A True Random Number Generator with Time Multiplexed Sources of Randomness

Vlastimil KOTE [1,2], *Patrik VACULA* [1,2], *Vladimir MOLATA* [1,2], *Ondrej VESELY* [2], *Ondrej TLASKAL* [2], *Dalibor BARRI* [1,2], *Jiri JAKOVENKO* [1], *Miroslav HUSAK* [1]

[1] Dept. of Microelectronics, Faculty of Electrical Engineering, Czech Technical University in Prague, Technicka 2, 166 27 Prague 6, Czech Republic
[2] STMicroelectronics, Pobrezni 620/3, 186 00 Prague 8, Czech Republic

kotevlas@fel.cvut.cz, patrik.vacula@st.com, vladimir.molata@st.com, ondrej.vesely@st.com, ondrej.tlaskal@st.com, dalibor.barri@st.com, jakovenk@fel.cvut.cz, husak@fel.cvut.cz

Submitted January 31, 2018 / Accepted May 1, 2018

Abstract. *A true random number generator (TRNG) with time multiplexed metastability-based sources of randomness, presented in this paper, is capable of generating random bit sequences formed from noise present in the electronic circuit. An incorporated time multiplexer interleaves digitized random signals coming from sources of randomness and increases output data rate. The proposed TRNG was fabricated in a STMicroelectronics 130 nm bulk CMOS technology on an area of 0.029 mm^2. The quality of all random bit sequences has been verified by the FIPS and NIST statistical test suites. The fabricated TRNG generates random bit sequences up to the data rate of 20 Mb/s without any corrective mechanisms at power consumption of 72.48 μW. The changing environmental conditions do not influence the quality of random bit sequences.*

Keywords

True random number generator (TRNG), source of randomness, time multiplexer, integrated circuit (IC), statistical test

1. Introduction

Cryptographic devices and secure communication systems use random numbers for their operations [1]. Random numbers are also basis for simulations of natural phenomena, are used in statistic science, and appear in commercial applications such as various lottery games. Random numbers can only be considered in case when the number is a part of a sequence of several other numbers which were generated independently within each other of a defined set or an interval with a probability distribution. Moreover, future values of random numbers are not possible to predict based on knowledge of present or past values of a number sequence.

Random numbers are usually generated by mathematical computational algorithms known as pseudorandom number generators (PRNGs). Number sequences generated by PRNGs only approximate to properties of true random number sequences and are not appropriate for use in cryptographic devices and communication systems because they are generated by an algorithm using present and past values. Thus, under certain conditions, future values could be theoretically predictable [2]. In these important systems and devices, there is essential to use true random number sequences, which are generated on the basis of a physical phenomenon random behavior in devices commonly called true random number generators (TRNGs). A fundamental parameter of these TRNGs used in the systems above-mentioned is sufficient output random data rate.

True random number sequences can be extracted from the chaos, which occurs in semiconductor lasers [3], from randomness occurring in Josephson junctions in superconductive integrated circuits [4]. Other TRNGs are based on radioactive decay [5] or magnetic tunnel junctions [6]. However, these principles are not suitable for TRNGs, which are parts of systems on chip (SoC), usually fabricated in standard complementary metal-oxide-semiconductor (CMOS) processes on silicon substrates. Thus TRNGs integrated into these SoCs are based on more principles appropriate such as thermal noise generated by resistors [7], oscillator jitter [8], [9] or circuits operating in a metastable state [10], [11]. In already published works [12], [13], devices combining above mentioned principles had been presented. Another type of TRNGs uses a chaotic based non-linear dynamic system. A TRNG described in [14] operates with a current mode skew tent map circuit and is suitable for low power mobile applications. The Sundarapandian-Pehlivan chaotic system is implemented in a high-speed TRNG, which is designed for embedded cryptographic applications [15]. Some published TRNGs [16], [17] use correctors such as the von Neumann corrector, the XOR corrector, or corrector based

on a linear feedback shift register. These correctors improve the properties of biased random sequences.

TRNGs provide unpredictable random number sequences but at a cost of lower data rate than the data rate of PRNGs. Rising level of system security increases demands on the amount of high-quality random numbers generated in the shortest possible time. Amount of generated random data is highly dependent on the used manufacturing process of TRNGs. For increasing generated random data at a given time, modern technologies such as superconductive circuits [4] or semiconductor lasers [3] can be used. However, TRNGs implemented in complex SoCs are proposed in technologies, which are suitable for these systems. It allows true random data generation directly in modern hand-held devices. Contemporary SoCs are usually designed in standard submicron CMOS technologies, in which the random data rate is limited. Therefore for these systems, the random data rate is increased by a transition from serial approach to the parallel use of more independent sources of randomness.

A TRNG working on the basis of electrical and thermal noises present in CMOS structures has been proposed and manufactured. Concretely for generation of random bits, metastability-based sources of randomness are used. Randomness is extracted from thermal noise and flicker noise, which are generated in metal-oxide-semiconductor field effect transistor (MOSFET) channels. For a random data rate increase, the parallelism of more sources of randomness is implemented. The designed TRNG is able to generate typically tens of random megabits per second, which is several times more than a conventional TRNG.

It is not possible to prove that an arbitrarily long sequence of random numbers is really random. So for assessment of random number properties, statistical tests are used. In this case, sequences of random numbers are evaluated by standard acclaimed suites of statistical tests called FIPS [18] and NIST [19].

The paper is organized as follows. The principle of the TRNG with time multiplexed sources of randomness is introduced in Sec. 2. Implementation of the proposed circuit in a 130 nm bulk CMOS technology is described in Sec. 3. The results obtained are presented in Sec. 4 and discussed in Sec. 5. Finally, the conclusion follows in Sec. 6.

2. Principle of Time Multiplexed TRNG

While increasing amount of random data at any defined time, technological limits of random data generation can be reached. Then while maintaining the same technology, the same way of random number generation and the same operating conditions of a TRNG - a further increase in the amount of random data in the defined time is impossible. Therefore it is necessary to find new options in TRNG architecture that allows increasing rate of random data whereas strict criteria of random sequences for cryptographic devices will be satisfied [19].

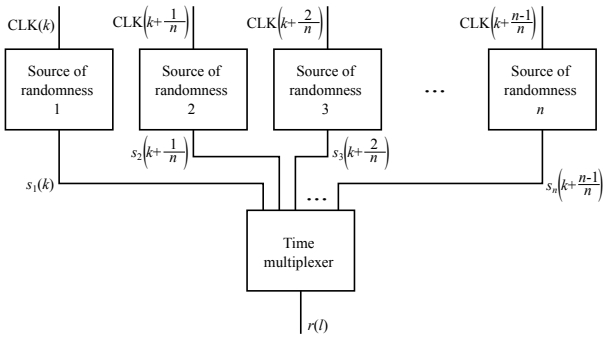

Fig. 1. Principle illustration of the TRNG with time multiplexed sources of randomness.

In this proposal, the parallelism of several sources of randomness is introduced. A composition of generated random data in parallel into a serial random data stream is defined on the hardware level. For the above composition, the pipeline principle is used. Thus, individual raw random data streams are generated synchronously with a phase shift and with a proportionally lower clock frequency rather than a clock frequency of an output random data stream. The output random data are created by sequential mixing of raw random data streams according to

$$\{r(l)\}_{l=0}^{\infty} = \left\{ s_1(k),\ s_2\left(k+\frac{1}{n}\right),\ s_3\left(k+\frac{2}{n}\right),\ \ldots,\ s_n\left(k+\frac{n-1}{n}\right) \right\}_{k=0}^{\infty} \tag{1}$$

where n is number of independent serial random data streams, $s_1,\ s_2,\ \ldots,\ s_n$ are random data streams, r is the output random data, k is a normalized period of the random data stream while $k = t/T_\mathrm{c}$, t is time, and T_c is a period of a clock signal, l is a normalized period of the output data stream while $l = nk$. Values of l and k are non-negative integers because random data is generated after the system starts at $t = 0$. The equation (1) may only be used under condition when the raw random data streams are generated in digitized sources of randomness independently. A bit rate of the raw random data stream is given by a clock frequency where $f_\mathrm{c} = 1/T_\mathrm{c}$. Naturally from the above-defined output random data, their bit rate can be derived according to

$$l = nk = n\frac{t}{T_\mathrm{c}} = nf_\mathrm{c}\,t \tag{2}$$

where f_c is a frequency of the clock signal. Thus the result of mixing is that the data rate of the output random data is n times higher than the bit rate of one digitized source of randomness. The used principle is depicted in Fig. 1 where CLK denotes a clock signal with the period T_c.

During mixing the randomness of generated data must be preserved. TRNGs do not generate a continuous analog signal but usually produce a digital signal with two voltage levels, which are represented by logic one and logic zero.

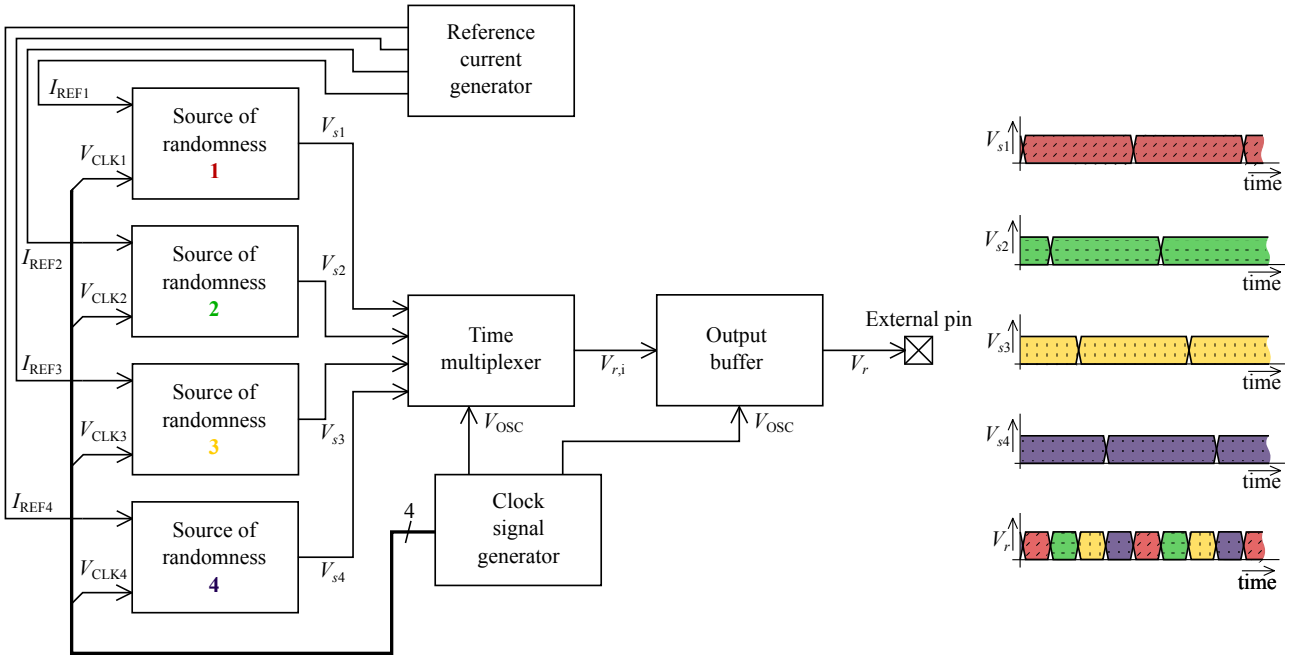

Fig. 2. Block diagram of the designed TRNG with time multiplexed sources of randomness.

Random data sequences composed of logic ones and zeros have significant properties, which must not be changed by data mixing. The most important one is the mean value of a bit sequence, which has to be close to $1/2$ in case of a random bit sequence. Thus, the mean value of output random data r has to be close $1/2$ after sequential mixing of single data streams $s_1,\ s_2,\ \ldots,\ s_n$, which are generated by fully functional and independent sources of randomness. One data stream can be described as

$$s_q = \left(s_1^q,\ s_2^q,\ s_3^q,\ \ldots,\ s_m^q\right) \tag{3}$$

where m is length of the data stream, s_i^q is a random value of the data stream, $i = 1, 2, 3, \ldots, m$, and q is an identifier of the source of randomness, $q = 1, 2, 3, \ldots, n$. In general, a random signal u with the mean value $\bar{u}$ is composed of consecutive data streams $s_1,\ s_2,\ \ldots,\ s_n$ so that

$$u = (s_1,\ s_2,\ s_3,\ \ldots,\ s_n) = \left(s_1^1,\ s_2^1,\ s_3^1,\ \ldots,\ s_m^n\right). \tag{4}$$

A random signal u_σ is formed by any permutation without repetition of values s_i^q of the random signal u and has the mean value $\bar{u}_\sigma$. At this point the distributive property of summation is used and then

$$\bar{u} = \frac{1}{n}\sum_{q=1}^{n}\left(\frac{1}{m}\sum_{i=1}^{m} s_i^q\right) = \frac{1}{nm}\sum_{q=1}^{n}\sum_{i=1}^{m} s_i^q. \tag{5}$$

Because the random signal u_σ is composed of values s_i^q and summation has the commutative property it can be shown that

$$\bar{u} = \frac{1}{nm}\sum_{q=1}^{n}\sum_{i=1}^{m} s_i^q = \bar{u}_\sigma. \tag{6}$$

Output random data r is a permutation of the random signal u, which means that the mean value of output random data is not changed during mixing.

3. Circuit Implementation

The introduced TRNG with time multiplexed sources of randomness has been designed and fabricated in the 130 nm CMOS technology from STMicroelectronics known as HCMOS9GP in a standard variant with power supply voltage 1.2 V. This design is based on four independent sources of randomness. Therefore output random data r are composed of four independent raw random data streams while equation (1) changes into the form

$$\{r(l)\}_{l=0}^{\infty} = \left\{s_1(k),\ s_2\left(k+\frac{1}{4}\right),\ s_3\left(k+\frac{2}{4}\right),\ s_4\left(k+\frac{3}{4}\right)\right\}_{k=0}^{\infty} \tag{7}$$

because n is a number of independent random data streams and is equal to 4 in this case. Thus the normalized period l of output random data stream r is four times higher than the normalized period k of each raw random data stream s_n. And according to (2) the bit rate of output random data r is four times higher than the bit rate of used sources of randomness. If a source of randomness works with a maximal bit rate the bit rate of output random numbers can be increased by implementation of above-described principle.

A block diagram of the proposed TRNG is shown in Fig. 2. In presented design, four independent sources of randomness based on metastable states of electronic circuits

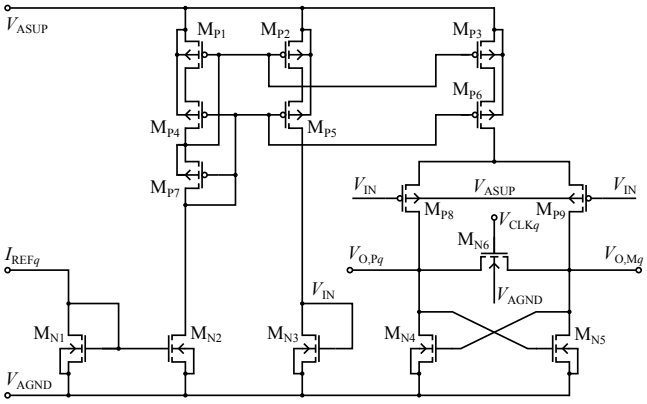

Fig. 3. Schematic diagram of the proposed metastability-based noise source.

are integrated and they generate digitized random signals V_{sq} where q is the identifier of the source of randomness and $q = 1, 2, 3, 4$. The source of randomness is composed of the noise source and the digitizer. Proposed noise sources use fine reference currents $I_{\mathrm{REF}q}$, which are generated in a reference current generator. All digitized random signals V_{sq} are mixed in a time multiplexer, which is together with all sources of randomness controlled by a clock signal generator where control signals are derived from the reference signal of external oscillator V_{OSC}. Output buffer processes output of time multiplexer and is able to drive external digital circuits or inputs of measuring instruments.

3.1 Noise Source

A noise source is a part of the presented circuit, which is able to extract randomness from a physical phenomenon appearing in silicon semiconductors. In this case the proposed noise source structure shown in Fig. 3 is based on a fast comparator, whose inputs are connected to the same voltage V_{IN} created on the transistor $\mathrm{M_{N3}}$. Thus a noise present in circuit decides on an output value. The noise source has to work periodically. Therefore the circuit is extended by a reset transistor $\mathrm{M_{N6}}$, which resets the proposed noise source and allows to generate a new random value each period. This is shown in Fig. 4 where a random value of the output signal $V_{\mathrm{O,M1}}$ is generated when the clock signal V_{CLK1} is in logic zero. The noise source is reset by the transistor $\mathrm{M_{N6}}$ when the clock signal V_{CLK1} is in logic one. Decision phase arises at the beginning of generating each random value when the circuit is in a metastable state. The noise present in the circuit causes a transition to a stable state. The final random value of the output signal $V_{\mathrm{O,M1}}$ is given by the noise in the circuit during decision phase.

Systematic errors can cause malfunction of the manufactured device. The noise source structure is proposed with maximum regard to symmetry, whose breach would adversely affect output random data, which would result in a deviation of the random data mean value. Therefore the important assumption is the conformity of dimensions of

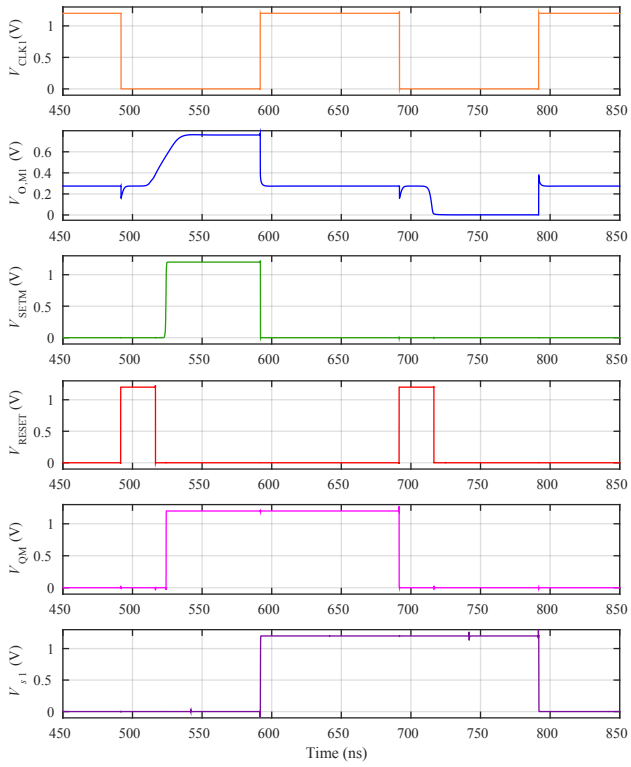

Fig. 4. Waveforms of signals inside the TRNG simulated by the Mentor Eldo simulator at the transistor level.

transistors $\mathrm{M_{P8}}$ and $\mathrm{M_{P9}}$ and also transistors $\mathrm{M_{N4}}$ and $\mathrm{M_{N5}}$. Not only the same dimensions of transistor pairs but also totally symmetric topology design including metal interconnections and well-matched elements are prerequisites for the properly functioning noise source. Hence the proposed noise source is composed of two equal branches, which differ only in small details namely vias among metal layers. To reduce appreciable mismatches among parameters of paired transistors, a common-centroid configuration is used so that first-order gradients are canceled [20], [21].

The noise source is designed to minimize the offset voltage, which is able to cause an undesirable distortion of output random data, such as a deviation of the mean value of output random data or even circuit locking in one logical value. Contribution to the offset voltage created by any difference between output voltages $V_{\mathrm{O,P}q}$ and $V_{\mathrm{O,M}q}$ is eliminated by proper circuit design and layout. Thus the offset voltage can be caused by a mismatch of MOSFET parameters such as a threshold voltage V_{TH} or a β parameter which is defined as

$$\beta = \mu C_{\mathrm{ox}} \frac{W}{L} \tag{8}$$

where μ is mobility of charge carriers in MOSFET, C_{ox} is the gate oxide capacitance per unit area, W is the gate width and L the gate length. The total offset voltage V_{OS} of the used circuit can be described by a sum of individual contributors and can be expressed as

$$V_{\mathrm{OS}} = \Delta V_{\mathrm{TH,P8,9}} + \Delta V_{\mathrm{TH,N4,5}} \sqrt{\frac{\mu_{\mathrm{N}} \frac{W_{\mathrm{N4,5}}}{L_{\mathrm{N4,5}}}}{\mu_{\mathrm{P}} \frac{W_{\mathrm{P8,9}}}{L_{\mathrm{P8,9}}}}} + \\ + \sqrt{\frac{I_{\mathrm{D}}}{2\mu_{\mathrm{P}} C_{\mathrm{ox}} \frac{W_{\mathrm{P8,9}}}{L_{\mathrm{P8,9}}}}} \left(\frac{\Delta\beta_{\mathrm{P8,9}}}{\beta_{\mathrm{P8,9}}} + \frac{\Delta\beta_{\mathrm{N4,5}}}{\beta_{\mathrm{N4,5}}} \right) \quad (9)$$

where $\Delta V_{\mathrm{TH,P8,9}}$ is a threshold voltage error between paired transistors $\mathrm{M_{P8}}$ and $\mathrm{M_{P9}}$ with the same widths $W_{\mathrm{P8,9}}$ and the same lengths $L_{\mathrm{P8,9}}$, $\Delta V_{\mathrm{TH,N4,5}}$ is the threshold voltage error between paired transistors $\mathrm{M_{N4}}$ and $\mathrm{M_{N5}}$ with the same widths $W_{\mathrm{N4,5}}$ and the same lengths $L_{\mathrm{N4,5}}$, μ_{N} is mobility of charge carriers in N-channel MOSFETs, μ_{P} is mobility of charge carriers in P-channel MOSFETs, $\frac{\Delta\beta_{\mathrm{P8,9}}}{\beta_{\mathrm{P8,9}}}$ is a normalized error of β parameter between paired transistors $\mathrm{M_{P8}}$ and $\mathrm{M_{P9}}$, $\frac{\Delta\beta_{\mathrm{N4,5}}}{\beta_{\mathrm{N4,5}}}$ is the normalized error of β parameter between paired transistors $\mathrm{M_{N4}}$ and $\mathrm{M_{N5}}$, and I_{D} is current flowing transistors $\mathrm{M_{P8}}$, $\mathrm{M_{N4}}$ or $\mathrm{M_{P9}}$, $\mathrm{M_{N5}}$. Thus according to the equation (9) for offset voltage minimization, the ratio $W_{\mathrm{P8,9}}/L_{\mathrm{P8,9}}$ should be maximized, the ratio $W_{\mathrm{N4,5}}/L_{\mathrm{N4,5}}$ minimized, and the bias current also minimized.

However, the proposed circuit has to be able to use non-deterministic noise as much as possible. In the CMOS technologies used for applications above-mentioned, thermal noise and flicker noise occur mainly [22]. Thermal noise is a noise with flat frequency spectrum so-called white noise, which can be modeled as an equivalent input noise voltage source of MOSFET [22] with the noise density

$$\frac{\mathrm{d}\overline{V_{\mathrm{n,th}}^2}}{\mathrm{d}f} = 4kT\gamma \frac{1}{g_{\mathrm{m}}} \quad (10)$$

where k is the Boltzmann constant, T is ambient temperature, g_{m} is MOSFET transconductance, and γ is a technology dependent coefficient, whose value depends on technology, often used value is 2/3, but increases for submicron CMOS technologies [20]. MOSFETs also exhibit flicker noise, which is noise with $1/f$ frequency spectrum and can be modeled as the equivalent input noise voltage source of MOSFET [22] with the noise density

$$\frac{\mathrm{d}\overline{V_{\mathrm{n,f}}^2}}{\mathrm{d}f} = \frac{K_{\mathrm{f}}}{WLC_{\mathrm{ox}}^2} \frac{1}{f} \quad (11)$$

where K_{f} is a coefficient with low technology dependence. Using above mentioned descriptions (10) and (11), the effect of noise in the circuit can be defined. Thus, the input-referred noise density is given by

$$\frac{\mathrm{d}\overline{V_{\mathrm{n,in}}^2}}{\mathrm{d}f} = 2 \left[4kT\gamma \frac{1}{g_{\mathrm{m,P8,9}}} \left(1 + \sqrt{\frac{\mu_{\mathrm{N}} \frac{W_{\mathrm{N4,5}}}{L_{\mathrm{N4,5}}}}{\mu_{\mathrm{P}} \frac{W_{\mathrm{P8,9}}}{L_{\mathrm{P8,9}}}}} \right) + \right. \\ \left. + \frac{K_{\mathrm{f,P}}}{W_{\mathrm{P8,9}} L_{\mathrm{P8,9}} C_{\mathrm{ox}}^2} \frac{1}{f} \left(1 + \frac{K_{\mathrm{f,N}}}{K_{\mathrm{f,P}}} \left(\frac{L_{\mathrm{P8,9}}}{L_{\mathrm{N4,5}}} \right)^2 \right) \right] \quad (12)$$

Influence	Recommendation	
Offset voltage ↓	$W_{\mathrm{P8,9}}/L_{\mathrm{P8,9}}$ ↑	$W_{\mathrm{N4,5}}/L_{\mathrm{N4,5}}$ ↓
Thermal noise ↑	$W_{\mathrm{P8,9}}/L_{\mathrm{P8,9}}$ ↓	$W_{\mathrm{N4,5}}/L_{\mathrm{N4,5}}$ ↑
Flicker noise ↑	$W_{\mathrm{P8,9}}L_{\mathrm{P8,9}}$ ↓	$L_{\mathrm{P8,9}}/L_{\mathrm{N4,5}}$ ↑

Tab. 1. Summary of found recommendations for the noise source proposal.

where $K_{\mathrm{f,P}}$ is the flicker noise coefficient for P-channel MOSFETs and $K_{\mathrm{f,N}}$ for N-channel MOSFETs. The first part of (12) represents thermal noise influence, which can be maximized by the ratio $W_{\mathrm{P8,9}}/L_{\mathrm{P8,9}}$ decreasing and the ratio $W_{\mathrm{N4,5}}/L_{\mathrm{N4,5}}$ increasing. Similarly, the second part of (12) describes flicker noise influence, which can be maximized by the product $W_{\mathrm{P8,9}}L_{\mathrm{P8,9}}$ decreasing and the ratio $L_{\mathrm{P8,9}}/L_{\mathrm{N4,5}}$ increasing.

Transistor dimensions of this most sensitive part are proposed in the light of the above considerations, which are summarized in Tab. 1. Naturally, dimensions of all used components are chosen to fit the designed circuit into a predefined area on a die. Recommendations for minimizing the offset voltage and prerequisites for maximizing thermal noise are contradictory. Therefore MOSFETs $\mathrm{M_{P8}}$, $\mathrm{M_{P9}}$, $\mathrm{M_{N4}}$, and $\mathrm{M_{N5}}$ have been proposed with the same dimensions. In order to maximize flicker noise influence, channel areas of transistors $\mathrm{M_{P8}}$ and $\mathrm{M_{P9}}$ have been minimized by shortening their lengths L_{P8} and L_{P9}. The proposed MOSFET dimensions have been fine-tuned by numerical simulations of the whole circuit, especially by a noise transient simulation in the Mentor Eldo simulator [23]. Thus the final transistors dimensions are $W_{\mathrm{P8,9}} = W_{\mathrm{N4,5}} = 5\ \mu\mathrm{m}$ and $L_{\mathrm{P8,9}} = L_{\mathrm{N4,5}} = 0.5\ \mu\mathrm{m}$.

The operating conditions of the circuit depicted in Fig. 3 are set by a reference current $I_{\mathrm{REF}q}$ of 1 μA. Current distribution inside the circuit is made by current mirrors. The current mirror composed of P-channel MOSFETs $\mathrm{M_{P1}}$, $\mathrm{M_{P2}}$, and $\mathrm{M_{P3}}$ is used in a cascode variant, which is created from transistors $\mathrm{M_{P4}}$, $\mathrm{M_{P5}}$, and $\mathrm{M_{P6}}$. A current branch generating the voltage V_{IN} for inputs of the differential pair copies the reference current in a ratio of 1 : 1. A ratio of the differential pair bias current created by MOSFET $\mathrm{M_{P3}}$ is proposed with regard to offset voltage minimization when according to (9) currents flowing transistors $\mathrm{M_{P8}}$ and $\mathrm{M_{P9}}$ should be decreased. The total current consumption of the whole circuit has to be also minimized due to minimization of the power consumption of the whole proposed TRNG. However, this bias current also has to be set with regard to a random data rate because higher data rate requires the higher bias current. The presented TRNG is designed for the output random data rate of 20 Mb/s. Therefore after verification by simulations, the current for the differential pair is set in the ratio 4 : 1 to the reference current $I_{\mathrm{REF}q}$.

The described TRNG is designed for use in complex ICs where a number of different parts work at the same time. In these ICs, deterministic disturbance arises and spreads across the chip, especially via the substrate and the power supply. This deterministic disturbance can occur in the output

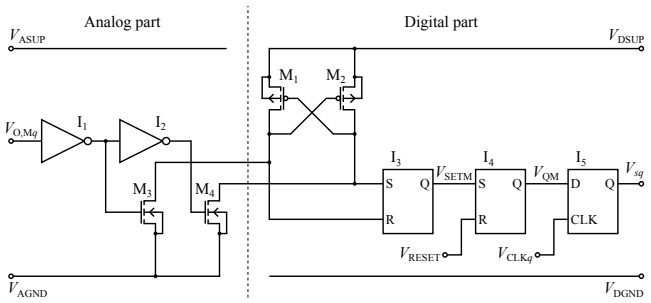

Fig. 5. Schematic diagram of the designed digitizer.

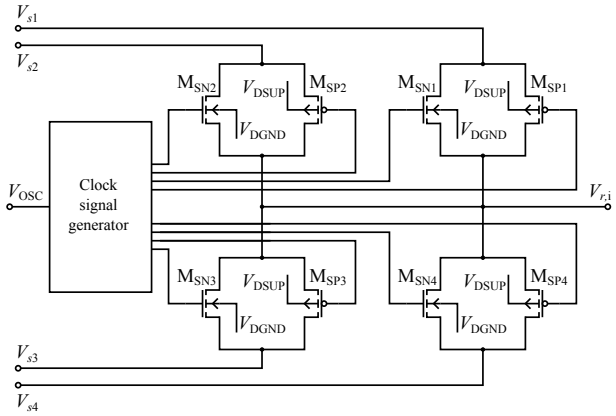

Fig. 6. Schematic diagram of the time multiplexer.

random data stream as a deterministic component. In other words, the output random data stream can be deliberately influenced and a system using some TRNG can be attacked in this way. Therefore systems with TRNGs must be resistant to this type of attack.

In presented design, the power supply and ground are divided into two separated parts. The first part is intended to supplying sensitive analog blocks, that is all noise sources. Moreover, the power supply and ground of each noise source are star-routed due to prevention of mutual distortion. Digital blocks of the proposed TRNG are supplied with the other part of the power supply and ground. These blocks operate with digital signals and cannot be easily affected by the described non-invasive attack. The presented generator does not show any deterministic components in the output random data stream during simulations of the power supply distortion as well as during on-chip verification. Nevertheless, to increase security, a low drop out regulator without any external capacitor integrated into the same die is suitable to use because there is no possibility to attack via external supply pin.

3.2 Digitizer

A block transferring random values from the analog part of the proposed TRNG to the digital part is marked as the digitizer and is shown in Fig. 5. The random signal $V_{\mathrm{O,M}q}$ in analog supply domain is transformed into the digital signal V_{sq} in digital supply domain. The random values have to be transferred without any damage and synchronized with rest of the system. This block separates both power supply domains as well. The sources of randomness shown in Fig. 2 are created by connection of the digitizer to the noise source.

The random values are transferred between both supply domains by differential signals, which are created by inverters I_1 and I_2. These standard cells also suitably shape the input signal $V_{\mathrm{O,M}q}$. A structure consisting of transistors M_1, M_2, M_3, M_4 and commonly used in level shifters perform the transmission itself and drives an RS latch I_3 composed of standard CMOS NOR gates. The large digital circuits in complex ICs usually create supply voltage distortion in form of voltage glitches, which can cause malfunctions of other circuits connected to the same supply voltage. The incorporated RS latch I_3 is able to hold a logic value during supply voltage fluctuations and thus increase system reliability.

The output signal V_{SETM} of the RS latch I_3 is still not suitable for further processing in the time multiplexor because a logic value of the signal cannot be reliably read in the next rising edge of the clock signal $V_{\mathrm{CLK}q}$ as it is shown in Fig. 4. Therefore the signal V_{RESET} is formed in the clock generator. Using this, a part of the signal V_{SETM} containing the random logic value is extended in the RS latch I_4 so that it can be synchronized with $V_{\mathrm{CLK}q}$ in the D latch I_5, which produces a further usable signal V_{sq}.

Slightly asymmetric or too large parasitic capacitances at the noise source outputs can cause the TRNG failure. The noise source can be locked in one logic value. Therefore from the layout point of view to maintain symmetry, the digitizer is connected to each noise source output but only one of them is used for next signal processing. The digitizer is also placed as near as possible to the noise source outputs due to parasitic capacitance reduction of interconnections $V_{\mathrm{O,M}q}$ and $V_{\mathrm{O,P}q}$. For the same reason, inverters with the smallest possible MOSFETs are used at the digitizer input.

3.3 Time Multiplexer

Each source of randomness is designed to generate the digital random signal V_{sq} with clock frequency 5 MHz. The time multiplexer depicted in Fig. 6 combines all four generated signals V_{sq} shown in Fig. 7 and forms the internal raw random signal $V_{r,\mathrm{i}}$ with four times higher clock frequency. Transmission gates are opened consecutively by signals $V_{\mathrm{SN}q}$ controlling N-channel MOSFETs $M_{\mathrm{SN}q}$ and their negations controlling P-channel MOSFETs $M_{\mathrm{SP}q}$. All control signals are generated in the clock generator and derived from the fast clock signal V_{OSC}.

More circuits switch on at the same time then peaks on the power supply can appear. The sources of randomness generate the digital random signals V_{sq} with a shift of one quarter the period $V_{\mathrm{CLK}q}$ among them. It reduces peak current consumption and thereby power supply distortion.

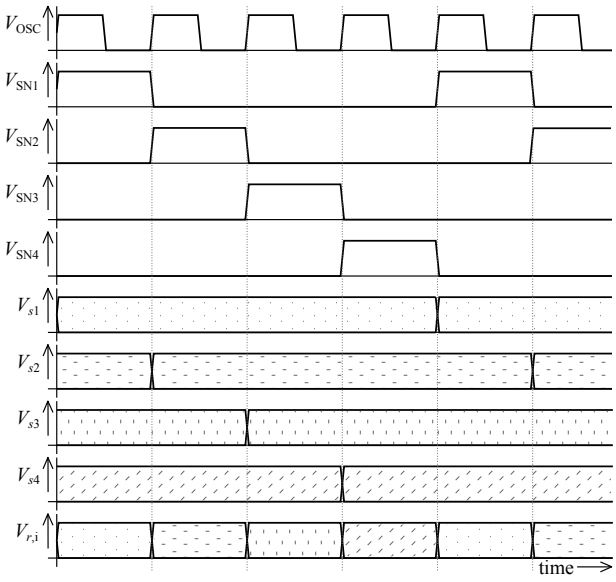

Fig. 7. Waveforms illustrating the function of the time multiplexer.

The internal raw random signal $V_{r,\mathrm{i}}$ does not have sufficient capability to drive external circuits or measuring instruments. Hence an output buffer increasing load-driving capability is incorporated into the proposed TRNG structure and produces the raw random signal V_r, which is accessible from a pin on the manufactured chip. The die photo is shown in Fig. 8. The designed device occupies an area of 0.029 mm^2 including all described parts - the noise sources, the digitizer, the time multiplexer, and the output buffer.

4. Measurement Results

The designed TRNG integrated on the fabricated chip has been characterized and validated both from the perspective of quality of generated random number sequences and from the perspective of integration into a system. All random sequences have been tested by statistical tests defined in the FIPS 140-2 standard [18] and by the stricter National Institute of Standards and Technology (NIST) tests [19]. These statistical test suites can reveal a bias, repeating patterns, or unbalanced distribution of random data.

The generator presented has been optimized for the random data rate of 20 Mb/s. However, the theoretical limit is slightly more than 60 Mb/s because the duration of the decision phase of the noise source output signal $V_{\mathrm{O,M}q}$ described in Sec. 3.1 does not depend on the clock frequency. While increasing the clock frequency, the duration of the stable state shortens but the duration of the decision phase does not change. For this reason, when increasing the clock frequency above the limit, the noise source is not able to move to the stable state and generate any random bits. In other words, when the output random data rate rises above the theoretical limit, there is not enough time after the decision phase for generation of a new random bit. Faultless FIPS and NIST tests have confirmed functionality of the designed TRNG at the output random data rate of 20 Mb/s and at the ambient

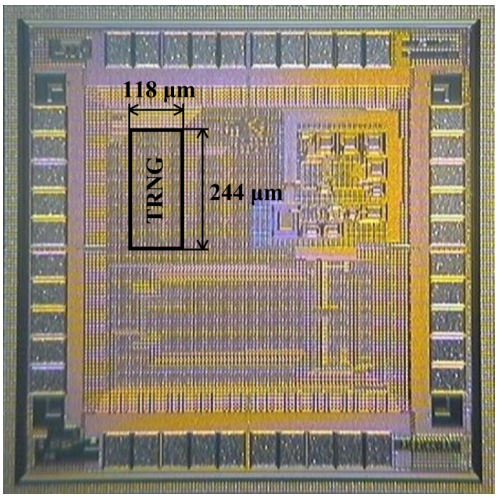

Fig. 8. Photo of the fabricated die. The proposed TRNG containing all described parts - the noise sources, the digitizer, the time multiplexer, and the output buffer - occupies the marked area.

temperature of 25 °C. Further at this temperature, the TRNG generated the output random data without failure even with the rate 60 Mb/s, which coincides with the assumed theoretical limit. These random data passed the FIPS tests but failed in some NIST tests as can be seen in Tab. 2.

As described above, deterministic disturbance of the power supply can affect the quality of generated random sequences. Hence the random sequences were generated with artificially created power supply distortion, which simulated power supply distortion with the voltage spike frequency in the range of 10 MHz to 100 MHz coming from digital circuits. Results of the FIPS and NIST tests are consistent with the results obtained from the output random data generated at the undistorted power supply. Of this comparison, it can be assumed that the above-described distortion of the power supply does not affect the output random data.

All 15 statistical NIST tests have been performed with a significance level α equal to 0.01. Thus each test has passed, if the computed *P-value* would have been equal or greater than 0.01. For NIST tests, 1 Mb long random sequences have been generated by the output random data rate from 10 Mb/s to 60 Mb/s in the temperature range of −35 °C to 85 °C. For FIPS tests, the random sequences have been shortened to the desired 20 kb.

From the results obtained, it is obvious that the proposed TRNG has worked well at the output random data rates 10 Mb/s and 20 Mb/s at all measured temperatures. However, at maximum data rates, the NIST tests have revealed periodic features in the output sequences, which has been caused by deterministic noise present in the circuit. At the highest measured temperatures and the highest data rates, an imbalance between zeros and ones has been observed, which has been demonstrated by unmet Frequency tests from both NIST and FIPS test suites. This bias has been caused by the occasional locking of the generator in one logic value. All results of NIST and FIPS tests at the ambient temperature 25 °C are listed in Tab. 2.

		Results at output random data rates					
		10 Mb/s	20 Mb/s	30 Mb/s	40 Mb/s	50 Mb/s	60 Mb/s
NIST test	Monobit	PASSED	PASSED	PASSED	PASSED	PASSED	PASSED
	Frequency	PASSED	PASSED	PASSED	PASSED	PASSED	PASSED
	Runs	PASSED	PASSED	PASSED	PASSED	FAILED	FAILED
	Longest runs	PASSED	PASSED	PASSED	PASSED	PASSED	PASSED
	Binary matrix rank	PASSED	PASSED	PASSED	PASSED	PASSED	PASSED
	Spectral DFT	PASSED	PASSED	PASSED	PASSED	FAILED	FAILED
	Non-overlapping template	PASSED	PASSED	FAILED	FAILED	FAILED	FAILED
	Overlapping template	PASSED	PASSED	PASSED	PASSED	PASSED	FAILED
	Universal statistical	PASSED	PASSED	FAILED	FAILED	FAILED	FAILED
	Linear complexity	PASSED	PASSED	PASSED	PASSED	PASSED	PASSED
	Serial	PASSED	PASSED	PASSED	PASSED	FAILED	PASSED
	Approximate entropy	PASSED	PASSED	PASSED	PASSED	PASSED	PASSED
	Cumulative sums	PASSED	PASSED	PASSED	FAILED	PASSED	PASSED
	Random excursions	PASSED	PASSED	PASSED	FAILED	FAILED	FAILED
	Random excursions variant	PASSED	PASSED	PASSED	FAILED	FAILED	FAILED
FIPS test	Monobit	PASSED	PASSED	PASSED	PASSED	PASSED	PASSED
	Poker	PASSED	PASSED	PASSED	PASSED	PASSED	PASSED
	Runs	PASSED	PASSED	PASSED	PASSED	PASSED	PASSED
	Long run	PASSED	PASSED	PASSED	PASSED	PASSED	PASSED

Tab. 2. Results of the NIST and FIPS tests at the ambient temperature 25 °C.

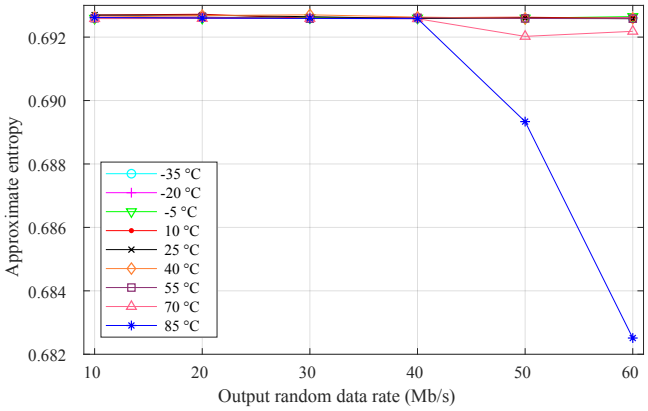

Fig. 9. The approximate entropy of output random number sequences generated in the temperature range of −35 °C to 85 °C.

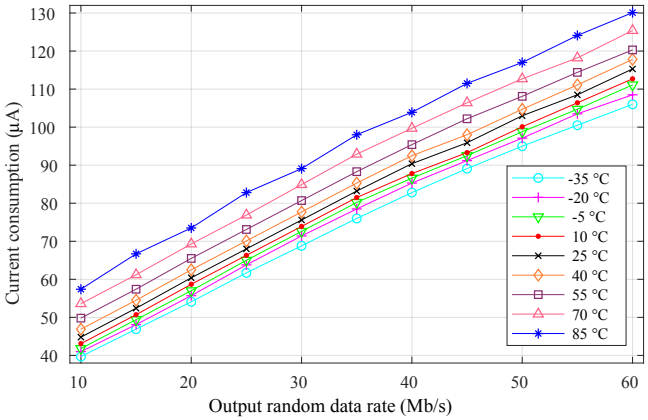

Fig. 10. The current consumption of the designed circuit measured in the temperature range of −35 °C to 85 °C.

The approximate entropy (ApEn) defined in [24] is used to determine the amount of regularity and the unpredictability of fluctuations in random bit streams. Thus small values of ApEn imply strong regularity or persistence, and large values of ApEn imply substantial fluctuation or irregularity [24]. The ApEn calculation is a part of the NIST test suite. The ApEn values of all sequences are shown in Fig. 9. The approximate entropy of sequences generated with the lower data rates is high and close to the maximum value of the natural logarithm of 2. However, the ApEn values of sequences generated with the higher data rates are lower. Even the ApEn values at high temperatures and data rates are very low, which proves the presence of the already mentioned bias in these data sequences.

This TRNG is designed for use within complex SoCs. From the point of view of the whole system therefore, it is necessary to know the consumption in defined operation points. Thus the current consumption was also measured in the temperature range of −35 °C to 85 °C for the output random data rate from 10 Mb/s to 60 Mb/s and results are in Fig. 10. At the ambient temperature of 25 °C and at the output random data rate of 20 Mb/s, the current consumption is 60.4 μA.

5. Discussion

Based on measurements and statistical test results, it might be said that the fabricated TRNG can be integrated inside SoCs for generation of random bit sequences with the bit rate up to 20 Mb/s at above-mentioned temperatures without incorporating any corrector. When used at higher random data rates, the TRNG should be supplemented with any above-mentioned corrector to improve the properties of random sequences. However, use of the generator at the highest random data rate is not appropriate because the results of both statistical test suites at the highest temperatures show failures.

The designed TRNG has been fabricated in the 130 nm bulk CMOS technology as a part of a multi-project test chip where occupies the area of 0.029 mm^2 and consumes 72.48 μW at the ambient temperature of 25 °C and the output random data rate of 20 Mb/s. A direct comparison can be done with the TRNG published in [10] and fabricated by the similar 130 nm bulk CMOS process. This generator occupies a larger area of 0.145 mm^2, has a higher power consumption of 1 mW, and a lower random data rate of 200 kb/s. Both TRNGs resist temperature and power supply variations.

Very high random data rates have been achieved by TRNGs published in [16], [11] which have been fabricated in advanced node CMOS processes optimized for higher clock frequencies of digital circuits. They have a much higher power consumption and that can be an obstacle for use in some applications such as hand-held devices due to energy saving.

6. Conclusion

This paper presents the true random number generator with time multiplexed metastability-based sources of randomness. Based on the principle described, the TRNG has been designed and fabricated in the 130 nm HCMOS9GP bulk CMOS technology from STMicroelectronics. All four noise sources extracting randomness from thermal and flicker noise present in CMOS circuits have been proposed to operate in changing environmental conditions. The time multiplexer expanding structure of the TRNG combines all digitized random signals and increases the output random data rate. Properties of random bit sequences have been verified by the FIPS and NIST statistical test suites. Their results have shown that the mentioned generator can operate without any correctors at the output data rate up to 20 Mb/s in temperature and power supply variations. With these achieved parameters, the presented TRNG is suitable for being integrated into noisy and low power environment of SoC. Because the TRNG does not contain passive analog devices it simplifies its migration to advanced process nodes to increase TRNG random data rate and decrease the power consumption even more.

Acknowledgments

This work has been supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS17/188/OHK3/3T/13 (Mikro a nanostruktury a soucastky).

References

[1] MENEZES, A. J., VAN OORSCHOT, P. C., VANSTONE, S. A. *Handbook of Applied Cryptography*. Boca Raton (USA): CRC Press, 1996. ISBN: 0849385237

[2] SCHINDLER, W., KILLMANN, W. Evaluation criteria for true (physical) random number generators used in cryptographic applications. In *Proceedings of the nternational Workshop on Cryptographic Hardware and Embedded Systems (CHES)*. Berlin, Heidelberg (Germany), 2003, p. 431–449. DOI: 10.1007/3-540-36400-5_31

[3] TANG, X., WU, Z.-M., WU, J.-G., et al. Tbits/s physical random bit generation based on mutually coupled semiconductor laser chaotic entropy source. *Optics Express*, 2015, vol. 23, no. 26, p. 33130–33141. DOI: 10.1364/OE.23.033130

[4] SUGIURA, T., YAMANASHI, Y., YOSHIKAWA, N. Demonstration of 30 Gbit/s generation of superconductive true random number generator. *IEEE Transactions on Applied Superconductivity*, 2011, vol. 21, no. 3, p. 843–846. DOI: 10.1109/TASC.2010.2092401

[5] ALKASSAR, A., NICOLAY, T., ROHE, M. Obtaining true-random binary numbers from a weak radioactive source. In *Proceedings of the International Conference on Computational Science and Its Applications (ICCSA)*. Singapore, 2005, p. 634–646. DOI: 10.1007/11424826_67

[6] OOSAWA, S., KONISHI, T., ONIZAWA, N., HANYU, T. Design of an STT-MTJ based true random number generator using digitally controlled probability-locked loop. In *Proceedings of the IEEE 13th International New Circuits and Systems Conference (NEWCAS)*. Grenoble (France), 2015, p. 1–4. DOI: 10.1109/NEWCAS.2015.7182089

[7] HOLMAN, W. T., CONNELLY, J. A., DOWLATABADI, A. B. An integrated analog/digital random noise source. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 1997, vol. 44, no. 6, p. 521–528. DOI: 10.1109/81.586025

[8] YANG, K., FICK, D., HENRY, M. B., et al. 16.3 A 23Mb/s 23pJ/b fully synthesized true-random-number generator in 28nm and 65nm CMOS. In *Proceedings of the IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*. San Francisco (USA), 2014, p. 280–281. DOI: 10.1109/ISSCC.2014.6757434

[9] TANG, Q., KIM, B., LAO, Y., et al. True random number generator circuits based on single- and multi-phase beat frequency detection. In *Proceedings of the IEEE Custom Integrated Circuits Conference*. San Jose (USA), 2014, p. 1–4. DOI: 10.1109/CICC.2014.6946136

[10] TOKUNAGA, C., BLAAUW, D., MUDGE, T. True random number generator with a metastability-based quality control. *IEEE Journal of Solid-State Circuits*, 2008, vol. 43, no. 1, p. 78–85. DOI: 10.1109/JSSC.2007.910965

[11] SRINIVASAN, S., MATHEW, S., RAMANARAYANAN, R., et al. 2.4GHz 7mW all-digital PVT-variation tolerant true random number generator in 45nm CMOS. In *Proceedings of the Symposium on VLSI Circuits*. Honolulu (USA), 2010, p. 203–204. DOI: 10.1109/VLSIC.2010.5560296

[12] PETRIE, C. S., CONNELLY, J. A. A noise-based IC random number generator for applications in cryptography. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 2000, vol. 47, no. 5, p. 615–621. DOI: 10.1109/81.84786

[13] BUCCI, M., GERMANI, L., LUZZI, R., et al. A high-speed oscillator-based truly random number source for cryptographic applications on a smart card IC. *IEEE Transactions on Computers*, 2003, vol. 52, no. 4, p. 403–409. DOI: 10.1109/TC.2003.1190581

[14] CICEK, I., PUSANE, A. E., DUNDAR, G. A novel design method for discrete time chaos based true random number generators. *Integration, the VLSI Journal*, 2014, vol. 47, no. 1, p. 38–47. DOI: 10.1016/j.vlsi.2013.06.003

[15] KOYUNCU, I., OZCERIT, A. T. The design and realization of a new high speed FPGA-based chaotic true random number generator. *Computers & Electrical Engineering*, 2017, vol. 58, p. 203–214. DOI: 10.1016/j.compeleceng.2016.07.005

[16] SURESH, V. B., BURLESON, W. P. Robust metastability-based TRNG design in nanometer CMOS with sub-vdd pre-charge and hybrid self-calibration. In *Proceedings of the Thirteenth International Symposium on Quality Electronic Design (ISQED)*. Santa Clara (USA), 2012, p. 298–305. DOI: 10.1109/ISQED.2012.6187509

[17] KINNIMENT, D. J., CHESTER, E. G. Design of an on-chip random number generator using metastability. In *Proceedings of the 28th European Solid-State Circuits Conference*. Florence (Italy), 2002, p. 595–598.

[18] Federal Information Processing Standards, National Institute of Standards and Technology. *Security Requirements for Cryptographic Modules*. 2001. NIST FIPS PUB 140-2. 69 pages. [Online] Cited 2018-01-11. Available at: http://csrc.nist.gov/publications/fips/fips140-2/fips1402.pdf

[19] RUKHIN, A., et al. *A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications*. 2010. National Institute of Standards and Technology. Rev 1a. 131 pages. [Online] Cited 2018-01-11. Available at: http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-22r1a.pdf

[20] RAZAVI, B. *Design of Analog CMOS Integrated Circuits*. New York (USA): McGraw-Hill, 2001. ISBN: 0071188150

[21] HASTINGS, A. *The Art of Analog Layout*. Upper Saddle River (USA): Prentice-Hall, 2001. ISBN: 0130870617

[22] SANSEN, W. M. C. *Analog Design Essentials*. Boston (USA): Springer, 2006. ISBN: 9780387257464

[23] Mentor Graphics Corporation. *Eldo User's Manual*. Wilsonville (USA), 2006.

[24] PINCUS, S., SINGER, B. H. Randomness and degrees of irregularity. In *Proceedings of the National Academy of Sciences of the United States of America*, 1996, vol. 93, no. 5, p. 2083–2088.

About the Authors . . .

Vlastimil KOTE, born in 1987, received his Bachelor's degree in Electrical Engineering from Czech Technical University in Prague in 2009, and Master's degree in Electrical Engineering from CTU in Prague in 2011. Currently, he works at STMicroelectronics as IC Layout Staff Engineer. He attends the Ph.D. study in the Department of Microelectronics at CTU in Prague where his research interests include structures of semiconductor devices, physical design, electrical circuit theory, and IC design methodology development.

Patrik VACULA was born in 1979. Master's degree in Electronics and Multimedia Communications Engineering, he completed at TU FEI in Kosice in 2003. Currently he works at STMicroelectronics as member of Technical Staff IC layout engineer responsible for entire BE IC development flow including power MOS integration. He is a Ph.D. student in the Department of Microelectronics at CTU.

Vladimir MOLATA received the Bachelor's degree in Electrical Engineering from Czech Technical University in Prague, in 2009 and the Master's degree in Electrical Engineering from Czech Technical University in Prague, in 2011. He is currently studying Ph.D. programme at the Czech Technical University where his research interests include electrical circuits theory, modelling and simulation. He works at STMicroelectronics as IC Design Senior Engineer of CMOS/BCD analog ICs for power management.

Ondrej VESELY graduated at CTU in Prague in 2011 with Master's degree in Electrical Engineering. Currently he works in STMicroelectronics in Prague as Senior validation engineer with specialization on measurement and automatic parametric validation. He is responsible for product testing during development stage.

Ondrej TLASKAL, born in Czech Republic, started his career in Racal Research Ltd, Reading, UK, where he worked on mixed-signal devices and systems, especially Sigma-Delta modulators for digital radio receivers. He received a Ph.D degree from the University of Surrey, UK, in the same subject. After several years in Silicon & Software Systems, which was a Philips Semiconductors company at that time, he joined STMicroelectronics where he is today leading a team developing power management products for data storage applications, specializing in DC/DC converters.

Dalibor BARRI, born in 1982 received his Bachelor's and Master's degreein Electronics from the Czech Technical University (CTU), Prague, in 2005 and 2007, respectively. He has worked in the company STMicroelectronics as an analog IC back-end designer. He is a Ph.D. student, and his topic of the thesis is to invent a novel tool for an automatic or semi-automatic layout of the analog integrated circuits.

Jiri JAKOVENKO received the Ph.D. degree in Microelectronics from the Czech Technical University in Prague, Faculty of Electrical Engineering CTU FEE in 2004. He works as Associate Professor at the Department of Microelectronics and vice-dean for education at CTU FEE. He is a member of Microsystems group. His research activities include analog integrated circuit design, MEMS design and reliability modeling. Since 2004 he is a leader of IC and MEMS design laboratory at CTU FEE. He is the author and co-author of more than 50 scientific publications, co-author of a chapter in Springer book; more than 30 publications are registered in WoS. He is a member of IMAPS EDS scientific committee, reviewer for scientific journals as *Microelectronics Reliability, Electron Device Letters, Radioengineering*, etc.

Miroslav HUSAK received his Ph.D. in 1984 from the Czech Technical University in Prague (CTU), branch Radioelectronics. He works as full professor in Electronics and medical engineering branch at the Department of Microelectronics of CTU in Prague (2000) and the Head of the Department of Microelectronics (from 1997). He is the head of the "Microsystems & integrated circuits" group. Research activities: Design and applications of microsystems, sensors, energy harvesting, electronic devices and their application in electronic instruments as well as diagnostics. He was the Applicant of 19 national grants, investigator of 4 European grants (research, 6th European Framework, 7.FP EU, ENIAC, NATO for Peace, Horizont 2020). Publications: Author of 1 monograph, 6 textbooks, more than 290 specialized publications in scientific journals, conference proceedings in the area of microsensors and microsystems. He is the author and co-author of more than 50 papers registered in WoS. He is a member of IEEE.