. Press, 1995); Пер. на русск. яз.: Мандель Э, Вольф Э *Оптическая когерентность и квантовая оптика* (М.: Физматлит, 2000)
11. Shannon C E *Bell Syst. Tech. J.* **27** 379 (1948)
12. Shannon C E *Bell Syst. Tech. J.* **27** 623 (1948)
13. Шеннон К *Работы по теории информации и кибернетике* (М.: ИЛ, 1963)
14. Cover T M, Thomas J A *Elements of Information Theory* (New York: Wiley, 1991)
15. Молотков С Н *Письма в ЖЭТФ* **105** 374 (2017); Molotkov S N *JETP Lett.* **105** 395 (2017)
16. Балыгин К А и др. *ЖЭТФ* **153** 879 (2018); Balygin K A et al. *J. Exp. Theor. Phys.* **126** 728 (2018)
17. Balygin K A et al. *Laser Phys. Lett.* **14** 125207 (2017)
18. Балыгин К А и др. *Письма в ЖЭТФ* **106** 451 (2017); Balygin K A et al. *JETP Lett.* **106** 470 (2017)
19. Von Neumann J, in *Applied Mathematics Series* Vol. 12 (Washington, DC: U.S. National Bureau of Standards, 1951) p. 36; Reprinted in *Neumann's Collected Works* Vol. 5 (Oxford: Pergamon Press, 1963) p. 768
20. Feller W *An Introduction to Probability Theory and Its Applications* 2nd ed. (New York: Wiley, 1957); Пер. на русск. яз.: Феллер В *Введение в теорию вероятностей и её приложения* Т. 1 (М.: Мир, 1964)
21. Бабкин В Ф *Проблемы передачи информации* **7** (4) 13 (1971)
22. Buzhan P et al. *Nucl. Instrum. Meth. Phys. Res. A* **567** 78 (2006)
23. Kalashnikov D A, Tan S-H, Krivitsky L A *Opt. Express* **20** 5044 (2012)
24. Computer Security esource Center, http://csrc.nist.gov/rng/SP800-22b.pdf
25. Knuth D E *The Art of Computer Programming* Vol. 2 (Cambridge: Addison Wesley, 1981)
26. Marsaglia G, http://stat.fsu.edu/pub/diehard
27. Cramer H *Mathematical Methods of Statistics* (Princeton, NJ: Princeton Univ. Press, 1946)

Extraction of quantum randomness

I.M. Arbekov [(1)], **S.N. Molotkov** [(1, 2, 3, 4, a)]
[(1)] *Academy of Cryptography of the Russian Federation, ul. Yartsevskaya 30, 121552 Moscow, Russian Federation*
[(2)] *Institute of Solid State Physics, Russian Academy of Sciences, ul. Akademika Osip'yana 2, 142432 Chernogolovka, Moscow region, Russian Federation*
[(3)] *Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics, Leninskie gory 1, str. 52, 119991 Moscow, Russian Federation*
[(4)] *Lomonosov Moscow State University, Quantum Technology Center, Leninskie gory 1, str. 35, 119991 Moscow, Russian Federation*
E-mail: [(a)] *molotkov@issp.ac.ru*

The nature of randomness and constructive and provable methods to obtain (extract) it from observations of physical systems are discussed. True randomness, which exists only in a microcosm in the quantum-mechanical description of physical systems, is a fundamental property of quantum systems, which manifests itself in the outcomes of measurements upon quantum systems. The classical description of physical systems does not include any randomness and, in fact, it is introduced 'manually' by means of uncertainty — unknown initial conditions. Methods to really 'feel' quantum randomness are discussed using the example of a quantum device, a random number generator. Issues related to the 'proof' of randomness — testing of numerical sequences — are reviewed, and logical constructions that underlie such testing are analyzed. A mathematical apparatus is used to this end, which does not require special academic training, so standard knowledge from university courses on quantum mechanics and probability theory is sufficient. The authors aim to track a unified logical path from the origin of randomness in the quantum domain to its extraction, physical implementation, and testing.

Keywords: quantum random number generators, randomness extraction

PACS numbers: 03.67.Dd, 42.50.Ex

Bibliography — 27 references

Received 15 May 2020, revised 28 October 2020

Uspekhi Fizicheskikh Nauk **191** (6) 651 – 669 (2021)

Physics – Uspekhi **64** (6) (2021)

DOI: https://doi.org/10.3367/UFNr.2020.11.038890

DOI: https://doi.org/10.3367/UFNe.2020.11.038890