Physical Random Number Generator

Yinghao Huang

***Abstract*—A method to generate physical random numbers by using reverse-bias P-N junction as the entropy source was proposed by Aaron Logue in 2002. Due to the phenomenon of quantum tunnelling, a small portion of electrons can cross the P-N junction. Base on this, in our project, we amplified the noise generated by using a high-gain amplifier, and the noise was then digitized by using an inverter. During the experiment, we found that the probabilities of logic 0 and 1 in the output of the inverter are not identical. For solving this issue, three steps of unweighting are applied to the signal, which is, making the probabilities of logic "0" and "1" are nearly the same (by using two diodes as an indicator) and implementing the von Neumann algorithm twice to the signal by a shift register and the computer. A high-speed physical random number generator, which passed the Diehard test, and could achieve a real-time bit rate of $160kbit/s$, is built.**

I. INTRODUCTION

THE random number generator is widely used in Monte-Carlo simulations, deep learning, cryptology etc. Especially, for secure communication, the random number generator is the most important component to guarantee the integrity and security of cryptographic protocols. With the development of the Internet, a high-speed secure random number generator is becoming more and more important.

At present, there are mainly two kinds of random number generators, i.e. pseudorandom number generator and physical random number generator. Pseudorandom numbers are generated by some algorithms (such as linear congruence). Although this kind of generators could generate random numbers at a high rate, it is not truly random. The periodicity of pseudorandom numbers makes it possible to be cracked; the security of the communication cannot be guaranteed. On the contrary, physical random number generators could generate genuine random numbers, which is absolutely secure for communication.

The physical random number generators use the physical random process as the entropy source, such as the thermal noise of a resistor, the jitter of signal and quantum processes. To generate random numbers from the thermal noise of a resistor, the current flows through the resistor needs to be very large. It is low efficient compared with the energy that it consumes. By using the jitter of signals, generating random numbers at a high rate is impossible. It cannot satisfy the demands about the generating speed.

The quantum process can be energy-saving, and it is possible to generate random numbers at a high rate. These processes have a solid theoretical background to guarantee the randomness of the process, such as quantum tunnelling and the decay of substance.

In 2002, Aaron Logue designed a truly random number generator which uses a reverse-bias P-N junction as the entropy source[2]. The generator is base on the phenomenon of quantum tunnelling. Our project is based on his idea. In our project, the reverse-bias P-N junction is used as a raw noise generator. The noise generated is amplified by a high-gain amplifier. Due to the existence of intrinsic bias (non-identical proportion of 0 and 1), we used both physical and computational methods to unweighting the random sequence. A high-speed ($160kbit/s$) genuine random number generator is finally built.

II. THEORY

A. P-N Junction

A P-N junction is a boundary or interface between two types of semiconductor materials, P-type and N-type, inside a single crystal of semiconductor. The P-type semiconductor has excess holes, while the N-type semiconductor has excess electrons. Some electrons diffuse into the P-type semiconductor due to the difference of concentration of electrons. These electrons are combined with the holes near the junction. Consequently, a layer of negative ions near the junction is formed inside P-type semiconductor, while a layer of positive ions is formed inside N-type semiconductor (shown in Fig.1). These layers are called a space charge region, which could also be perceived as a potential barrier for the electrons in N-type semiconductor. The space charge region generates an electric field which drifts the electrons inside P-type semiconductor to N-type semiconductor. The system will reach an equilibrium state, i.e. the rate of electrons leaving N-type semiconductor (due to diffusion) is equal to the rate of electrons enter N-type semiconductor (due to the electric field in the space charge region).

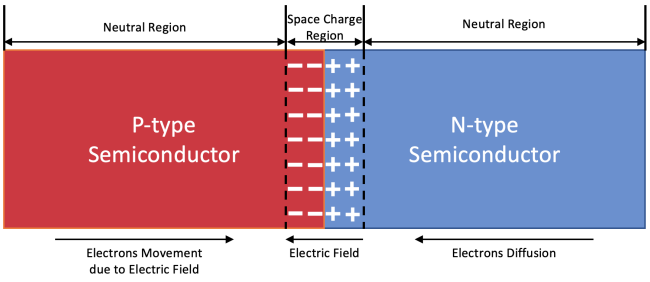


Fig. 1: Structure of a P-N Junction

There are two ways to supply an external electric field to a P-N junction, that is, forward bias and reverse bias. In forward bias, the positive side of a voltage source is connected to the P-type semiconductor. In this case, the electric field, which is generated by the space charge region, is weakened, and the width of the space charge region is shortened. Electrons in N-type semiconductor could cross the junction easily. In the case of reverse bias, the positive side of a voltage source

is connected to the N-type semiconductor. The electric field generated by the space charge region is enhanced, and the width of the space charge region is lengthened. Classically, no electrons inside N-type semiconductor can cross the junction. However, a small portion of electrons could cross the potential barrier because electrons act as quantum particles. Quantum tunnelling allows the small particle to cross a potential barrier, which is higher than the kinetic energy of the particle itself, with a certain probability. It is completely unpredictable and random. The small current flowing through the junction fluctuates significantly in a short time because the time taken for each quantum tunnelling is extremely small. Thus, connecting a P-N junction in reverse bias is an excellent way to produce high-frequency random signals.[5]

B. Zener Diode

A Zener diode is a type of diode that allows current to flow from cathode to anode when a certain voltage, which is called Zener voltage, is reached. And it is a kind of diode which current change significantly when the reverse-biased voltage across it is near to the Zener voltage. Due to this property, the maximum voltage across a reverse-biased Zener diode is the Zener voltage when the current in the circuit is limited (if the current is not limited, the Zener diode will be destructed). Thus, the output voltage in Fig. 2 is limited and the maximum of it is the Zener voltage. The Zener diode could be used as a voltage regulator to clip the voltage higher than a certain value by connecting it in parallel between the voltage input and the voltage output.[5]

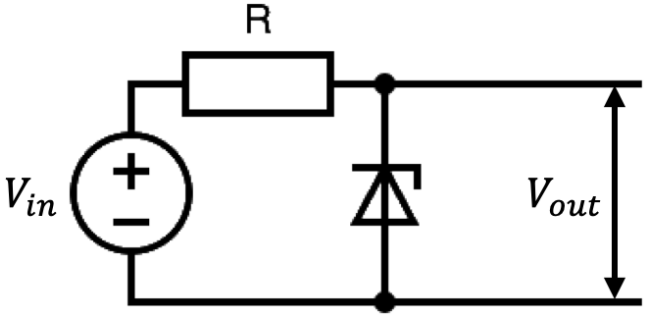


Fig. 2: A Zener Diode as a Voltage Regulator

C. NPN Transistor Amplifier

An NPN transistor contains three parts, which are emitter, base and collector. The emitter and the collector are made of N-type semiconductor, while the emitter is highly negatively doped. The base is a thin layer of P-type semiconductor. The diagram of an NPN transistor is shown in Fig. 3.

If the collector connected to the positive terminal of a high voltage source and the emitter is connected to the negative terminal of the source, the current flows in the transistor can be ignored. Under this circumstances, if a new small voltage source is added and the positive terminal is connected to the base and the negative terminal is connected to the emitter, a small current flows through the P-N junction from the base to

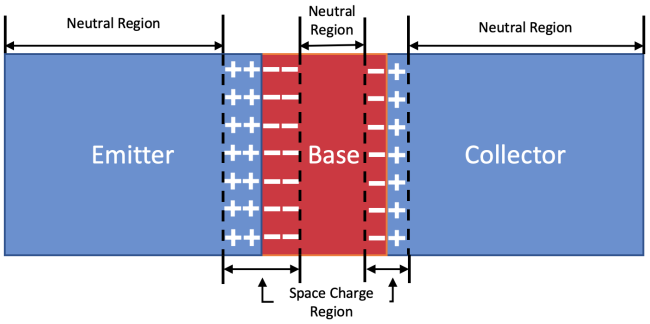


Fig. 3: Structure of a Transistor

the emitter. The potential barrier between emitter and base is weakened. Under the effect of the larger voltage source and the fact that the emitter is highly negatively doped, a large number of electrons in the emitter is able to cross the junction and enter the base. Because the base is thin, the electrons are attracted by the positive ions in the space charge region inside the collector, and most of it is able to cross the barrier between collector and base. As a result, a large current is generated from the collector to the emitter. Due to this property, A NPN transistor could be used as a current amplifier. By adding loads to the circuit of the current amplifier, the current amplifier could be used as a voltage amplifier. The simplest voltage amplifier is shown in Fig.4.

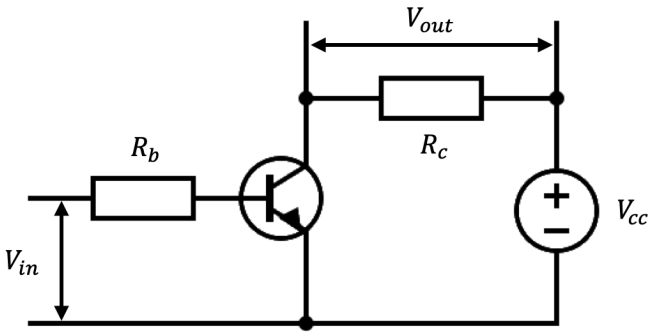


Fig. 4: A NPN Transistor as a Voltage Amplifier

The gain of the voltage amplifier is:

$$g_V = \frac{V_{out}}{V_{in}} = \frac{R_c I_c}{R_b I_b} = \beta \frac{R_c}{R_b} \tag{1}$$

where I_b is the current flow through resistor R_b while I_c is the current flow through resistor R_c, and β, the gain of the current, is a characteristic constant of an NPN transistor itself. Equation. (1) shows that the gain of voltage is independent of the magnitude of current and the gain of the voltage amplifier could be adjusted by changing the magnitude of the load.[5]

D. Von Neumann Algorithm

A Bernoulli process is a discrete-time stochastic process that takes only two values, canonically ”0” and ”1”. The sequence generated by this process is called the Bernoulli sequence which consists of ”0” and ”1”. The probabilities of ”0” and ”1” in the sequence are not necessarily equal. To produce a random binary sequence of identical probabilities of ”0” and ”1”, the

von Neumann algorithm was designed by von Neumann in 1951. Qualitatively speaking, the algorithm is based on a very simple fact, that the probability of the pair of "01" and "10" are equal. Thus, an even binary random sequence can be produced by extracting the first digit of the pairs "01" and "10" from the original Bernoulli sequence and ignoring the pairs "00" and "11".[4]

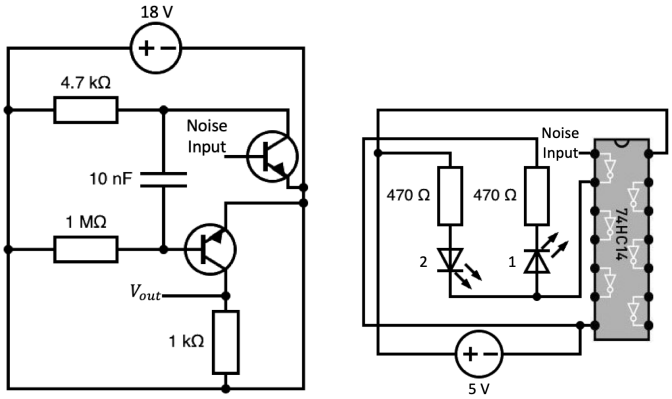
For the logic signals in a circuit, the von Neumann algorithm could be implemented by using an integrated circuit–shift register. A shift register consists of 8 D-type flip-flops. The first flip-flop uses the data input of the shift register, while others use the output of the previous flip-flop as the input. The flip-flops use the clock to sample the data input. Thus, by controlling the clock input to the shift register, the frequency of the output sequence can be adjusted.

III. METHOD

A. Set-up of Circuit

To set up the circuit, a reverse-biased P-N junction in an NPN transistor is used as the raw random noise generator. The voltage source is chosen to be $18V$ to make the noise generated not too small. To limit the current and protect the transistor, this transistor is connected with a $4.7k\Omega$ resistor in series.

The high-gain voltage amplifier (shown in Fig. 5(a)) consists of two NPN transistor, a $10nF$ capacitor, and three resistors which are $1k\omega$, $4.7k\Omega$ and $1M\Omega$ respectively. The amplifier amplifies the noise twice to make sure that the resultant voltage is large enough to analyse. The $1M\Omega$ is used to limit the current flows in the circuit to an extremely small value, to prevent the effect of current to the randomness of the signal. The capacitor in the circuit is used to blocks the DC voltage supply by the source so that the amplified voltage noise could be passed directly from the first transistor to the second transistor. For reducing the effect of noise generated by the circuit, every component of the amplifier are placed as close as possible.



(a) High-Gain Amplifier (b) Indicator of Weighting

Fig. 5: Two Important Components of the Circuit

A Zener diode $1N750$ with Zener voltage $4.7V$ is connected parallel to the high-gain amplifier. It regulates the output voltage signal from the amplifier and forbids the voltage signal high than $4.7V$.

For converting the analogue signal to digital signal, a hex inverting Schmitt trigger 74HC14 is used. The operating voltage of the integrated circuit is chosen to be $5V$, so that the input voltage which is larger than $2.5V$ results in the $5V$ output (logic "1"), while the input voltage which is smaller than $2.5V$ results in the $0V$ output (logic "0").

During the experiment, the phenomenon, that there is a large difference between the probability of logic "1" and "0" in the output sequence of the inverter, was observed. If the von Neumann algorithm applied to the sequence, there is a huge loss of information, i.e. most of the pairs in the sequence are "00" or "11" which are ignored by the algorithm directly. After some rigorous tests, the reason, that the amplifier overamplified the signal, was found. The amplified voltage signal in most of the time is larger than $4.7V$. The Zener diode clips these parts of the signal and results in more "1" output from the inverter. To solve this issue, the load of resistance $1k\Omega$ in Fig. 5(a) is replaced by an adjustable load, which are made of two $1k\Omega$ resistor and a $10k\Omega$ slide resistor. By adjusting the value of the load, the gain of the amplifier can be changed. To make the effect of the change of the value of load be controllable, an indicator which consists of two diodes and two resistors is added to the circuit. (shown in Fig. 5(b)). The diode 1 is connected forward-bias to the output gate of the inverter. The diode 2 is reverse-biased to the output gate of the inverter, and forward-bias to a $5V$ voltage source. Thus, the resultant voltage across diode 2 is equal to $5V$ minus the voltage across the diode 1. The brightness of diodes is proportional to the average value of the voltage on it when the fluctuation of the voltage is high-frequency. Thus, the probabilities of logic "0" and "1" in the output digital signal of the inverter can be adjusted to be nearly equal by making the brightness of two diodes to be nearly equal through the slide resistor.

To implement the von Neumann algorithm, a shift register 74HC164 is used. The operating voltage is chosen to be $5V$, and the active LOW is chosen to be $0V$. The clock input of the shift register is generated by an oscilloscope RTB2004 which could generate at most $10MHZ$ clock input. For generating a high-frequency random binary sequence, the input clock frequency is chosen to be the maximum frequency of $10MHZ$. 74HC164 is not an ideal integrated circuit. The data input gate and operating voltage gate are affected by the clock input, i.e. the voltage of data input gate and operating voltage gate fluctuate due to the clock input. This could also influence other parts of the circuit. For reducing the effect of flaws of the shift register, a voltage regulator LM7805 is used. The voltage regulator worked with a $100nF$ capacitor and $9V$ voltage supply. The $5V$ voltage output from the regulator supplies to the operating voltage gates of integrated circuits and the diodes which are used for unweighting. Two $10nF$ capacitors are mounted parallel to the operating voltage gate of the integrated circuits. The final set-up of the circuit is shown in Fig. 6.

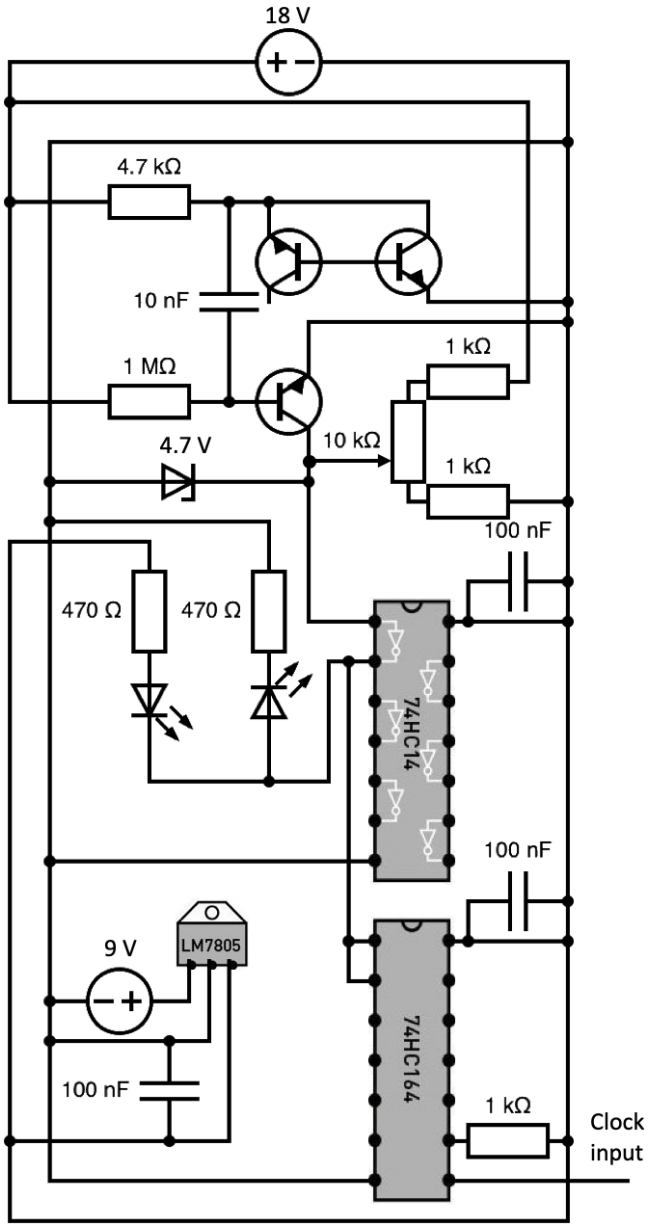


Fig. 6: The Final Set-up of Circuit

B. Data Processing

The final output of the shift register is measured by the oscilloscope. The oscilloscope could transfer the real-time data to the computer. The data is then converted from the values of voltages to binary numbers by setting the voltage signal higher than the average value to be "1" and setting the voltage signal less than the average value to be "0". Due to the flaws of the shift register, the von Neumann algorithm needs to be applied once more on the computer to guarantee the good statistical property of the final binary sequence. For conducting these processes, a Python program, which could convert the voltage signal to binary sequence and apply the von Neumann algorithm to the sequence, is built.

C. The Diehard Test

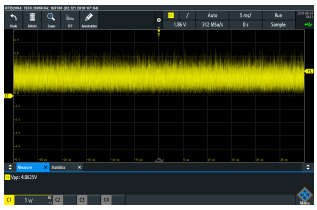
The Diehard test is one of the most canonical tests for testing the randomness of a random sequence. A standard Diehard test consists of 16 different tests (for a clear explanation of each test, please read [1]). The 16 tests check almost all the requirements that random sequences need to satisfy. Therefore, it is not necessary to discuss and check the properties of the random sequences one by one.

The program which is used to conduct this test is designed by George Marsaglia, and the code is open-source[3]. Each of these tests returns a p-value. If the p-values of all tests are ranged from 0.01 to 0.99, the Diehard test is passed.

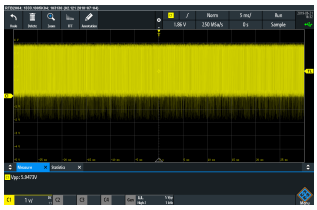
IV. Results

A. Waveforms of Signals

The waveform of signals is shown in Fig. 7. As the graph shows, the amplified, clipped voltage noise seems behaves randomly. The digitized noise and the final digital signal output from the shift register seem to have identical probabilities of logic "0" and "1".

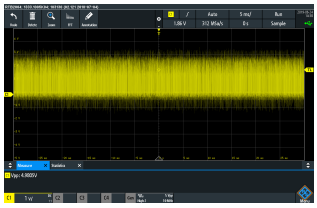


(a) The Clipped and Amplified Noise



(b) The Output Signal from the Inverter

(c) The Output Signal from the Shift Register (frequency of clock is 1kHZ)



(d) The Output Signal from the Shift Register (frequency of clock is 10MHZ)

Fig. 7: Waveforms of Singals

B. Generating Rate of Random Numbers

Under the condition that the frequency of clock input is $10MHZ$, the time interval on the oscilloscope is set to vary from $-120ms$ to $+120ms$. The minimum distance between logic "0" and "1" in the output signal of the shift register is $0.1ns$ when the frequency of the clock is $10MHZ$, t. Each output file contains $120,000$ data. Thus, the time interval that was chosen is the smallest one that the data output is discrete enough.

Every $240ms$, an file contains $120,000$ voltage data is generated. By implementing the von Neumann algorithm to each file, on average, a sequence contains about $40,000$ binary bits is generated from the original file. The rate of the generation of random numbers is about $160kbits/s$. In other words, it generates random numbers with a rate of $19.5Mb/s$.

C. The Diehard Test

The Diehard test needs at least $3,200,000$ binary bits. Thus, 82 files of raw voltage data were processed. The result of the Diehard test is shown in the TABLE. I.

TESTS	P-VALUE	STATUS
Birthday Spacing Test	0.684805	PASS
Overlapping 5-Permutation Test	0.931053	PASS
32×32 Binary Rank Test	0.743923	PASS
6×8 Binary Rank Test	0.924107	PASS
The Bitstream Test	0.179661	PASS
OPSO Test	0.942323	PASS
OQSO Test	0.924021	PASS
DNA Test	0.173164	PASS
Count the 1s Test	0.349382	PASS
Parking Lot Test	0.189101	PASS
Minimum Distance Test	0.974853	PASS
$3D$ Spheres Test	0.493209	PASS
The Squeeze Test	0.701253	PASS
Overlapping Sums Test	0.274532	PASS
Run Test	0.242144	PASS
The Craps Test	0.403718	PASS

TABLE I: Results of the Diehard Test

The random sequences pass all the tests. Therefore, random number generator generates truly random numbers.

D. Discussion

As Fig. 7(b) shows the maximum output voltage of the inverter is not the operating voltage of it. This might be caused by the flaws of the shift register because the clock input effects the operating voltage gate of the shift register, and the operating gate of the inverter is connected with that of the shift register. This phenomenon does not affect the randomness of the generator due to the following two reasons. Most of the data points still locate near the logic "0" and "1" (discrete). And the von Neumann algorithm is implemented once on the computer.

The rate of the generation of random numbers could be improved. The data lost around $\frac{2}{3}$ due to implementation of the von Neumann algorithm on the computer. For avoiding the loss of data due to the von Neumann algorithm, the shift register needs to be replaced by a better one. Additionally, by enhancing the sensitivity of the gauging equipment, more information could be extracted. Furthermore, by increasing the clock frequency, the rate of the generation of random numbers could be enhanced.

The generator passed the Diehard test, which means it generates real random numbers. However, there is an issue, that is the randomness comes from the quantum process? By placing the components of the high-gain amplifier as close as possible, the effect of the disturbances of the circuit itself is kept as small as possible. In spite of this, the contamination of the noise of the circuit to the original noise is not avoidable. Due to the limitation of the gauge instruments, the current from the P-N Junction cannot be measured directly. Therefore, there is not a way to determine that to what extent the noise generated by the circuit itself affect the final output.

V. CONCLUSION

The aim of this project is to build a physical random number generator. Basing on the phenomenon of quantum tunnelling, and using a reverse-bias P-N junction as an entropy source, a high-speed physical random number generator is built. The randomness of the sequence generated by this generator is tested and approved by the Diehard test. The generator generates random numbers with a rate of $160kbit/s$. The rate of generating random numbers could be improved by enhancing the frequency of clock input, enhancing the sensitivity of gauge instrument and using better circuit components.

However, due to the limitation of the instruments, a question, which is about the effect of the noise generated by the circuit on the original noise, cannot be answered in this project.

In my point of view, using P-N junction as an entropy source to generate random numbers is a very efficient and convenient method. Due to the limitation of instruments, we cannot build the best random number generator basing on this theory. However, I believe that the potential of this technique will be explored in the further.

REFERENCES

[1] Walter Anderson. *Dieharder Test Descriptions*. 2008. URL: https://sites.google.com/site/astudyofentropy/background-information/the-tests/dieharder-test-descriptions.

[2] Aaron Logue. *Hardware Random Number Generator*. 2002. URL: http://www.cryogenius.com/hardware/rng/.

[3] George Marsaglia. *The Marsaglia Random Number CDROM including the Diehard Battery of Tests of Randomness*. 1995. URL: https://web.archive.org/web/20161114211602/http://stat.fsu.edu/pub/diehard/.

[4] John von Neumann. "Various techniques used in connection with random digits". In: *John von Neumann, Collected Works* 5 (1963), pp. 768–770.

[5] David J Roulston and J David. *Bipolar semiconductor devices*. Vol. 11. McGraw-Hill New York, 1990.