Письма в ЖЭТФ, том 105, вып. 6, с. 374 – 380

О предельных характеристиках квантовых генераторов случайных чисел при различных группировках фотоотсчетов

С. Н. Молотков[1)]

Институт теоретической физики им. Л.Д. Ландау РАН, 142432 Черноголовка, Россия

Академия Криптографии Российской Федерации, 121552 Москва, Россия

Факультет вычислительной математики и кибернетики Московского государственного университета им. М.В. Ломоносова, 119991 Москва, Россия

Поступила в редакцию 19 декабря 2016 г.

После переработки 7 февраля 2017 г.

Рассмотрены различные способы группировки фотоотсчетов, из которых формируется последовательность случайных чисел. Показано, что группировка фотоотсчетов, приводящая к распределению Ферми–Дирака, позволяет выйти на теоретический предел по скорости генерации случайных чисел.

DOI: 10.7868/S0370274X17060108

Введение. Генераторы случайных чисел широко используют в различных областях науки и техники, например, в физике при моделировании методом Монте-Карло. В криптографии, в том числе и квантовой, генератор случайных чисел является одним из основных элементов, характеристики которого определяет криптостойкость системы. Статистические свойства случайной последовательности и скорость генерации являются главными критериями качества таких генераторов. Генераторы случайных чисел условно делятся на два типа – математические и физические. Первый тип – генераторы, в которых случайная последовательность получается в результате некоторого математического преобразования, обычно рекурсивного, требующего для инициализации на первом шаге начальное затравочное случайное число. Выходные последовательности являются псевдослучайными, так как полностью зависят от начальных условий.

В серьезных криптографических системах с симметричным шифрованием для генерации ключей используют только физические генераторы случайных чисел. Физические генераторы также можно разделить на два типа – классические и квантовые. Физические классические генераторы случайных чисел основаны на извлечении случайности из некоторого физического процесса, эволюция которого во времени описывается законами классической физики. Эволюцию любой, даже сколь угодно сложной классической системы, описывают дифференциальными уравнениями. Последовательности, получаемые на выходе такого генератора, также не являются истинно случайными, поскольку полностью определяются начальными условиями. Их случайность связана только с неопределенностью начальных условий для системы. Поэтому такие генераторы нельзя считать истинно случайными. Второй тип физических генераторов – квантовые [1]. Извлечение случайной последовательности чисел основано на измерении квантовой системы. *Нет фундаментальных запретов на то, что эволюция любой классической системы, в принципе, может быть предсказана, если известны начальные условия. В отличие от классической физики измерения над квантовой системой, каждый раз приготовленной в определенном, 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.*

Последовательность результатов измерений $(x_1, x_2, \dots x_n)$ над квантовой системой в общем случае не является статистически независимой, т.е. функция распределения не распадается на произведение $P(x_1, x_2, \dots x_n) \neq P(x_1)P(x_2)\dots P(x_n)$. В этом случае число истинно случайных бит дается минимальной энтропией Реньи $H_{\min} = -\log(\max_{P(x_1,x_2,\dots x_n))} P(x_1, x_2, \dots x_n))$. Существует целое математическое направление – экстракторы случайности, а также процедуры, которые позволяют извлечь случайность из распределения $P(x_1, x_2, \dots x_n) \neq P(x_1)P(x_2)\dots P(x_n)$ [2]. Однако проблема состоит в том, что в реальной ситуации функция распределения $P(x_1, x_2, \dots x_n)$ неизвестна. Кроме того, сама процедура получения

[1)]e-mail: sergei.molotkov@gmail.com

истинно случайной последовательности требует затравочной случайной последовательности для случайного выбора хэш-функции.

Поэтому более удобно выбрать другой путь, а именно, такие квантово-механические измерения, которые обеспечивают статистическую независимость последовательных актов измерений. Разумеется, что проверка правильности выбора и экспериментальной реализации физического процесса, который обеспечивает статистическую независимость последовательных результатов квантово-механических измерений, проверяется на финальной стадии по полученной случайной последовательности 0 и 1. Это, впрочем, имеет место в обоих упомянутых случаях.

Такими квантовыми процессами могут быть, например, α-распад, фотоэффект и т.д. Фотоэффект был открыт Столетовым еще до создания квантовой механики, поэтому не был объяснен. Последовательное объяснение фотоэффекта было дано Эйнштейном с использованием понятия квантов. Как известно, первопричина пуассоновской статистики фотоотсчетов при детектировании лазерного излучения носит принципиально квантовый характер и обусловлена поглощением фотонов атомами [3]. В реальной ситуации при создании квантовых генераторов случайных чисел приходится ослаблять лазерное излучение до квазиоднофотонного уровня, чтобы фотоотсчеты не были слишком частыми. Это связано с техническими ограничениями работы лавинных детекторов, точнее конечным временем восстановления детекторов после регистрации, что дает ограничения на скорость формирования случайной последовательности. Восстановление детектора до следующего акта регистрации обеспечивает статистическую независимость последовательных фотоотсчетов. *Квантовые генераторы случайных чисел (КГСЧ), хотя и выдают случайные последовательности с хорошими статистическими свойствами, являются достаточно медленными по сравнению с генераторами псевдослучайных последовательностей. Возникает принципиальный вопрос о нахождении предельных характеристик КГСЧ. Прояснению этого вопроса посвящено данное сообщение.*

Вероятность обнаружить m фотонов во временно́м окне T при пуассоновской статистике есть [3]

$$P_T(m) = e^{-\mu}\frac{\mu^m}{m!},$$

где μ – среднее число фотонов в излучении.

Поскольку лавинные детекторы не различают число фотонов, то случайными событиями являются: отсутствие фотоотсчета во временно́м окне (такте) T – $\sqcup$ или фотоотсчет – $*$ (рис. 1a)). При этом вероятность фотоотсчета (от одного, двух и более фотонов в окне T) есть $P(*) = 1 - e^{-\mu}$, соответственно, отсутствие отсчета $P(\sqcup) = e^{-\mu}$.

Для дальнейшего важно, что процедуре извлечения случайности требуется только статистическая независимость фотоотсчетов и не требуется знание самих вероятностей $p = P(\sqcup)$ *и* $1 - p = P(*)$. Под случайной последовательностью понимается последовательность 0 и 1, которые независимы в каждой позиции, и вероятность $P(0) = P(1) = 1/2$.

Возможны различные методы извлечения случайности, основанные на разных способах группировки фотоотсчетов. Например, метод получения случайной последовательности, предложенный фон Нейманом еще в 1951 году [4], состоит в следующем. Пошагово просматривается последовательность $\{*, \sqcup\}^n$. Если две последовательные позиции $*\sqcup$, то они заменяются на $*\sqcup \to 0$. Если встречается комбинация $\sqcup*$, то она заменяется $\sqcup* \to 1$. Любые другие парные комбинации $**$, $\sqcup\sqcup$ отбрасываются. Полученная последовательность 0 и 1 является равномерно распределенной случайной последовательностью. Вероятность 0 и 1 в новой последовательности строго равна $1/2$; 0 и 1 равновероятны при любом значении исходных вероятностей $1 - p \to *$ и $p \to \sqcup$. В данном методе, даже в самом лучшем случае, когда $p \approx 1 - p$, теряется половина исходной последовательности, также еще не использованы методы теории информации, поэтому не получается извлечь максимально возможную длину случайной последовательности, содержащейся в физическом процессе – последовательности фотоотсчетов.

Качественные соображения. Метод фон Неймана состоит в группировке исходных физических событий в различные классы таким образом, чтобы все представители из одного класса имели одинаковую вероятность. Далее каждому представителю класса сопоставляется битовая последовательность 0 и 1. Точнее говоря, если перенумеровать в лексикографическом порядке все последовательности в одном классе, то бинарное представление номера будет блоком выходной случайной последовательности.

Как следует из теории информации [5], для источников без памяти имеет место асимптотическое равнораспределение выходных последовательностей. Неформально данное свойство означает следующее. При длинной последовательности имеется $\approx 2^{nh(p)}$ типичных последовательностей, которые имеют равную вероятность ($h(p) = -p\log(p) - (1-p)\log(1-p)$ – бинарная энтропийная функция Шеннона [5]). Множество типичных последовательностей при $n \to \infty$

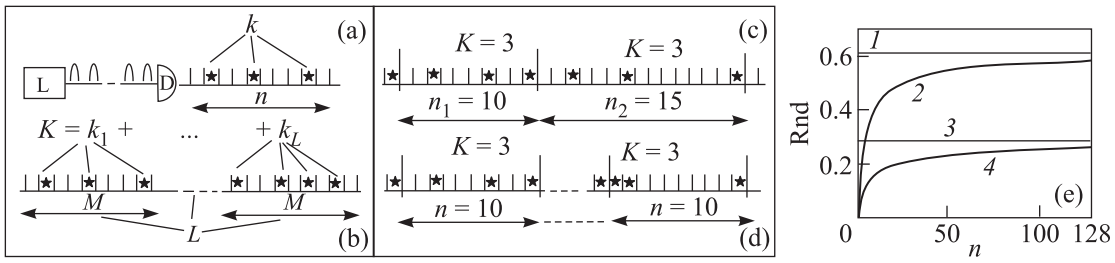

Рис. 1. (a) – Схематическое представление группировки фотоотсчетов, приводящих к распределению Ферми–Дирака. L – лазер, D – фотодетектор. (b) – Группировка по Ферми–Дираку с несколькими системами. (c) – Пример группировки по распределению Бозе–Эйнштейна ($k = 3$, длина блока n переменная. (d) – Пример группировки по распределению Бозе–Эйнштейна, пример блоков с фотоотсчетами, входящими в один класс. (e) – Зависимости числа случайных бит Rnd в пересчете на один такт от длины блока n для группировки по Ферми–Дираку. Горизонтальные линии отвечают теоретическому пределу по числу случайных бит. Вероятность отсутствия фотоотсчета $p(\sqcup) = 0.85$ – кривые *1*, *2*, и $p(\sqcup) = 0.95$ – кривые *3*, *4*

реализуется с вероятностью единица. В асимптотическом пределе каждая типичная строка содержит $\approx n(1-p)$ отсчетов $*$ и соответственно $\approx np$ пропусков $\sqcup$.

Отсюда фактически вытекает рецепт извлечения случайной последовательности 0 и 1. Для этого достаточно перенумеровать в лексиграфическом порядке все типичные последовательности. Присвоить свой номер $0 \le \mathrm{Num}(x) \le [2^{nh(p)}]-1$ ($x = x_1, x_2, \ldots x_n$ последовательность * и $\sqcup$) каждой последовательности. Поскольку все последовательности равновероятны, то бинарное представление номера и есть случайная последовательность 0 и 1. Длина случайной последовательности равна числу двоичных разрядов, необходимых для записи максимального значения номера $\mathrm{Num}(x)$. Следовательно, в асимптотическом пределе из последовательности фототсчетов $\{\sqcup, *\}^n$ длины n можно извлечь случайную последовательность длины $[nh(p)]$ бит.

В реальной ситуации всегда приходится ограничиваться конечной длиной последовательности с фотоотсчетами, т.е. разбивать последовательности фотоотсчетов на блоки длиной n. При этом возможны различные способы группировки последовательностей в равновероятные классы.

Сразу возникает вопрос – какой способ группировки при заданном n дает максимальный выход случайных 0 и 1 в пересчете на такт? Второй вопрос – обеспечивает ли данный метод группировки выход на асимптотический режим, т.е. позволяет ли данная группировка в асимптотическом пределе ($n \to \infty$) извлечь предельно допустимое количество случайных бит? В обзоре [1] приведены ссылки на практически все работы, которые используют извлечение случайности из последовательностей фотоотсчетов. Однако до сих пор так и неясно, имеется ли конструктивный способ, который позволяет дотянуться до теоретического предела по случайности, и каких вычислительных ресурсов это требует? *Важно также, чтобы метод группировки позволял обрабатывать последовательность фотоотсчетов и извлекать случайную последовательность вычислительно эффективно, т.е. с полиномиальными вычислительными ресурсами по длине обрабатываемой последовательности.*

Интуитивно понятно, что пуассоновский процесс содержит некоторое максимальное количество случайных 0 и 1, которое не может быть превышено. Число случайных бит на такт может быть получено из последовательности фотоотсчетов и пропусков $\{*, \sqcup\}^n$ и, очевидно, зависит от способа группировки фотоотсчетов – от способа отображения последовательности $\{*, \sqcup\}^n \to \{0,1\}^l$ ($l \le n$).

Формально возможно неограниченное число способов группировки, приводящее к различным распределениям. Однако удивительным образом в разных разделах физики, математики и теории информации естественным образом возникает лишь небольшое число одних и тех же распределений. Это связано с одними и теми же ограничениями, хотя контекст задачи может быть разным. Например, распределение Ферми–Дирака возникает в физике в задаче о размещении фермиевских частиц по уровням энергии с дополнительным ограничением – одна частица в одном состоянии. При фотодетектировании распределение Ферми–Дирака возникает при "размещении" фотоотсчетов по временны́м интервалам с дополнительным ограничением: в каждом временно́м окне – один акт регистрации детектора (детектор не различает число фотонов, регистрирует-

ся только импульс тока). Фактически это та причина, по которой при группировке по распределению Ферми–Дирака достигается максимум извлекаемой случайности.

Фотодетектирование по своей природе является квантовым процессом, поэтому он используется многими авторами при разработке квантовых генераторов случайных чисел. Например, в работе [1a] использована группировка относительных времен регистрации фотоотсчетов (аналогичный способ применен в работе [7]), приводящая к распределению Бозе–Эйнштейна. При этом способ выравнивания последовательности требовал экспоненциально большой таблицы по длине последовательности. В работе [1b] использована группировка фотоотсчетов по Ферми–Дираку, при этом применены два детектора. И до конца неясно, какие ресурсы нужны для выхода на предельные характеристики.

В каждой из работ приводится свой способ, чем-то отличный от других. Однако до сих пор внятно не обозначен теоретический предел по извлечению случайности при заданной вероятности фотоотсчетов, и не ясно каких ресурсов это стоит – например, объем памяти при обработке.

Ниже простыми средствами показано, что количество максимальной случайности – это энтропия Шеннона, на которую выходит группировка по Ферми–Дираку. И самое важное, требуются только полиномиальные (квадратичные) ресурсы по объему памяти (сложности обработки). Это позволяет выйти на теоретический предел уже при таблице размером 128×128, что дает конструктивный завершенный ответ на поставленный вопрос. При реализации требуется только один детектор, что снимает ряд технических сложностей.

Группировка фотоотсчетов, приводящая к распределению Ферми–Дирака: одна система. Первым рассмотрим способ группировки фотоотсчетов, который приводит к распределению Ферми–Дирака (Ф–Д) [6]. Пусть вся последовательность разбита на блоки длиной n, т.е. содержит n тактов (ящиков) – аналог уровней энергии (см. рис. 1a, b). Пусть произошло $k \le n$ фотоотсчетов (фотоотсчет – аналог частицы). Все последовательности длины n с k фотоотсчетами имеют одинаковую вероятность и относятся к одному классу. Число последовательностей в классе (статистический вес) равно числу способов размещения k частиц по n уровням (ящикам) так, чтобы на каждом уровне было не более одной частицы (не более одного фотоотсчета в такте). Удобно воспользоваться известной формулой для биномиальных коэффициентов

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

Число последовательностей в классе есть C_n^k, и каждая из них имеет вероятность $P_n(k) = (1-p)^k p^{n-k}$. Оценим предельное число случайных бит, которые могут быть извлечены при таком способе группировки. Количество бит (пусть не целое, но максимальное), которое можно получить из каждой последовательности из класса k есть $\log(C_n^k)$, соответственно из всех последовательностей в классе $C_n^k \log(C_n^k)$. С учетом того, что вероятность появления каждой последовательности в классе $P_n(k) = (1-p)^k p^{n-k}$, среднее число случайных бит из всех классов

$$H_n^{\mathrm{F-D}}(p) = \sum_{k=0}^{n} P_n(k) C_n^k \log(C_n^k). \tag{2}$$

Покажем, что в асимптотическом пределе ($n \to \infty$) $H_n^{\mathrm{F-D}}(p)$ стремится к предельному значению $nh(p)$, т.е. метод группировки по Ф-Д асимптотически точный и позволяет получить максимальное число случайных бит, содержащихся в физическом процессе. Первое, что нужно показать, что данный способ группировки в асимптотическом пределе позволяет извлечь полное число случайных бит в пересчете на такт, которое дается энтропией Шеннона[2)]

$$H_n^{\mathrm{Sh}}(p) = -\sum_{k=0}^{n} C_n^k P_n(k) \log(P_n(k)), \tag{3}$$

где $[-\log(P_n(k))]$ – количество истинно случайных бит, содержащихся в последовательности длины n с k отсчетами. Асимптотический предел по числу случайных бит на такт есть $H_n^{\mathrm{Sh}}(p)/n = h(p)$, и не зависит от длины последовательности. Хотя данный факт интуитивно очевиден, тем не менее покажем это. Учитываем, что

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

$$= -\sum_{k=0}^{n} C_n^k (1-p)^k p^{n-k} \left(k\log(1-p) + (n-k)\log(p)\right) =$$

[2)] Отметим, что для случая независимых отсчетов распределение вероятностей распадается на произведение $P(x_1, x_2, \dots x_n) = P(x_1)P(x_2)\dots P(x_n)$. В этом случае $H_{\min}$ энтропия, имеющая смысл числа случайных бит в последовательности длины n, совпадает с энтропией Шеннона, т.к. $\max_{P(x_1,x_2,\dots x_n)} P(x_1, x_2, \dots x_n)$ в нашем случае достигается на типичных последовательностях, которые имеют вероятность $(1-p)^{n(1-p)}p^{np}$, в итоге $H_{\min} = nh(p)$.

$$= -\sum_{k=0}^{n} C_n^k \left(p^{n-k} \log(1-p) \frac{\partial}{\partial(1-p)} (1-p)^k + \right.$$

$$\left. + (1-p)^k \log(p) \frac{\partial}{\partial p} p^{n-k} \right) = nh(p). \qquad (4)$$

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

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

$$\log(N_n(k)) \le -\log(P_n(k)), \qquad (5)$$

откуда следует, что

$$H_n^{\text{F–D}}(p) = \sum_{k=0}^{n} P_n(k) N_n(k) \log(N_n(k)) \le$$

$$\le H_n^{\text{Sh}}(p) = -\sum_{k=0}^{n} P_n(k) N_n(k) \log(P_n(k)). \qquad (6)$$

При $n \to \infty$ число фотоотсчетов $*$ в последовательности длины n равно $k \approx (1-p)n$, соответственно число пропусков $\sqcup$ $n-k \approx np$. При этом $\log(N_n(k)) = \log(C_n^k) = \log(\frac{n!}{(n-k)!k!})$. С учетом формулы Стирлинга $n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$, находим в главном приближении $\log(C_n^K) \approx n \log(\frac{1}{p(1-p)}) = nh(p)$, подставляем в (2), и учитывая (4), получаем $H_n^{\text{F–D}}(p) \to H_n^{\text{Sh}}(p) = nh(p)$.

Группировка фотоотсчетов, приводящая к распределению Ферми–Дирака: несколько систем. Каждый интервал из M тактов с k_i отсчетами $(*)$ соответствует одной из L разных систем, $i = 1, 2 \dots L$ (см. рис. 1b)). Разные расположения отсчетов внутри интервала из M тактов отвечают размещениям полного числа частиц $K = \sum_{i=1}^{L} k_i$ по разным L системам. Данная группировка отвечает группировке неразличимых фермионов по нескольким различным системам. Для сравнения с предыдущим способом надо привести ситуацию к общему числу тактов, т.е. положить $n = LM$. Аналогично предыдущему, имеем

$$\sum_{K=0}^{LM} C_{LM}^K (1-p)^K p^{LM} =$$

$$= (1-p+p)^{LM} = ((1-p+p)^M \dots (1-p+p)^M) =$$

$$= \sum_{k_1=0}^{M} \sum_{k_2=0}^{M} \dots \sum_{k_L=0}^{M} C_M^{k_1} C_M^{k_2} \dots C_M^{k_L} (1-p)^{k_1} \times$$

$$\times p^{M-k_1} (1-p)^{k_2} p^{M-k_2} \dots (1-p)^{k_L} p^{M-k_L}. \qquad (7)$$

Выделим в левой и правой частях (7) слагаемые, которые имеют одинаковую вероятность $(1-p)^K p^{LM-K}$,

$$C_{LM}^K = \sum_{k_1=0}^{M} \sum_{k_2=0}^{K-k_1} \dots \sum_{k_L=0}^{K-k_1-k_2-\dots-k_{L-1}} C_M^{k_1} C_M^{k_2} \dots C_M^{k_L},$$

$$K = \sum_{i=1}^{L} k_i, \quad 0 \le K \le LM = n. \qquad (8)$$

При группировке по распределению Ф–Д имеется один класс с числом C_{LM}^K последовательностей, а при группировке с несколькими типами имеется несколько классов, равных числу слагаемых в сумме (8). Причем все последовательности имеют одинаковую вероятность $(1-p)^K p^{LM-K}$. Покажем теперь, что при таком разбиении число случайных бит в пересчете на один такт меньше, чем при группировке по Ф–Д в предыдущем способе. Для краткости будем нумеровать число классов индексом i. Пусть число классов есть n_{class}, число последовательностей в классе – $N_i(K)$. В каждом i-м классе имеется $N_i(K) = C_M^{k_1} C_M^{k_2} \dots C_M^{k_L}$ последовательностей (с такими k_j, что $K = \sum_{i=j}^{L} k_j$). Далее

$$\sum_{i=1}^{n_{\text{class}}} N_i(K) = C_{LM}^K = N(K). \qquad (9)$$

Покажем теперь, что можно извлечь меньшее число случайных бит. Пусть $q_i(K) = \frac{N_i(K)}{N(K)}$, и $\sum_{i=1}^{n_{\text{class}}} q_i(K) = 1$ имеет смысл вероятности. Из того, что $q_i(K) \log(q_i(K)) \le 0$, следует

$$H_n^{\text{F–D}_{\text{multi}}}(p) - H_n^{\text{F–D}}(p) = \sum_{K=0}^{LM} P(K) N(K) \times$$

$$\times \left(\sum_{i=1}^{n_{\text{class}}} \frac{N_i(K)}{N(K)} \left(\log(N_i(K)) - \log(N(K)) \right) \right) =$$

$$= \sum_{K=0}^{LM} P(K) N(K) \sum_{i=1}^{n_{\text{class}}} q_i(K) \log(q_i(K)) \le 0. \qquad (10)$$

Из (10) следует, что $H_n^{\text{F–D}_{\text{multi}}}(p) \le H_n^{\text{F–D}}(p)$, т.е. группировка отсчетов (K частиц) по L различным системам менее эффективна, чем группировка K частиц по статистике Ф–Д в одной системе, что является следствием выпуклости шенноновской энтропии.

В асимптотическом пределе $M \to \infty$ $H_n^{\text{F–D}_{\text{multi}}}(p)$ выходит на шенноновский предел. С учетом того,

что $k_i \approx M(1-p)$ и $M-k_i \approx Mp$, $C_M^{k_i} \approx 2^{Mh(p)}$, $H_n^{\mathrm{F-D_{multi}}}(p) \to \prod_{i=1}^{L} C_M^{k_i}(1-p)^{LM(1-p)}p^{LM}\log((1-p)^{LM(1-p)}p^{LM}) \to LMh(p) = nh(p)$.

Группировка фотоотсчетов, приводящая к распределению Бозе–Эйнштейна. Возможна следующая группировка фотоотсчетов, которая приводит к распределению Бозе–Эйнштейна. В этой статистике роль частиц играют пропуски ($\sqcup$ – отсутствие отсчета во временно́м окне T). Роль перегородок между ящиками (уровнями энергии) – фотоотсчеты $*$. Заранее фиксируется число R – отсчетов $*$. Начало отсчета начинается с конца предыдущего отсчета $*$ (см. рис. 1c, d). Пусть $W = R+1$ – число перегородок – $*$ с учетом последнего отсчета, соответственно число ящиков (уровней) $W-1 = R$ (рис. 1c, d). Как только возникает R-отсчет, процесс прерывается и начинается отсчет следующего процесса. Пусть полное число частиц ($\sqcup$) между отсчетами $*$ (перегородками) по всем ящикам есть D. При этом полное число тактов есть $D+W-1 = D+R$.

Все последовательности с одинаковым числом частиц D и числом ящиков $W-1 = R$ имеют одинаковую вероятность $(1-p)^{W-1}p^D$. Отрицательное биномиальное распределение (распределение Паскаля, оно же распределение Бозе–Эйнштейна) возникает при подсчете числа комбинаций до появления k успеха, при этом k задано, а число попыток n – переменное. Такое же распределение возникает при подсчете числа способов размещения D тождественных частиц ($\sqcup$) по $W-1 = R$ уровням (ящикам). Имеем

$$C_{R+D-1}^{D} = \frac{(R+D-1)!}{D!(R-1)!)} = C_n^k, \qquad (1-p)^{W-1}p^D = (1-p)^k p^{n-k}. \tag{11}$$

Для подсчета вероятностей отдельных событий уточним множество элементарных событий Ω – бесконечный набор последовательностей, которые входят в отдельные классы с одинаковой вероятностью. Классы нумеруются двумя индексами – $k-1$ число отсчетов $*$ (фиксировано), $n-k$ – полное число пустых тактов $\sqcup$ (частиц), $\Omega = \{(k, n=0), (k, n=1), \ldots (k, n\to\infty)\}$. Для выяснения вероятности каждого элементарного события воспользуемся условием нормировки полной вероятности. Пусть вероятность элементарного события есть $P_k(n)$. Число элементарных событий, имеющих одинаковую вероятность в каждом классе, равно C_n^k. Условие нормировки

$$\sum_{n\geq k}^{\infty} C_n^k P_k(n) = 1. \tag{12}$$

Производящая функция для распределения Бозе–Эйнштейна имеет вид

$$\frac{p^k}{(1-p)^{k+1}} = \sum_{n\geq k}^{\infty} C_n^k p^n, \tag{13}$$

откуда следует, что вероятность последовательности из класса (k,n) есть $P_k(n) = (1-p)^{k+1}p^{n-k}$.

Найдем предельное значение энтропии при группировке фотоотсчетов, приводящей к распределению Бозе–Эйнштейна. В асимптотическом пределе $(k\to\infty)$

$$H_k^{\mathrm{B-E}}(p) = -\sum_{n\geq k}^{\infty} C_n^k P_k(n)\log(P_k(n)). \tag{14}$$

Учитывая, что $C_n^{k+1} = \frac{n-k}{k+1}C_n^k$, и обозначая $k' = k+1$, можно переписать (14):

$$H_k^{\mathrm{B-E}}(p) = \tag{15}$$

$$= -\sum_{n\geq k'}^{\infty} \frac{k'}{n-k'+1} C_n^{k'}(1-p)^{k'}p^{n-k'}\log((1-p)^{k'}p^{n-k'}).$$

Сравнивая (15) с (4), и учитывая, что главный вклад дают типичные последовательности, т.е. такие, что $k' \approx (1-p)n$ и $n-k' \approx pn$, и при этом $C_n^{k'} \approx 2^{nh(p)}$, находим

$$H_k^{\mathrm{B-E}}(p) \to \frac{1-p}{p} nh(p), \tag{16}$$

что несколько меньше (так как p близко к 1 в реальной ситуации), чем шенноновский предел. Формула (16) носит асимптотический характер (np должно быть велико), поэтому предельный переход $p\to 0$ надо делать на более раннем этапе в формуле (14). Кроме того, данная группировка неудобна при технической реализации, потому что ожидание последнего k-го отсчета $(*)$ (см. рис. 1d) может потребовать большого числа тактов (хотя такие события могут быть и редкими), что технически неприемлемо. В предыдущем случае, когда число тактов n в каждом блоке заранее задано, а число отсчетов не фиксируется, их число такое, какое реально произошло.

Извлечение случайной последовательности 0 и 1 из последовательности фотоотсчетов $*$ и $\sqcup$. Необходимо эффективно по последовательности фотоосчетов $i_1, i_2, \ldots i_n$ ($i_j = *$ или $\sqcup$) длиной n построить выходной блок случайной последовательности 0 и 1. В нашей работе [7] при экспериментальной реализации КГСЧ использован по существу способ группировки по Ферми–Дираку с несколькими подсистемами. Получение случайной последовательности по фотоотсчетам происходило при помощи таблицы, содержащей последовательности фото-

отсчетов как адрес блока выходной случайной последовательности. Скорость генерации случайных чисел при таком способе технически ограничена размером таблицы, который растет экспоненциально по длине последовательности.

Возможен способ нумерации последовательностей фотоотсчетов, который требует лишь полиномиальных ресурсов по длине последовательности и позволяет обрабатывать практически последовательности любой длины. Для этого можно воспользоваться методами и результатами из теории арифметического кодирования. Это происходит в два этапа. На первом этапе выходной последовательности $i_1, i_2, \dots i_n$ сопоставляется номер, который определяется полиномиальным алгоритмом [8]. Пусть в последовательности имеется K отсчетов $*$. Обозначим номер позиции индексом j_m. Тогда имеется взаимно однозначное соответствие между последовательностью фотоотсчетов $(i_1, i_2, \dots i_n)$ и ее номером $\mathrm{Num}(i_1, i_2, \dots i_n)$ $(0 \le \mathrm{Num}(i_1, i_2, \dots i_n) \le C_n^K - 1)$

$$\mathrm{Num}(i_1, i_2, \dots i_n) = C^1_{j_1-1} + C^2_{j_2-1} + \dots + C^k_{j_k-1}, \quad (17)$$

$$C^l_j = 0, \quad j < l.$$

Биномиальные коэффициенты могут быть вычислены заранее и помещены в таблицу размером $n \times n$. Биномиальный коэффициент выбирается на ходу по мере появления фотоотсчетов $(*)$. При появлении первого отсчета в позиции j_1 выбирается число (биномиальный коэффициент) на пересечении первой строки j_1 и первого столбца матрицы. При появлении второго отсчета – берется коэффициент в матрице на пересечении j_2 строки и второго столбца, и т.д. В итоге получается номер последовательности $\mathrm{Num}(i_1, i_2, \dots i_n)$.

Пусть последовательность принадлежит к некоторому классу, полное число последовательностей в классе – N_k. Следующий шаг – по номеру последовательности выдается блок случайных 0 и 1. Номера последовательностей находятся в диапазоне $0 \le \le \mathrm{Num} \le N_k - 1$. Далее, берем номер текущей последовательности Num. Пусть бинарное представление числа последовательностей в классе $N_k = \sum_{i=0}^{i_{\max}} 2^{k_i}$. Процедура выполняется рекурсивно. Если номер текущей последовательности Num находится в интервале $2^{k_0} + 2^{k_1} + \dots + 2^{k_{i-1}} \le \mathrm{Num} \le 2^{k_0} + 2^{k_1} + \dots + 2^{k_{i-1}} + 2^{k_i} - 1$ $(i \le i_{\max})$, тогда выходной случайной последовательностью будет k_i младших разрядов бинарного представления Num. Число номеров последовательностей в этом диапазоне равно 2^{k_i}.

Например, $N_k = 6 = (110)_2$, пусть номер последовательности $\mathrm{Num} = 3 = (011)_2$, тогда на выход выдается (11). При $\mathrm{Num} = 5 = (101)_2$ на выходе будет (01). На рис. 1e приведены зависимости числа случайных бит на такт Rnd как функции длины блока n для различных значений вероятности p. Как видно, выход на теоретическое значение достигается уже при длинах блока $n = 128$ независимо от вероятности p.

Заключение. Таким образом, группировка по распределению Ферми–Дирака является самой эффективной. Выход на асимптотический предел при различных вероятностях $P(*)$ фотоотсчетов происходит с хорошей точностью при длине блока $n \approx 128$. Обработка последовательностей такой длины требует таблицы размером 128×128, что легко достижимо, в отличие от прямой адресации, где размер таблицы экспоненциально большой $\approx 2^{128}$, поэтому практически недостижим. Выбор вероятности $P(*)$ – темпа фотоотсчетов зависит от свойств лавинного детектора (времени рассасывания заряда после регистрации, величины afterpulsing) и выбирается таким образом, чтобы обеспечить статистическую независимость отсчетов. Следовательно, при надлежащем выборе лавинного детектора можно выйти на предельную скорость генерации случайных чисел, близкую к тактовой частоте. Экспериментальная реализация будет представлена в отдельном сообщении.

Выражаю благодарность К.А. Балыгину, А.Н. Климову, С.П. Кулику, К.С. Кравцову, И.В. Радченко за обсуждения, а также коллегам из Академии Криптографии Российской Федерации за обсуждения и поддержку.

1. M. Herrero-Collantes, J. Carlos Garcia-Escartin, Rev. Mod. Phys., accepted 17 October (2016); Q. Yan, B. Zhao, Z. Hua, Q. Liao, and H. Yang, Rev. Sci. Instrum. **86**, 073113 (2015); F.-X. Wang, C. Wang, W. Chen, S. Wang, F.-S. Lv, D.-Y. He, Z.-Q. Yin, H.-W. Li, G.-C. Guo, and Z.-F. Han, J. Light Wave Techn. **33**, 3319 (2015).
2. R. König, R. Renner, and C. Schaffner IEEE Transactions on Information Theory **55**, 4337 (2009).
3. Д. Н. Клышко, *Фотоны и нелинейная оптика*, Наука, М. (1980).
4. J. von Neumann, Appl. Mathem. Ser. **12**, 36 (1951).
5. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley (1991).
6. Л. Д. Ландау, Е. М. Лифшиц, *Статистическая физика*, т. V, Наука, М. (1995).
7. K. S. Kravtsov, I. V. Radchenko, S. P. Kulik, and S. N. Molotkov, J. Optical Soc. Amer. **32**, 1743 (2015).
8. В. Ф. Бабкин, Пробл. Перед. Информ. **VII**, 13 (1971).