Randomness and device independence*

Jędrzej Kaniewski†

Abstract

In these lectures we explore the notion of randomness in the context of device-independent information processing. In the first part we discuss the concept of randomness and various ways of formalising it. In the second part we explain why quantum mechanics is a convenient framework for generating randomness and discuss specific protocols that achieve this goal.

Contents

*Lecture notes for the KIAS Summer School in Quantum Information Science in Seoul, Korea on 16–20 July 2018.

†Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland

1 The concept of randomness

While we all have decent intuition for the concept of randomness, formalising it is not so straightforward and generally speaking there are two distinct approaches. The first one, which we call information-theoretic randomness, is defined on the grounds of probability theory and allows us to think of randomness as a resource which can be stored, extracted or transferred. This theory is intuitive and convenient to work with, but it requires us to have a reliable probabilistic model of our system. The other approach to randomness, which we call algorithmic randomness, is based on complexity theory. This approach is also quite intuitive (at least at first glance) and it does not require a probabilistic. We start by explaining the two alternative approaches to randomness. At the end we briefly discuss how these are related to typical applications.

1.1 Information-theoretic randomness

In everyday speech the word "random" is often used to mean "unpredictable". Typical examples of random events include traffic jams, flight cancellations or price fluctuations. However, the concept of unpredictability suffers from an inherent problem: it is strongly observer-dependent. The ending of a book might be quite unpredictable to the reader, but is well-known to the author. This is often expressed through the infamous question: "random for whom?". The natural way of formalising these concepts uses the standard framework of probability theory and this is precisely the information-theoretic approach to randomness.

We start with a probability distribution over some (finite) set $\mathcal{X} = \{1, 2, \ldots, d\}$, i.e. a sequence $p(x)$ of numbers satisfying

$$\begin{aligned} \forall x \in \mathcal{X} \quad p(x) \geq 0 \quad &\text{(non-negativity)}, \\ \sum_{x \in \mathcal{X}} p(x) = 1 \quad &\text{(normalisation)}. \end{aligned}$$

Clearly, the least random probability distribution is the deterministic distribution, e.g.

$$p(x) = \begin{cases} 1 & \text{for} \quad x = 1, \\ 0 & \text{otherwise.} \end{cases}$$

The most random is the uniform distribution, i.e.

$$p(x) = \frac{1}{d}.$$

Given a probability distribution one can propose many measures of randomness. The most important measure is called the *Shannon entropy* and is defined as

$$\mathrm{H}(X) := -\sum_x p(x) \log p(x),$$

where we adopt the convention that "$0 \log 0 = 0$" (because $x \log x \to 0$ as $x \to 0$) and the logarithm is base 2. The Shannon entropy measures the amount of randomness of the probability distribution in the asymptotic limit (we will later sketch an argument why that is the case).

The min-entropy, on the other hand, has a more natural definition as it is related to the *guessing probability*

$$\mathrm{H}_{\min}(X) := -\log \mathrm{p}_{\mathrm{guess}}(X)$$

for

$$\mathrm{p}_{\mathrm{guess}}(X) := \max_x p(x).$$

It is not hard to prove that $\mathrm{H}_{\min}(X) \leq \mathrm{H}(X)$ (hence the name) and often this difference can be significant, e.g. consider the probability distribution defined as

$$p(x) = \begin{cases} \frac{1}{2} & \text{for} \quad x = 1, \\ \frac{1}{2(d-1)} & \text{for} \quad x \geq 2. \end{cases}$$

It is easy to check that $\mathrm{H}_{\min}(X) = 1$, while

$$\mathrm{H}(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{d-1}{2(d-1)}\log\frac{1}{2(d-1)} = 1 + \frac{1}{2}\log(d-1).$$

For large d the two quantities are substantially different, but we should not think of this as a problem since they describe distinct aspects of the randomness present in X. The security guarantees in cryptography are usually stated in terms of the min-entropy. This simply shows that there is not a single universal measure of randomness.

So far we have only considered a single random variable X and the resulting entropies quantify unpredictability from the point of view of an observer who knows the distribution of X, but nothing else. To include the possibility of additional side information we must consider two random variables X and Y drawn from the joint distribution $p(xy)$. This gives rise to *conditional entropies* and the conditional Shannon entropy is defined as the difference between the *joint entropy* and the entropy of the additional system, i.e.

$$H(X|Y) := H(XY) - H(Y).$$

The conditional min-entropy is defined in terms of the conditional probabilities

$$p(x|y) := \frac{p(xy)}{p(y)},$$

where $p(y) := \sum_x p(xy)$ and we assume that $p(y) > 0$. The conditional min-entropy is defined as

$$\mathrm{H}_{\min}(X|Y) := -\log \mathrm{p}_{\mathrm{guess}}(X|Y),$$

where the conditional guessing probability is given by

$$\mathrm{p}_{\mathrm{guess}}(X|Y) := \sum_y p(y) \max_x p(x|y).$$

It is easy to see that the last definition captures precisely the concept of guessing with side information.

Since having access to additional side information can be very beneficial, the quantities $\mathrm{H}(X)$ and $\mathrm{H}(X|Y)$ can be vastly different. Consider the maximally correlated distribution, i.e.

$$p(xy) = \frac{1}{d}\delta_{xy},$$

where

$$\delta_{xy} := \begin{cases} 1 & \text{if} \quad x = y, \\ 0 & \text{otherwise.} \end{cases}$$

Since $p(x) = \frac{1}{d}$, we have $\mathrm{H}(X) = \log d$, but $\mathrm{H}(X|Y) = 0$ (and the same for the min-entropy). The distribution of X itself is random, but the knowledge of Y allows us to predict it with perfect accuracy. This clearly shows that the amount of randomness depends on the observer. Alternatively, one can think that randomness is simply a consequence of ignorance.

1.2 Randomness extraction

We have seen how to quantify the amount of randomness in a random variable. For most applications what we want is strong (close to uniform) randomness. On the other hand, what is often available are weak sources of randomness which contain entropy but they are not close to the uniform distribution. Assume that we have access to a source of weak randomness and we can use it several times (and the different uses are independent from each other). Can we use it to generate strong randomness? In other words, is randomness a *resource* that can be "concentrated"?. The answer is "yes" and such protocols go under the name of *randomness extraction*.

1.2.1 von Neumann extractor

Let us start with the simplest procedure known as the von Neumann extractor. Consider a source producing random bits such that

$$\begin{aligned} p(0) &= q, \\ p(1) &= 1-q \end{aligned}$$

for some $q \in (1/2, 1)$. If we use the source twice, it is easy to check that

$$\begin{aligned} p(00) &= q^2, \\ p(01) &= q(1-q), \\ p(10) &= q(1-q), \\ p(11) &= (1-q)^2. \end{aligned}$$

The proposal of von Neumann is based on the observation that $p(01) = p(10)$ regardless of the value of q. If we observe "00" or "11" the protocol aborts, whereas the other two possible output strings are interpreted as

$$\begin{aligned} 01 &\to 0, \\ 10 &\to 1. \end{aligned}$$

This extractor succeeds with probability $2q(1-q)$, but when it succeeds, it produces one bit of perfect randomness regardless of the quality of the source. With some non-zero probability we have turned weak randomness into strong randomness.

1.2.2 Asymptotic extraction

In the second example we consider the asymptotic limit, i.e. the source is used a large number of times and we are interested in the amount of randomness that can be extracted per use.

Consider a random variable X described by the probability distribution $p(x)$. Suppose we have n independent copies of this random variable, which we denote by $X_1 X_2 \dots X_n$. It is easy to see that the expected number of appearances of the symbol $x \in \mathcal{X}$ equals $np(x)$. Of course, in general this is not an integer, but for large n it will be close enough to some integer. The probability of observing a string $x^* = x_1 x_2 \dots x_n$ in which symbol x appears approximately $np(x)$ times equals approximately

$$p(x^*) \approx \prod_x \big[p(x)\big]^{np(x)} = \prod_x 2^{np(x)\log p(x)} = 2^{n\sum_x p(x)\log p(x)} = 2^{-nH(x)}.$$

In the limit of large n the probability of producing a string whose composition significantly differs from the one described above is small, i.e. with high probability we will end up with one of the special strings. Since all these strings have roughly the same probability, we end up with $2^{nH(x)}$ almost equally likely strings, i.e. $nH(X)$ bits of nearly perfect randomness. This is precisely why Shannon's entropy quantifies the amount of randomness produced by the source per one use in the asymptotic limit.

1.3 Algorithmic randomness

Information-theoretic randomness is a powerful framework in which many natural measures of randomness can be defined and related to operational tasks. However, it crucially relies on the existence of some underlying probability distribution. In other words, we need a reliable probabilistic description of our problem.

Can we make any statements in the absence of such a probabilistic description? What if we are just given a string and we want to assess whether it is random or not?

Such considerations lead to the notion of *algorithmic* or *Kolmogorov complexity* and these concepts are conveniently discussed using the example of n-bit strings, i.e. elements of the set $x \in \{0,1\}^n$. The intuition is simple: the strings "0000000000" and "1111111111" do not look random, but the string "0011010010" does. Note that to make such claims we do not need any underlying distribution, we simply look at a single string $x \in \{0,1\}^n$. In the algorithmic formulation a string is considered random if it does not have a simple description. For instance strings "0000000000" and "0101010101" have simple descriptions ("output all zeroes" and "alternate between 0 and 1"), whereas the string "0011010010" is more complicated to describe. This might seem like an attractive formulation, but it is not particularly convenient to work with: the Kolmogorov complexity of a string, defined as the length of the shortest program on a universal Turing machine which produces that string, is a deeply uncomputable quantity. Therefore, it is hard to make any computations in this framework.

The idea that random objects should not admit simple description is precisely what stands behind the various statistical tests used to assess the quality of random number generators. If we see a string that is particularly simple, we conclude that the random number generator does not function correctly. If we are given a device that is supposed to generate an n-bit string chosen uniformly at random, how do we verify it? Well, we use the machine to generate an n-bit string x and analyse it. We expect to see approximately equal number of zeroes and ones, so our first check is to simply count the number of symbols. If these numbers deviate significantly from $\frac{n}{2}$, e.g. we see strings like "0000000000" or "1111111111", we should become suspicious.[1] In the second check we should compare the number of "changes" ($x_{j+1} \neq x_j$) with the number of no-changes ($x_{j+1} = x_j$). Again, one would expect that the these numbers should be roughly the same, so if we see a string like "0000011111", then, again, something is not quite right. These two checks are useful, but they are clearly not enough: the string "001100110011" satisfies both of them, but does not look particularly random.

Let us conclude with a simple example that demonstrates the difference between information-theoretic and algorithmic randomness sometimes referred to as the *memory-stick attack*. Suppose we are given access to a source of perfectly random bits. We generate an n-bit string x and whatever statistical tests we run on it, we will find that our string passes them. Now, suppose we build a device which deterministically outputs string x and advertise it as a perfect random number generator. This device will pass all the statistical tests, but we know that the output is not random at all. In particular, a copy of our "random" number can be stored on somebody else's hard drive. This shows that the two way of formalising randomness are essentially incomparable.

For cryptographic purposes it is the information-theoretic approach which is relevant and we will focus on that for the rest of these notes. However, one should remember that there exist other approaches to randomness which are useful for other applications.

1.4 Why do we need randomness?

With the advent of digital information generating randomness has become an important task which is performed regularly by all the digital devices you own (including the mp3 player: think about the shuffle option). Let us just quickly go through 3 typical applications.

[1] Clearly, this is in stark contradiction to the previous model in which every string is equally likely and, therefore, seeing "0000000000" should not be any more surprising than "0011010010".

- Cryptography: generating cryptographic keys or running online casinos.
- Algorithms: often probabilistic algorithms are more efficient than deterministic ones. Suppose you are given two functions $f, g : \{1, 2, \ldots, d\} \to \{1, 2, \ldots, d\}$ and you want to check whether $f = g$. A deterministic algorithm must look at all the points. A probabilistic algorithm can choose a random subset and if the functions agree, one can with high probability conclude that the functions are the same. The point is that for every deterministic algorithm there is a strategy that hides the point at which $f \neq g$ until the end of the search. Probabilistic algorithms are harder to fool.
- Numerical simulations: choosing a random starting point or making random decisions in the process (e.g. Monte Carlo algorithms).

In the first case we require that our randomness is unpredictable from the point of view of the adversary (e.g. a hacker trying to guess our cryptographic key). In the other cases we require randomness, but this could be publicly known randomness, e.g. we could use the NIST beacon available at:

https://beacon.nist.gov/home.

Note the explicit warning against using this beacon for cryptographic applications.

2 Randomness in physics

In this section we discuss the role of randomness in physics. We start by describing classical physics which has no inherent randomness. We then proceed to quantum physics which is a non-deterministic theory and, therefore, is useful for generating randomness. Finally, we explain the device-independent approach to randomness in quantum physics.

2.1 Classical physics

Prequantum theories like Newtonian mechanics or electromagnetism are completely deterministic. Given the rules that govern the evolution of the system and the initial conditions, we can at least in principle simulate its evolution until arbitrary large time (although the computational cost of such a simulation might quickly become prohibitive). Therefore, a classical machine cannot generate any "new" randomness: the best it can do is to collect data which are believed to be random and then scramble them together to produce an output which is even more random. Such algorithms are known as *pseudorandom number generators* and from the mathematical point of view they are simply (deterministic) functions which take an input string known as the *seed* and generate another string. The output string can be significantly longer than the seed, but if we think of the input and output strings as random variables we immediately see that the entropy of the output cannot be higher than the entropy of the input. In other words, these algorithms do not generate any new randomness, they just shuffle the existing randomness around. This can still be useful as shown by the following example. Suppose I have access to two pseudorandom number generators produced by different companies and I use them to generate two random n-bit strings x_1 and x_2. It could be the case that x_1 is perfectly known to the first company and x_2 to the second company, but if I compute the XOR of the two strings $x_1 \oplus x_2$, the companies would need to talk to each other to figure out the resulting string. This is a well-known trick in classical randomness generation: collect random data from multiple source (ideally you should have a good reason to believe that these sources are uncorrelated, e.g. devices produced by different companies or physical measurements on systems far apart) and then scramble it all together. In a personal computer these sources can range from the

system clock and signals from your keyboard to the external noise recorded by the microphone. While these solutions seem to work well, in the sense that we do not experience major security breakdowns on a regular basis, assessing the level of security in a quantitative fashion is quite challenging. In fact, a study in 2016 revealed that poor quality randomness affects a non-negligible fraction of RSA keys used online [BSTPS16].

An obvious way to eliminate full determinism is to artificially inject a little bit of randomness into our otherwise fully deterministic theory. The first attempt would be to blame everything on our ignorance: since we do not have complete knowledge about the system, its behaviour appears random. This, however, only solves the problem for us and if a more knowledgeable observer comes along, he will see perfectly deterministic behaviour. A more compelling argument in this direction has been recently put forward by Nicolas Gisin [Gis18]. In classical physics the position of a particle moving on a line is described by a real number. However, since almost all real numbers are extremely complicated to describe (in the sense that they do not have a short description like $1/2$ or $\sqrt{2}$; there are just many more real numbers than short descriptions), Gisin argues that we can never hope to learn the exact position of the particle: we simply do not have the capacity to process that amount of information. Therefore, he postulates that from our point of view the position is only specified up to certain accuracy, say 100 decimal places, whereas everything further down is beyond our reach. Imposing such a fundamental limitation will indeed render our theory somewhat non-deterministic and this might be particularly important for chaotic systems, in which a small difference in the initial conditions leads to vastly different trajectories.

2.2 Quantum physics

In quantum physics, on the other hand, non-determinism is inherent in the theory: when a measurement is performed, the Born rule does not tell us which outcome we will observe, it only gives us the probabilities. This feature might seem quite disturbing and this is precisely the reason why Einstein disliked quantum mechanics as expressed in his famous quote "God does not play dice". A solution to this "problem" motivated by classical physics is to attribute all non-determinism to our ignorance. In the famous paper by Einstein, Podolsky and Rosen (EPR) this reasoning is used to conclude that the "quantum-mechanical description" should not be considered complete [EPR35].

The presence of inherent randomness in quantum mechanics, which seriously troubled some great minds of the 20th century, is precisely what makes it the dream theory for cryptographers. Suddenly, we have a theory that allows us to generate randomness from scratch and all within a rigorous mathematical framework. What else could you possible ask for?

2.3 Quantum random number generator

The simplest quantum random generator consists of a source of photons, a beamsplitter and two detectors. The source emits the state $|\psi_0\rangle = |0\rangle$, the action of the beamsplitter is represented by the Hadamard matrix

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

where the kets $|0\rangle$ and $|1\rangle$ correspond to the two different output ports of the beamsplitter. Clearly, the measurement operators corresponding to the two detectors are given by $M_0 = |0\rangle\langle 0|$ and $M_1 = |1\rangle\langle 1|$. It is clear that the probability of either detector clicking equals $\frac{1}{2}$.

Of course, in a real experiment we only have an approximate description of our devices. For instance, we believe that the actual state emitted by the device, which we denote by ρ, is *close* to the ideal state $\psi_0 := |\psi_0\rangle\langle\psi_0|$ as quantified by the trace distance

$$\|\rho - \psi_0\|_1 \leq \varepsilon.$$

Then, we can still show that probability of each detector clicking is close to $\frac{1}{2}$. More specifically, we have

$$\frac{1-\varepsilon}{2} \leq \Pr[\text{detector 0 clicks}] \leq \frac{1+\varepsilon}{2},$$
$$\frac{1-\varepsilon}{2} \leq \Pr[\text{detector 1 clicks}] \leq \frac{1+\varepsilon}{2}.$$

As long as $\varepsilon < 1$ we get some randomness and recall that we already know how to turn weak randomness into strong randomness.

A photon hitting a beamsplitter is the paradigmatic example of a quantum process that can be used to generate randomness [SGG+00], but obviously there are many other processes that can be used for the same purpose, e.g. spontaneous emission [MSL+15]. Quantum random number generators are commercially available, but they do not seem particularly popular. This is usually blamed on their low randomness generation rate and unreliability. These are, of course, typical problems of every new technology and progress is continuously being made to eliminate them, e.g. see a recent report from Toshiba [MPL+18]. However, for a new technology to be successful it is not sufficient to simply match the specifications of the existing one: it must provide an advantage significant enough to outweigh the cost of transitioning. Although the classical ways of generating randomness are not perfect, they do not cause any major problems (at least no such information is available to the public). Therefore, it might be hard for quantum solutions to challenge the existing technology.

The existence of quantum processes that generate true randomness should not be surprising and the only necessary assumption is that we must know what our quantum devices are doing. What is surprising is that even that assumption can be dropped. This leads to the concept of *device-independence* which from now on will become our main focus. However, to get a grip on device-independence, we must first discuss the basics of Bell nonlocality.

2.4 Bell nonlocality

In classical physics every measurable quantity can be assigned a particular value in an objective manner and if this value is not known to us, we can only blame it on our ignorance. In the seminal paper of Einstein, Podolsky and Rosen these objective values are referred to as the "elements of reality". The crucial observation is that in quantum mechanics these elements of reality cannot be assigned in a consistent way unless we are willing to give up the notion of locality, i.e. we allow for superluminal signalling. This led the authors to believe that the quantum-mechanical description cannot be considered not complete.

This argument was picked up almost 30 years later by John Bell who formalised the notion of elements of reality consistent with locality [Bel64]. He found that if a theory admits a local-realistic description, then the strength of correlations in that theory is limited. Moreover, he showed that quantum mechanics does not obey these restrictions, which implies that there does not exist a local-realistic description of quantum mechanics.

In the simplest Bell scenario we consider two correlated devices (see Fig. 1). Each device is capable of performing at least two different measurements and we denote the measurement settings by x and y. Then, each device produces an outcome denoted by a and b. The devices might be correlated, but they do not communicate during the experiment. Assuming that we can repeat the experiment multiple times and that every time the devices behave in the same manner (the so-called i.i.d. assumption, see Sec. 2.7), we can estimate the conditional probability distribution

$$P(a, b|x, y).$$

Assuming that the devices are governed by a particular theory, we may ask which probability distribution can be observed within that theory. For example if the two devices are governed by quantum

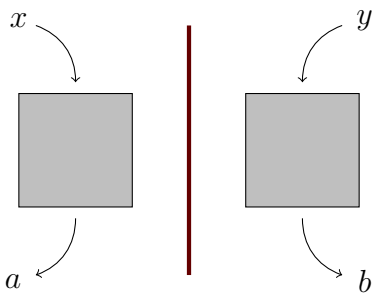

Fig. 1: The simplest Bell scenario.

mechanics we know that

$$P(a,b|x,y) = \mathrm{tr}\left[(F_a^x \otimes G_b^y)\rho_{AB}\right],$$

where ρ_{AB} is the state shared between Alice and Bob, $\{F_a^x\}$ are the measurement operators of Alice and $\{G_b^y\}$ are the measurement operators of Bob. Formalising the notion of local elements of reality, on the other hand, leads to probability distributions which can be written as

$$P(a,b|x,y) = \sum_\lambda p(\lambda)\, p_A(a|x,\lambda)\, p_B(b|y,\lambda).$$

Such probability distribution are called *local* and the breakthrough result of Bell was to rigorously show that quantum mechanics can generate probability distribution which are *not* of this kind. Such points are called *nonlocal* (in the sense of Bell) and we often say that quantum mechanics is nonlocal (although some find this name rather misleading). Local distributions are precisely the *convex hull* of deterministic strategies, i.e. strategies where the outputs of Alice and Bob can be written as deterministic functions:

$$a = f(x) \quad \text{and} \quad b = g(y).$$

Since there is only a finite number of deterministic strategies, the set of local strategies is a polytope. The quantum set is strictly larger and it is not a polytope.

Given a probability point $P(a,b|x,y)$ the easiest way to show that P does not belong to the local set is to find a separating hyperplane. Consider a real functional B given by coefficients c_{abxy} such that

$$\langle B, P\rangle := \sum_{abxy} c_{abxy} P(a,b|x,y),$$

where c_{abxy} are some real coefficients. Denote the maximal value of B over the local set by

$$\beta_L := \max_{P\in\mathcal{L}} \langle B, P\rangle.$$

If a probability point P gives $\langle B, P\rangle > \beta_L$, then we immediately conclude that $P \notin \mathcal{L}$. The functional B is the *Bell functional* or *Bell expression*, whereas the value β_L is known as the *local* or *classical value*. Together they form a Bell inequality:

$$\langle B, P\rangle \leq \beta_L,$$

which is valid for all local points P. A point which does not satisfy this inequality, i.e. when $\langle B, P\rangle > \beta_L$, is said to *violate* this Bell inequality. Note that when considering quantum-mechanical systems, the Bell value β can be alternatively written as $\beta = \mathrm{tr}(W\rho_{AB})$, where

$$W := \sum_{abxy} c_{abxy} F_a^x \otimes G_b^y$$

is the *Bell operator*.

The simplest Bell inequality is due to Clauser, Horne, Shimony and Holt (CHSH) [CHSH69] and applies to a scenario with two settings and two outcomes on each side. Measurements with two outcomes are conveniently represented as *observables*. If the measurement operators are $\{F_0, F_1\}$, the corresponding observable is given by

$$A := F_0 - F_1.$$

It is easy to check that A is Hermitian, $A = A^\dagger$ and satisfies $-\mathbb{1} \leq A \leq \mathbb{1}$. The CHSH functional is conveniently expressed in terms of the observables. The corresponding Bell operator reads

$$W = A_0 \otimes (B_0 + B_1) + A_1 \otimes (B_0 - B_1).$$

Computing the local value corresponds to replacing the operators by numbers $a_0, a_1, b_0, b_1 \in \{+1, -1\}$ and finding the maximal value. It is easy to see that in this case $\beta_L = 2$. On the other hand, a larger value can be achieved in quantum mechanics. Writing the two-qubit maximally entangled state $|\Phi^+\rangle_{AB} = (|0\rangle_A|0\rangle_B + |1\rangle_A|1\rangle_B)/\sqrt{2}$ in the Pauli basis gives

$$\Phi^+_{AB} = \frac{1}{4}\big(\mathbb{1} \otimes \mathbb{1} + \sigma_x \otimes \sigma_x - \sigma_y \otimes \sigma_y + \sigma_z \otimes \sigma_z\big).$$

The observables of Alice and Bob are given by

$$A_0 = \sigma_x, \quad B_0 = \frac{\sigma_x + \sigma_z}{\sqrt{2}},$$
$$A_1 = \sigma_z, \quad B_1 = \frac{\sigma_x - \sigma_z}{\sqrt{2}}.$$

It is easy to verify that the Bell operator reads

$$W = \sqrt{2}(\sigma_x \otimes \sigma_x + \sigma_z \otimes \sigma_z)$$

and therefore

$$\beta = \mathrm{tr}(W\Phi^+_{AB}) = 2\sqrt{2}.$$

Tsirelson showed that this is the maximal value achievable in quantum mechanics [Tsi80]. Moreover, one can also show that this is essentially the only way of achieving it (up to some well-understood equivalences) [Tsi87, SW87, PR92, Tsi93]. This phenomenon is called *self-testing*.

Our brief discussion of Bell nonlocality might seem like a detour, but we are now ready to appreciate its consequences for randomness generation. Since achieving $\beta = 2\sqrt{2}$ is possible in essentially one way, if Alice and Bob observe the maximal violation they essentially know what their quantum systems are doing, even if previously they had no knowledge about their inner workings. Since the marginal states of the maximally entangled state are maximally mixed, it is easy to verify that all the local expectation values vanish

$$\langle A_0 \rangle = \langle A_1 \rangle = \langle B_0 \rangle = \langle B_1 \rangle = 0.$$

This means that locally the outcomes $+1$ and -1 are equally likely. Since Alice and Bob share a pure entangled state, they cannot be correlated with anyone else. Therefore, if some external adversary, Eve, tries to guess the local outcome of Alice (or Bob), she cannot do any better than a random guess. This is precisely the idea of *device-independent randomness certification*: we start with completely uncharacterised devices, use them to violate some Bell inequality and use the violation to certify randomness. The first such proposal based on the tripartite GHZ state and the Mermin inequality appeared in the PhD thesis of Roger Colbeck [Col06].

2.5 Device-independent randomness

Let us begin by explaining the usual formulation of the problem. Alice and Bob share an unknown state ρ_{AB} and perform some uncharacterised measurement denoted by $\{F_a^x\}$ and $\{G_b^y\}$ to obtain statistics

$$P(a,b|x,y) = \mathrm{tr}\left[(F_a^x \otimes G_b^y)\rho_{AB}\right],$$

which violate some Bell inequality, i.e. $\beta = \langle B, P\rangle > \beta_L$. Moreover, there is an external adversary who holds the purification of the state, i.e. the overall state is given by Ψ_{ABE}. The goal of Eve is to guess the outcome of some measurement, e.g. the measurement of Alice corresponding to $x = 0$. To achieve this she performs a measurement $\{H_e\}$ on her part of the system. The probability of guessing successfully is given by

$$\mathrm{p}_{\mathrm{guess}} := \sum_a \mathrm{tr}\left[(F_a^0 \otimes \mathbb{1} \otimes H_a)\Psi_{ABE}\right].$$

Note that since Eve always performs the same measurement, her attack is equivalent to sending in a classically correlated state

$$\rho_{ABE} := \sum_e \rho_{AB}^e \otimes |e\rangle\langle e|,$$

where

$$\rho_{AB}^e := \mathrm{tr}_E\left[(\mathbb{1}_{AB} \otimes H_e)\Psi_{ABE}\right].$$

Since her attack is essentially classical, we will see that the violation vs. guessing trade-off is fully determined by the relation between the violation and the randomness in the local outcomes. Finally, we are interested in deriving bounds of the form:

$$\mathrm{p}_{\mathrm{guess}} \leq f(\beta)$$

for some explicit function f.

Let us consider the simplest example: the CHSH inequality. We are interested in the trade-off between the CHSH violation β and Eve guessing the outcome of Alice corresponding to $x = 0$. As mentioned before, this reduces to understanding the trade-off between β and the local marginal $\langle A_0\rangle$. Fortunately, thanks to the tilted CHSH inequality [AMP12] we know that

$$|\langle A_0\rangle| \leq \sqrt{2 - \beta^2/4}. \tag{1}$$

This result is not hard to prove, but it is beyond the scope of these notes. Clearly, without any side information the guessing probability and the marginal distribution are related by:

$$\mathrm{p}_{\mathrm{guess}}(A_0) = \frac{1 + |\langle A_0\rangle|}{2}.$$

The attack of Eve essentially introduces an extra classical random variable E distributed according to probability distribution $p(e)$. Conditional on $E = e$, the marginal is given by $\langle A_0\rangle_e$, whereas the violation is given by β_e and note that the observed violation satisfies

$$\beta = \sum_e p(e)\beta_e.$$

Since Eq. (1) holds for every value of e, we can write

$$\begin{aligned}\mathrm{p}_{\mathrm{guess}}(A_0|E) &= \sum_e p(e) \cdot \frac{1 + |\langle A_0\rangle|_e}{2} \leq \frac{1}{2} + \frac{1}{2}\sum_e p(e)\sqrt{2 - \beta_e^2/4} \\ &\leq \frac{1}{2} + \frac{1}{2}\sum_e p(e)\sqrt{2 - \beta_e^2/4} \leq \frac{1}{2} + \frac{1}{2}\sqrt{2 - \beta^2/4},\end{aligned}$$

where in the last step we have used the fact that the function under the square root is concave. This fundamental trade-off was first proven by Pironio et al. [PAM+10] and for clarity, let us restate it:

$$\mathrm{p}_{\mathrm{guess}}(A_0|E) \leq \frac{1}{2} + \frac{1}{2}\sqrt{2 - \beta^2/4}. \tag{2}$$

It is easy to check that $\beta = 2\sqrt{2}$ gives maximal randomness, which should not be surprising in the light of the previous self-testing results. The maximal violation requires Alice and Bob to perform anticommuting Pauli measurements on the pure, maximally entangled state of two qubits. This implies that the local distributions of outcomes are uniform and uncorrelated from Eve. The classical bound $\beta = \beta_L$ gives no randomness at all. This is not surprising, because any local distribution can be written as a convex combination of deterministic strategies and in the adversarial scenario Eve could have access to the random variable λ. Therefore, from her point of view all the outcomes are deterministic and no randomness can be certified. This shows that one needs nonlocal correlations to certify randomness.

Note that if we generate randomness from the maximally entangled state of two qubits, then 1 bit of randomness (locally) is the most we can hope for if we are restricted to projective measurements. It is natural to ask, whether more randomness can be certified if we employ non-projective measurements. Generally, a quantum system of local dimension d can be used to certify at most $\log d$ bits of randomness using projective measurements, but this bound increases to $2\log d$ for non-projective measurements. Acín et al. have shown that indeed using non-projective measurements can increase the amount of certified randomness [APVW16]. By combining multiple variants of the CHSH inequality, they have shown that there exists measurements on the two-qubit maximally entangled state which certify 2 bits of randomness. However, their construction is not particularly robust to noise: already a small amount of noise brings the certified randomness below 1 bit, i.e. below what is possible with projective measurement. In other words, it seems that non-projective measurements can in principle be useful, but the improvement is immediately lost in any practical scenario.

The scenario described above is the most common model of device-independent randomness generation, because it matches the scenario of quantum key distribution (see Sec. 2.6). However, one can also look at other versions of this problem. For example instead of protecting randomness of Alice against Eve, one might want to protect it against Bob. This scenario is a bit more complicated as Bob can now act in two possible ways: either he cooperates to produce the highest Bell violation possible or he is malicious and he tries to guess the outcomes of Alice. Clearly, the CHSH inequality in the standard setup is not a suitable candidate here, because the maximal violation is achieved using the maximally entangled state. We know that in such a case for every projective measurement of Alice, there exists a projective measurement of Bob, which produces perfectly correlated outcomes. In other words, even if Alice observes the maximal violation, she has no guarantee that her outcomes are secure against Bob. The problem can be resolved by introducing an extra restriction motivated by the so-called bounded storage model. Suppose that in the guessing round Alice flips a fair coin and measures either A_0 or A_1. However, Bob does not have a quantum memory and he must perform a measurement on his system *before* he finds out which measurement Alice performed. He stores his measurement outcome, then he learns the measurement setting of Alice and finally he is challenged to guess her outcome. In this scenario the guessing probability averaged over the two settings of Alice satisfies [KW16]

$$p^{\mathrm{ave}}_{\mathrm{guess}} \leq \frac{1}{2} + \frac{1}{4}\Big(\beta/2 + \sqrt{2 - \beta^2/4}\Big).$$

The bound is non-trivial (meaning the the right-hand side is strictly smaller than 1) whenever $\beta > 2$. Note that the maximal violation $\beta = 2\sqrt{2}$ only guarantees that

$$p^{\mathrm{ave}}_{\mathrm{guess}} \leq \frac{1}{2} + \frac{1}{2\sqrt{2}}.$$

This should not be particularly surprising: this is precisely the guessing probability achieved if Bob performs the measurements B_0. In fact, this bound turns out to be tight for all $\beta \in [2, 2\sqrt{2}]$.

Instead of placing such a memory restriction, we can look for inequalities which are maximally violated by non-maximally entangled states, e.g. the family of *tilted CHSH inequalities* introduced by Acín, Massar and Pironio [AMP12]. These inequalities are self-tests, just like the CHSH inequality, i.e. the maximal violation is achieved by a unique quantum realisation. In this quantum realisation there is one measurement of Alice, which Bob cannot guess perfectly. In fact, by choosing the right inequality from the family we can make Bob's guessing as hard as we want, i.e. we can ensure that

$$\mathrm{p}_{\mathrm{guess}}(A_1|B) \leq \frac{1}{2} + \varepsilon$$

for any $\varepsilon > 0$.

Let us conclude with a brief discussion of another security model proposed recently by Miller and Shi [MS17]. In the trade-offs discussed above, each device was used only once. Miller and Shi considered a two-round variant of randomness generation. In their scenario in the first round Alice and Bob perform the standard Bell test, except that Bob's measurement can be a partial measurement, i.e. it can output some leftover quantum state. In the second round Bob is given the output of Alice and is challenged to guess her outcome. They show that whenever Alice and Bob observe some violation, Bob's guessing in the second round cannot be perfect. This implies that after performing a Bell test, the local outcomes are random even for the other party, which opens up the possibility of randomness "recycling".

2.6 Device-independent quantum key distribution

Having discussed various aspects of device-independent randomness generation, let us conclude with the other great discovery of this field. Quantum key distribution [BB84, Eke91] is one of the two reasons why the field of quantum information gained such popularity in the late '90s (the other being Shor's algorithm [Sho94]). In the '00s quantum distribution became even it better: it was made device-independent [BHK05, AGM06, ABG+07].

We assume familiarity with the general idea of QKD and the standard (device-dependent) protocols, so let us go directly to describing the protocol proposed in Ref. [ABG+07].

In entanglement-based QKD Alice and Bob share an unknown state ρ_{AB}. In the usual (device-dependent) scenario this state is guaranteed to a be a two-qubit state on which Alice and Bob perform trusted measurements. In the device-independent setting the state is arbitrary (local dimensions are not known) and the local observables are uncharacterised. We only know that Alice has two binary observables to choose from (denoted by A_0, A_1), whereas Bob has three (B_0, B_1, B_2). In the ideal setup the state shared between Alice and Bob is the maximally entangled state of two qubits $\rho_{AB} = \Phi^+_{AB}$, while the observables A_0, A_1, B_0, B_1 are the ones giving the maximal CHSH violation

$$A_0 = \sigma_x, \quad B_0 = \frac{\sigma_x + \sigma_z}{\sqrt{2}},$$
$$A_1 = \sigma_z, \quad B_1 = \frac{\sigma_x - \sigma_z}{\sqrt{2}}.$$

The last observable of Bob is simply $B_2 = \sigma_x$. The protocol consists of two types of rounds: the testing rounds in which Alice and Bob check the CHSH violation on the observables A_0, A_1, B_0, B_1 and key generation rounds in which Alice measures A_0 and Bob measures B_2. At the end Alice and Bob announce the measurement outcomes for some of the rounds to estimate the CHSH violation

$$\beta := \langle A_0 B_0 \rangle + \langle A_0 B_1 \rangle + \langle A_1 B_0 \rangle - \langle A_1 B_1 \rangle$$

and the quantum bit error rate

$$Q := P(A_0 \neq B_2) = \frac{1 - \langle A_0 B_2 \rangle}{2}.$$

Clearly, if the devices adhere to the ideal specification we will observe $\beta = 2\sqrt{2}$, $Q = 0$ and each key round generates 1 bit of perfect key. To show that this protocol is indeed device-independent, we must demonstrate that security can be deduced based solely on observable quantities, in this case β and Q. More specifically, we allow Eve to distribute an arbitrary state $|\psi\rangle_{ABE}$ and we allow the observables of Alice and Bob to be arbitrary. We do assume, however, that Eve performs a collective attack, i.e. she attacks every round in the same way independent from the other rounds. As explained in Section 2.7 this restriction can be lifted, but it makes the argument significantly more complicated.

After performing the quantum part of the protocol and estimating β and Q Alice announces information which allows Bob to reconstruct A_0 (using his knowledge of B_2). It is well-known that in such a one-way postprocessing scenario the optimal key rate is given by the Devetak-Winter rate [DW05]

$$r_{\mathrm{DW}} := I(A_0 : B_2) - \chi(A_0 : E).$$

Intuitively, we simply subtract Eve's knowledge (the second term) from the correlations initially shared by Alice and Bob (the first term). The first term is directly related to the quantum bit error rate

$$I(A_0 : B_2) = 1 - h(Q),$$

where

$$h(x) := -x \log x - (1 - x) \log(1 - x)$$

is the binary entropy. The information of Eve is captured by the Holevo quantity which for the classical-quantum state

$$\rho_{XQ} := \sum_x p_x |x\rangle\langle x| \otimes \rho_x$$

is defined as

$$\chi(X : Q) := H\Big(\sum_x p_x \rho_x\Big) - \sum_x p_x H(\rho_x).$$

Holevo quantity tells us how much information about the classical variable X can be obtained by measuring the quantum system Q. If the quantum system carries no information about the classical variable, i.e. all the ρ_x are the same, we obtain $\chi(X : Q) = 0$. The last step of the argument is to bound the Holevo information $\chi(A_0 : E)$ in terms of the CHSH violation β. Through a direct calculation Acín et al. prove that [ABG+07]

$$\chi(A_0 : E) \leq h\left(\frac{1 + \sqrt{(\beta/2)^2 - 1}}{2}\right)$$

and this trade-off turns out to be tight. Clearly, in spirit this bound is similar to Eq. (2): we use the CHSH violation observed between Alice and Bob to place an upper bound on the amount of information that Eve has about the outcomes. The only difference is the precise quantity that we bound. As expected setting $\beta = 2\sqrt{2}$ gives $\chi(A_0 : E) = 0$ and, in fact, we get a non-trivial bound whenever $\beta > 2$. Clearly, if there is no violation, no security can be guaranteed and indeed setting $\beta = 2$ gives the trivial bound $\chi(A_0 : E) \leq 1$.

Finally, we have obtained a lower bound on the Devetak-Winter rate as a function of Q and β. To perform a fair comparison with the device-dependent QKD, let us fix a particular noise model. A particularly popular model is the depolarising noise, i.e.

$$\rho_{AB} = p\Phi^+_{AB} + (1 - p)\frac{\mathbb{1}_{AB}}{4}$$

for some $p \in [0,1]$. It is easy to check that for this noise model the CHSH violation and the error rate are related by

$$\beta = 2\sqrt{2}(1-2Q).$$

It is easy to check numerically that a positive rate is guaranteed whenever $Q < 7.1\%$. The analogous threshold for device-dependent QKD equals 11%, i.e. device-indepedendent QKD is less resistant to noise, but the difference is not striking.

2.7 Beyond i.i.d. models

In the previous section we have discussed cases in which Bell violation certifies randomness. However, if we want to construct an actual scheme which produces randomness, we immediately realise that we have until now completely missed an important point, namely that *performing a Bell test requires randomness*. For instance the input to a single round of the CHSH test consists of one bit for each system, i.e. two bits in total.

One of the fundamental assumptions of science is that the basic laws governing the behaviour of the universe should not change in time. This means that if we perform an experiment today and then an identical experiment tomorrow, we should not see any difference. Our ability to repeat the experiment many times and our hope that the different experimental trials are independent allows us to collect statistics and estimate probabilities of various events. Note that this fundamental assumption is by construction *unverifiable*: the basic laws of the universe could change tomorrow (rendering all research efforts up to day completely worthless) and there is no way of ruling out this possibility.

That is why in most experiments one implicitly assumes that the different experimental runs are independent. In quantum information this is often referred to as the *i.i.d. assumption*, which stands for *independent and identically distributed*. This comes from probability theory in which two random variables X_1 and X_2 over a finite set $\mathcal{X}$ are independent if

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2)$$

for all $x_1, x_2 \in \mathcal{X}$. Not surprisingly, random variables X_1 and X_2 are identically distributed if

$$P(X_1 = x) = P(X_2 = x)$$

for all $x \in \mathcal{X}$. A sequence of random variables $(X_1, X_2, \ldots, X_n)$ which are i.i.d. can be thought of as n independent copies of the same random variable.

In the context of nonlocality and device-independence the i.i.d. assumption states that our devices have a well-defined behaviour which does not change in time. Every time we push the button the source emits a particular bipartite state ρ_{AB}. We might not know what it is, but it is always the same state.

The i.i.d. assumption feels natural from the physics point of view, but it might cause some discomfort in the context of Bell nonlocality and device-independence. For instance a crucial aspect of a Bell test is that the input of Alice is not available to the device of Bob and vice versa. On the other hand, if we are truly convinced that the i.i.d. assumption holds, we could collect statistics corresponding to distinct input pairs $\{0,0\}, \{0,1\}, \{1,0\}, \{1,1\}$ on different days. In fact, since switching between different measurement settings takes time, this is precisely what was done in in the early Bell experiments. In the context of Bell tests this is nowadays considered a serious shortcoming known as the *locality loophole* (the choice of measurement settings for Alice can in principle be communicated to the device of Bob during the experiment).

Proving security of a scheme under the i.i.d. assumption is in most cases precisely as hard as proving the fundamental trade-off of the type presented in Section. 2.5. For instance if we want to use the CHSH scheme to generate randomness, we first extensively test the device to obtain a good

estimate of the CHSH value. Since the device is assumed to behave in an i.i.d. fashion, this testing can be done deterministically. In the second step we simply fix a pair of measurement settings, e.g. $\{0, 0\}$ and use it to generate randomness. The amount of randomness is guaranteed by the fundamental bound given in Eq. (2) and since every use of the device is independent, we get a fresh portion of randomness every time. In other words, we have a scheme for generating randomness from scratch.

Dropping the i.i.d. assumption makes the problem significantly harder. First of all, the testing must be randomised (any deterministic testing procedure can be fooled by preprogrammed devices), i.e. we must initially invest some randomness. Therefore, in this context one usually speaks of *randomness expansion*.

This also means that when designing a protocol there is a non-trivial trade-off between the number of rounds used for testing (which costs randomness) and the rounds used to generate randomness. Clearly, the testing and the randomness generation must be mixed up together so that the devices do not know whether they are being tested or used for randomness generation. This prevents the "Volkswagen attack", i.e. "act honestly when being tested and cheat when not tested". Security analysis for non-i.i.d. devices is much harder and until recently gave significantly weakened security guarantees. However, a recent development known as the "entropy accumulation theorem" suggests that in many situations the non-i.i.d. bounds are in fact comparable with the i.i.d. ones [AFDF+18].

3 Conclusions

In these notes we have presented a compressed introduction to randomness, particularly in the context of quantum physics. The selection of topics reflects entirely author's personal taste and many important results had to be skipped due to time limitations. Nevertheless, we hope that it serves as a useful introduction to the field. For further reading we recommend two short reviews on randomness: Ref. [ER14] and Ref. [AM16], which focus on quantum key distribution and randomness generation, respectively. For a more comprehensive review on randomness (from philosophy to technology), we refer the reader to Ref. [BAK+17].

References

[ABG+07] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani. Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.*, 98(230501), 2007.
DOI: 10.1103/PhysRevLett.98.230501.

[AFDF+18] R. Arnon-Friedman, F. Dupuis, O. Fawzi, R. Renner, and T. Vidick. Practical device-independent quantum cryptography via entropy accumulation. *Nat. Commun.*, 9(459), 2018.
DOI: 10.1038/s41467-017-02307-4.

[AGM06] A. Acín, N. Gisin, and L. Masanes. From Bell's theorem to secure quantum key distribution. *Phys. Rev. Lett.*, 97(120405), 2006.
DOI: 10.1103/PhysRevLett.97.120405.

[AM16] A. Acín and L. Masanes. Certified randomness in quantum physics. *Nature*, 540(213), 2016.
DOI: 10.1038/nature20119.

[AMP12] A. Acín, S. Massar, and S. Pironio. Randomness versus nonlocality and entanglement. *Phys. Rev. Lett.*, 108(100402), 2012.
DOI: 10.1103/PhysRevLett.108.100402.

[APVW16] A. Acín, S. Pironio, T. Vértesi, and P. Wittek. Optimal randomness certification from one entangled bit. *Phys. Rev. A*, 93(040102(R)), 2016.
DOI: 10.1103/PhysRevA.93.040102.

[BAK+17] M. N. Bera, A. Acín, M. Kuś, M. W. Mitchell, and M. Lewenstein. Randomness in quantum mechanics: philosophy, physics and technology. *Rep. Prog. Phys.*, 80(124001), 2017.
DOI: 10.1088/1361-6633/aa8731.

[BB84] C. H. Bennett and G. Brassard. Quantum cryptography: public key distribution and coin tossing. *Proc. International Conference on Computers, Systems and Signal Processing*, 1984.

[Bel64] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1(195), 1964.

[BHK05] J. Barrett, L. Hardy, and A. Kent. No signaling and quantum key distribution. *Phys. Rev. Lett.*, 95(010503), 2005.
DOI: 10.1103/PhysRevLett.95.010503.

[BSTPS16] M. Barbulescu, A. Stratulat, V. Traista-Popescu, and E. Simion. RSA weak public keys available on the internet. *Innovative Security Solutions for Information Technology and Communications. SECITC 2016. Lecture Notes in Computer Science*, 10006(92), 2016.
DOI: 10.1007/978-3-319-47238-6_6.

[CHSH69] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23(880), 1969.
DOI: 10.1103/PhysRevLett.23.880.

[Col06] R. Colbeck. *Quantum and relativistic protocols for secure multi-party computation.* PhD thesis, University of Cambridge, 2006.
arXiv: 0911.3814.

[DW05] I. Devetak and A. Winter. Distillation of secret key and entanglement from quantum states. *Proc. R. Soc. A*, 461(207), 2005.
DOI: 10.1098/rspa.2004.1372.

[Eke91] A. Ekert. Quantum cryptography based on Bell's theorem. *Phys. Rev. Lett.*, 67(661), 1991.
DOI: 10.1103/PhysRevLett.67.661.

[EPR35] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47(777), 1935.
DOI: 10.1103/PhysRev.47.777.

[ER14] A. Ekert and R. Renner. The ultimate physical limits of privacy. *Nature*, 507(443), 2014.
DOI: 10.1038/nature13132.

[Gis18] N. Gisin. Indeterminism in physics, classical chaos and Bohmian mechanics. Are real numbers really real? 2018.
arXiv: 1803.06824.

[KW16] J. Kaniewski and S. Wehner. Device-independent two-party cryptography secure against sequential attacks. *New J. Phys.*, 18(055004), 2016.
DOI: 10.1088/1367-2630/18/5/055004.

[MPL+18] D. G. Marangon, A. Plews, M. Lucamarini, J. F. Dynes, A. W. Sharpe, Z. Yuan, and A. J. Shields. Long-term test of a fast and compact quantum random number generator. *J. Light. Technol.*, 36(3778), 2018.
DOI: 10.1109/JLT.2018.2841773.

[MS17] C. A. Miller and Y. Shi. Randomness in nonlocal games between mistrustful players. *Quant. Inf. Comp.*, 17(595), 2017.
arXiv: 1610.05140.

[MSL+15] A. Martin, B. Sanguinetti, C. C. W. Lim, R. Houlmann, and H. Zbinden. Quantum random number generation for 1.25-GHz quantum key distribution systems. *J. Light. Technol.*, 33(2855), 2015.
DOI: 10.1109/JLT.2015.2416914.

[PAM+10] S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe. Random numbers certified by Bell's theorem. *Nature*, 464(1021), 2010.
DOI: 10.1038/nature09008.

[PR92] S. Popescu and D. Rohrlich. Which states violate Bell's inequality maximally? *Phys. Lett. A*, 169(411), 1992.
DOI: 10.1016/0375-9601(92)90819-8.

[SGG+00] A. Stefanov, N. Gisin, O. Guinnard, L. Guinnard, and H. Zbinden. Optical quantum random number generator. *J. Mod. Opt.*, 47(595), 2000.
DOI: 10.1080/09500340008233380.

[Sho94] P. W. Shor. Algorithms for quantum computation: discrete logarithms and factoring. *Proceedings 35th Annual Symposium on Foundations of Computer Science*, 1994.
DOI: 10.1109/SFCS.1994.365700.

[SW87] S. J. Summers and R. F. Werner. Maximal violation of Bell's inequalities is generic in quantum field theory. *Commun. Math. Phys.*, 110(247), 1987.
DOI: 10.1007/BF01207366.

[Tsi80] B. S. Tsirelson. Quantum generalizations of Bell's inequality. *Lett. Math. Phys.*, 4(93), 1980.
DOI: 10.1007/BF00417500.

[Tsi87] B. S. Tsirelson. Quantum analogues of the Bell inequalities. The case of two spatially separated domains. *J. Soviet Math.*, 36(557), 1987.
DOI: 10.1007/BF01663472.

[Tsi93] B. S. Tsirelson. Some results and problems on quantum Bell-type inequalities. *Hadronic J. Suppl.*, 8(329), 1993.