doi:10.11887/j.issn.1001-2486.25040001 http://journal.nudt.edu.cn

面向加速粒子输运随机模拟的概率可调真随机数生成器

傅思清,黎铁军*,吴利舟,张春元,马 胜,张建民,任睿轩
(国防科技大学 计算机学院,湖南 长沙 410073)

摘 要:粒子输运问题的随机模拟在传统冯·诺依曼架构上面临随机事件分支和不规则访存带来的挑战,其根源在于随机算法与确定性硬件之间的不匹配。为此,设计了一种基于自旋和铁电器件的概率可调真随机数生成器。基于自旋器件的物理随机性,为架构提供物理随机源,并通过优化的控制逻辑和写入机制提高随机比特吞吐率;基于铁电器件的忆阻特性,设计了可编程和具有非易失连续存储权重的概率可调突触。实验表明,该设计求解示例输运问题时性能相比通用处理器提高171~1 028倍。进一步地,相较现有的基于自旋转移矩磁隧道结的真随机数生成器,其不仅唯一具有生成可调概率随机采样的能力,且产生均匀分布随机序列时吞吐率达到303 Mbit/s,具有更高的随机比特吞吐率。

关键词:粒子输运;磁隧道结;铁电隧道结;真随机数生成器;概率计算
中图分类号:TP331.1 **文献标志码**:A **文章编号**:1001-2486(2025)06-036-10

论文拓展

Probability tunable random number generator for random simulation of accelerated particle transport

FU Siqing, LI Tiejun, WU Lizhou, ZHANG Chunyuan, MA Sheng, ZHANG Jianmin, REN Ruixuan*
(College of Computer Science and Technology, National University of Defense Technology, Changsha 410073, China)

Abstract: Particle transport simulations using stochastic methods face significant challenges on conventional von Neumann architectures, particularly due to random branching events and irregular memory access patterns. These limitations stem from the fundamental mismatch between probabilistic algorithms and deterministic computing paradigms. To bridge the gap between architecture and algorithms, a probabilistically tunable true random number generator was developed based on spintronic and ferroelectric devices. The physical randomness of spintronic devices was leveraged to provide a physical random source for the architecture, and the throughput of random bits was enhanced through optimized control logic and writing mechanisms. Next, programmable synapses were designed based on the memristive properties of ferroelectric devices, enabling non-volatile continuous weight storage with tunable probabilities. The experimental results indicate that the proposed approach achieves performance improvements ranging from 171 to 1 028 times compared to a general-purpose CPU when solving a sample transport problem. Furthermore, compared to existing spin-transfer torque magnetic tunnel junction based true random number generators, the developed method not only enables tunable probability random sampling but also achieves a throughput of 303 Mbit/s when generating uniformly distributed random sequences.

Keywords: particle transport; magnetic tunnel junction; ferroelectric tunnel junction; true random number generator; probabilistic computing

粒子输运问题的广泛应用涵盖从原子能物理学和天体物理学到材料科学和放射医学工程等多个学科领域,是现代科学与工程计算的重要组成部分[1]。由于输运问题的非线性、高维度等特点,随机模拟即蒙特卡罗(Monte Carlo, MC)方法是在高性能计算(high performance computing, HPC)系统上求解粒子输运问题时使用最广泛的方法[2]。

然而,传统的基于冯·诺依曼架构的HPC系统在执行MC模拟时效率低下。MC粒子输运程序中跟踪粒子过程的随机分支,引起了粒子截面数据的不规则访存,这是在HPC上求解粒子输运问题的主要瓶颈[3]。而从架构上看,指令控制的确定性程序执行以及存算分离设计直接导致上述问

收稿日期:2025-04-01
基金项目:国家自然科学基金资助项目(62304257,62472435,62172430);湖南省科技创新计划资助项目(2022RC3066,2022RC3065)
第一作者:傅思清(1996—),男,湖北襄阳人,博士研究生,E-mail:fusiqingnudt@nudt.edu.cn
*通信作者:黎铁军(1977—),男,湖南邵阳人,研究员,博士,博士生导师,E-mail:tjli@nudt.edu.cn
引用格式:傅思清,黎铁军,吴利舟,等. 面向加速粒子输运随机模拟的概率可调真随机数生成器[J]. 国防科技大学学报,2025,47(6):36-45.
Citation: FU S Q, LI T J, WU L Z, et al. Probability tunable random number generator for random simulation of accelerated particle transport[J]. Journal of National University of Defense Technology, 2025, 47(6): 36-45.

题,归根到底是确定性计算架构与随机模拟算法不匹配带来的问题。近十年来众核处理器和图形处理器(graphics processing unit, GPU)的进步,使 MC 方法的潜在并行性得到利用,计算效率和可扩展性得到了显著提升[4-5],但其核心挑战,也就是用确定性计算架构执行非确定随机算法带来的不确定分支、访存开销问题仍然没有得到本质解决[6]。

在架构层面,超越冯·诺依曼架构的随机模拟求解加速早期以现场可编程门阵列(field-programmable gate array, FPGA)为主,但片上资源限制和硬件伪随机特点限制其实际推广。近十年来,神经形态计算用于求解科学计算问题的前景已开始得到研究[7-10]。Smith 等[9]利用 Loihi 和 TrueNorth 芯片上的随机游走求解包括粒子输运问题在内的多个领域的随机模拟问题,评估了在神经形态架构上求解 MC 问题的显著功耗优势。但由于使用的计算平台仍然是基于互补金属氧化物半导体(complementary metal-oxide-semiconductor, CMOS)的确定性架构,并没有获得性能优势。随着材料科学的进步,具有物理随机性的新兴器件的出现为开发随机计算架构提供了机会[11]。磁隧道结(magnetic tunnel junction, MTJ)具有两种非易失自旋状态,并可通过自旋转移矩[12](spin-transfer torque, STT)和自旋轨道矩[13](spin orbit torque, SOT)等方式驱动状态翻转。由于热扰动的存在,MTJ 的动态翻转过程存在物理随机性,目前基于自旋器件设计的真随机数生成器(true random number generator, TRNG)已经得到广泛研究,但大都仅限于生成均匀分布随机序列[14-20]。

为进一步耦合硬件与随机算法,设计了用于加速粒子输运随机模拟的自旋电子学体系结构。首先,利用铁电隧道结(ferroelectric tunnel junction, FTJ)的非易失存储特性实现输出概率调控,通过写电压幅值精确控制 MTJ 翻转概率,避免了传统均匀采样 TRNG 所需的后处理开销。然后,设计了 TRNG 的双向概率读写和反馈机制,优化随机比特生成周期,同时延长器件寿命。最后,以一个粒子输运问题为例,评估 TRNG 加速粒子输运模拟的能力。

1　随机模拟方法与新兴器件原理

1.1　粒子输运的随机游走求解方法

玻耳兹曼(Bolzmann)方程描述高能粒子在介质中的输运过程。定义 t 时处于位置 x,运动方向 Ω 的粒子角通量密度 $\Phi:=\Phi(t,x,\Omega)$ 满足玻耳兹曼粒子输运方程:

$$\frac{1}{v}\frac{\partial}{\partial t}\Phi(t,x,\Omega)+\Omega\frac{\partial}{\partial x}\Phi(t,x,\Omega)+\Sigma_t(x,\Omega)\Phi(t,x,\Omega)=\int\Phi(t,x,\Omega')\sigma_s(x,\Omega'\to\Omega)\mathrm{d}\Omega'+R(t,x,\Omega)\quad(1)$$

式中:

$$\Sigma_t(x,\Omega)=\Sigma_a(x,\Omega)+\Sigma_s(x,\Omega)\quad(2)$$

其中,Σ_a 和 Σ_s 分别为粒子吸收和散射速率的函数,R 为粒子源项,v 为粒子速度,$\sigma_s(x,\Omega'\to\Omega)$ 为微分散射截面。

解输运方程的随机游走方法[21]通过 Feynman-Kac 公式[22]建立偏微分方程(3)与随机微分方程(4)之间的联系。

$$\frac{\partial u}{\partial t}+\mu(x,t)\frac{\partial u}{\partial x}+\frac{1}{2}\sigma^2(x,t)\frac{\partial^2 u}{\partial x^2}+f(x,t)u=0\quad(3)$$

式中,$u=u(x,t)$ 是要解的函数,$\mu(x,t)$ 和 $\sigma(x,t)$ 是漂移和扩散项,$f(x,t)$ 是一个给定的函数。

对应的随机微分方程如方程(4)所示:

$$u(x,t)=E[\phi(X_t)\mid X_t=x]\quad(4)$$

式中,X_t 表示从时间 t 开始的布朗运动,条件期望表示终止条件下对所有可能随机轨迹的数学平均。用于执行随机游走的马尔可夫链如图 1 所示,定义一个包含 N 个离散位置的有限一维空间,间距为 Δx,粒子状态空间为 $S=\{X_1,X_2,\cdots,X_N\}$。$P_{i,j}(j=i-1,i,i+1)$ 规定了从状态 i 转移到 j 的概率:

$$\begin{cases}P_{i,i+1}=P_{i,i-1}=P_g\\P_{i,i}=P_s\end{cases}\quad(5)$$

通过在马尔可夫链上执行随机游走,模拟随机微分方程,可以得到近似方程(3)的解。

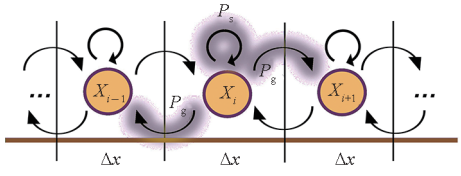


图 1　用于执行随机游走的马尔可夫链

Fig. 1　Markov chain for random walks

1.2　自旋磁隧道结和铁电隧道结

MTJ 作为一种新兴自旋电子学器件,已经开始得到广泛研究[20, 23]。MTJ 具有固定层(pinned layer, PL)、隧穿势垒(tunnel barrier, TB)层以及自由层(free layer, FL)结构。PL 磁化方向固定,而 FL 磁化方向可以与 PL 的磁化方向平行(parallel, P)或反平行(anti-parallel, AP),分别代表逻辑“0”和“1”。图 2 描述了热扰动下 STT 翻转及其动态概率翻转过程。在 AP 状态下,自旋

极化电流 $I_{AP\to P}$ 将电子从固定层导向自由层，进入自由层的极化电子的磁矩诱导其磁化翻转。同样，在 P 状态下，自旋极化电流 $I_{P\to AP}$ 诱导反向的磁化翻转。在适当的电流幅值和作用时间下，将由随机动态翻转主导翻转结果。在幅度为 I、持续时间为 t 的偏置电流下，MTJ 的翻转概率为：

$$\begin{cases} P(I,t) = 1 - \exp\left(-\dfrac{t}{\tau}\right) \\ \tau = \tau_0 \exp\left[\Delta\left(1 - \dfrac{I}{I_{c0}}\right)^2\right] \end{cases} \tag{6}$$

其中，τ 代表平均翻转时间，τ_0 是尝试时间因子，Δ 是温度相关的热稳定因子，I_{c0} 是在 0 K 时的临界翻转电流。

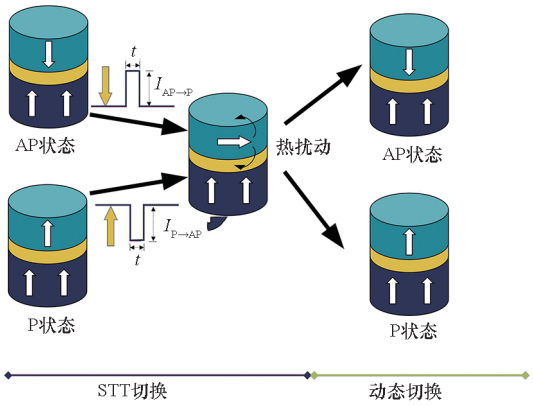


图 2 STT 翻转及其动态概率翻转过程

Fig. 2 STT switching and dynamic probability switching

最近的研究在 FTJ 中发现了忆阻行为，铁电极化翻转可以引起电阻的连续变化[24]。典型的 FTJ 结构由两个电极层和它们之间的超薄铁电势垒层组成，如图 3 中由 Co、$BaTiO_3$（BTO）、LSMO 组成的 FTJ。施加电压可以诱导 BTO 铁电层的极化翻转，改变电子的隧穿概率，从而导致电阻发生变化。在完全向上极化的铁电薄膜中，施加正向编程电压会引发向下极化畴的成核和拓展，进而导致连续的电阻变化，即图 3 中从左向右的过程。反向电压则驱动向上极化畴的成核与拓展。利用 FTJ 的忆阻特性，可以将其编程到稳定阻值，从而实现非易失且连续编码 TRNG 写入电压，以实现 TRNG 输出概率可调。需要说明的是，极化阻值可能受到温度以及势垒层尺寸等参数影响，但本文重点考虑自旋切换过程，因此通过对 MTJ 引入电压偏差来考虑 FTJ 工艺偏差。

2 电路设计

所提出的设计将粒子输运求解过程的粒子跟踪阶段卸载到硬件 TRNG 电路，以避免 CPU 体系结构在随机跟踪过程中产生分支延迟和访存瓶颈。系统框架如图 4 所示。

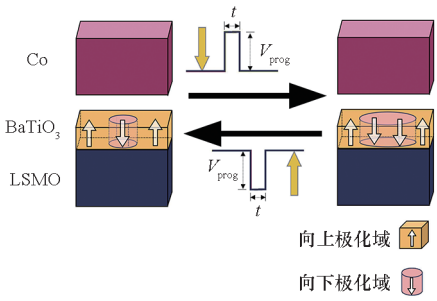


图 3 编程电压诱导下 FTJ 极化畴成核和拓展的过程

Fig. 3 Domain nucleation and propagation in FTJ under programming voltage

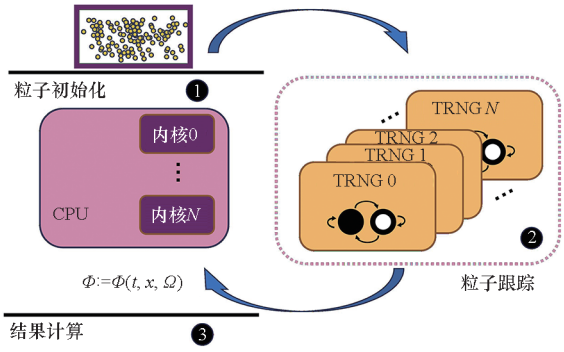


图 4 求解粒子输运的系统框架

Fig. 4 Computational system frameworks for particle transport

2.1 总体架构设计

基于 MTJ 的 TRNG 的基本控制逻辑[14-17, 19]如图 5 所示。每个控制周期首先重置 MTJ 状态，然后进行随机写入，最后读出写入的随机比特。相比读延迟，自旋器件具有更高的写延迟。因此控制周期的两次写入操作限制了随机比特生成速度。所提出的设计如图 6 所示，MTJ 的初始状态首先被读出，随后将反相值反馈作为写驱动器的写入值。在第二阶段，写驱动器执行随机写入操作。这让 TRNG 在每个周期只执行一次写入操作，而不需进行重置，缩短了控制周期。

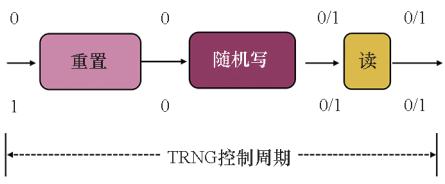


图 5 基于 MTJ 的 TRNG 的控制逻辑

Fig. 5 Control logic for MTJ-based TRNG

电路架构设计如图 7 所示。当 Rd 信号被启用时，感测放大器[25]（sense amplifier，SA）读出

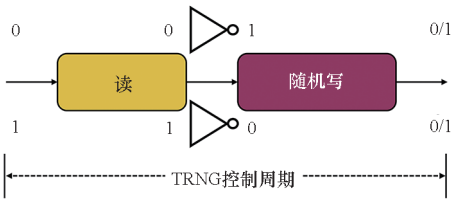


图 6　可调概率 TRNG 的控制逻辑

Fig. 6　Control logic for tunable probability TRNG

MTJ 的逻辑状态(标记为 Out)及其反相值$\overline{\text{Out}}$。随后,$\overline{\text{Out}}$信号作为输入 Data_in 反馈给写驱动器[20]。当 Wr 信号被触发时,写驱动器将 Data_in 写入 MTJ。对于固定的写入周期,MTJ 的翻转概率由电流幅值控制,即由概率调节突触 S_1 和 S_2 保存的权值控制。

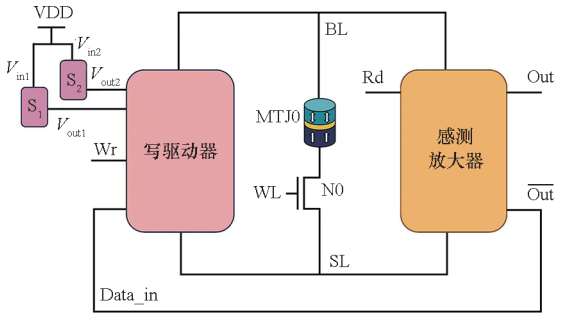


图 7　可调概率 TRNG 的电路设计

Fig. 7　Circuit design for a tunable probability TRNG

2.2　概率调节突触设计

图 8 展示了基于 FTJ 的概率调节突触结构。突触有编程模式和写驱动模式。在编程模式下,当晶体管 P_0 的栅极信号 V_{0g} 置为低电平时,晶体管开启,从而允许编程电压 V_{prog} 被施加到 FTJ 上。V_{in} 表示概率调节突触输入电压。在写驱动模式下,当晶体管 P_1 的栅极信号 V_{1g} 置为低电平时,晶体管开启,输出电压 V_{out} 根据当前 FTJ 的电阻值 R_{FTJ} 进行调节,如式(7)所示。

$$V_{out}=\frac{R_{FTJ}}{R_{FTJ}+R}\cdot V_{in} \tag{7}$$

较低的写驱动电压不会改变铁电薄膜的极化状态,这使得它能够作为突触稳定的保留权重。

2.3　PVT 容忍设计

图 7 中写驱动电路由两组突触后点位分别控制两个方向写入电流的幅值。当写控制信号 Wr 被启用并且 Data_in 上的写数据为“1”时,激发从位线(bit line, BL)流向源线(source line, SL)的写入电流。电流的幅度和持续时间决定了 MTJ 被设置为状态“1”的概率,在固定写入时间下,概率由 V_{out1} 控制也就是突触 S_1 的 FTJ 阻值控制。同样,当 Data_in 为“0”时,电路将输出相反方向的写入电流。

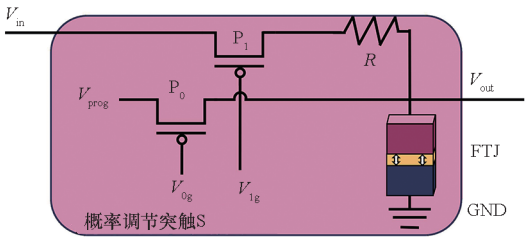


图 8　基于 FTJ 的概率调节突触结构

Fig. 8　Structure of tunable probability synapse based on FTJ

下面考虑对工艺 - 电压 - 温度(process-voltage-temperature, PVT)偏差的容忍情况。MTJ 器件初始可能处于 P 状态(概率为 P_P)或 AP 状态(概率为 P_{AP})。当 MTJ 处于 P 状态时,设在写入电流作用下翻转到 AP 状态的概率为 P_1。当 MTJ 处于 AP 状态时,在相反方向的写入电流作用下翻转到 P 状态的概率为 P_2。即有:

$$\begin{cases}P_{AP}=P_{AP}\cdot(1-P_2)+P_P\cdot P_1\\ P_{AP}+P_P=1\end{cases} \tag{8}$$

设每个 TRNG 单元输出比特“1”的概率为 P^1_{out},则其可以表示为:

$$P^1_{out}=P_{AP}=\frac{P_1}{P_1+P_2} \tag{9}$$

图 9 展示了式(9)中 P^1_{out} 的变化情况,其中 X 轴代表 P_2,Y 轴代表 P_1。当 P_1 和 P_2 理想上都等于 50% 时,随机数生成单元输出“1”的概率也是 50%。PVT 偏差可能导致 P_1 和 P_2 偏离。当 P_1 和 P_2 在同一方向上以相同的幅度移动时,P^1_{out} 仍然可以保持在 50%,如图中黑色实线所示。实际上,可能会影响输出概率的几种偏差来源(例如环境温度)确实倾向于造成 P_1 和 P_2 沿同一方向移动。所提出的电路设计使得 TRNG 能够对这类偏差保持容忍。

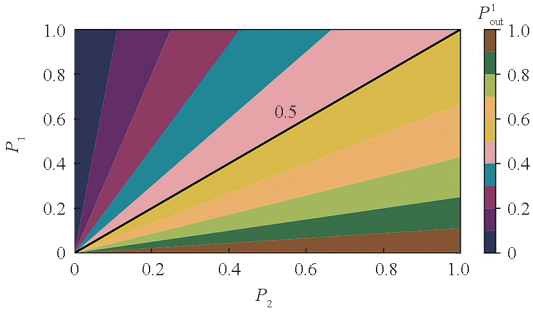


图 9　P_1 和 P_2 控制下 TRNG 输出“1”的概率

Fig. 9　Probability of TRNG outputting “1” under the control of P_1 and P_2

3 实验评估

3.1 实验验证方法概述

在实验中,构建从底层器件特性、电路功能到系统应用的全栈验证体系。在器件层面,通过集成电路仿真程序(simulation program with integrated circuit emphasis, SPICE)仿真验证 FTJ 突触的阻值连续可调特性,以时间 - 阻值曲线为指标,预期实现不同编程电压驱动下的阻值变化。电路层面重点评估 TRNG 的随机序列质量,采用 NIST SP800 - 22 随机数测试集测试通过率和熵值指标,通过与传统三阶段控制逻辑的对比实验,验证反馈写入机制、随机比特质量和 PVT 容忍性优势。系统层面以粒子输运问题为验证案例,对比软件模拟与硬件实现的精度差异,预期实现软硬件一致的平方误差。

3.2 实验设置

电路级模拟使用 Cadence Virtuoso 工具完成,采用 MTJ[23] 和 FTJ[26] 的紧凑模型以及 Cadence 通用工艺设计套件(generic process design kit, GPDK) 45 nm 工艺库。关键的 MTJ 参数列于表 1,主要的 FTJ 参数与模型[26] 预设值一致。除讨论温度偏差外,所有电路级的模拟均在 300 K(约 27 ℃)温度下进行。系统模拟器使用 Python 3.7 开发,在使用 Ubuntu 20.04.1 的 Intel i9 - 12900 CPU 上运行。

表 1 MTJ 紧凑模型的关键参数

Tab. 1 Key device parameters for MTJ compact model

参数	描述	值
t_{FL}	自由层厚度	1.3 nm
$\sigma_{t_{\mathrm{FL}}}$	t_{FL}的标准差	3% ×1.3 nm
d_{CD}	直径	32 nm
t_{TB}	绝缘层厚度	0.85 nm
$\sigma_{t_{\mathrm{TB}}}$	t_{TB}的标准差	3% ×0.85 nm
r_{TMR}	隧道磁阻率 TMR	200%
$\sigma_{r_{\mathrm{TMR}}}$	r_{TMR}的标准差	3% ×200%

3.3 电路仿真

3.3.1 概率存储

图 10 展示了基于 FTJ 的概率调节突触的 SPICE 仿真结果。图中从上到下分别为:编程电压 V_{prog}、概率调节突触输入电压 V_{in}、FTJ 畴壁中向下极化畴所占比例 r_{domain}以及 FTJ 的电阻 R_{FTJ}。

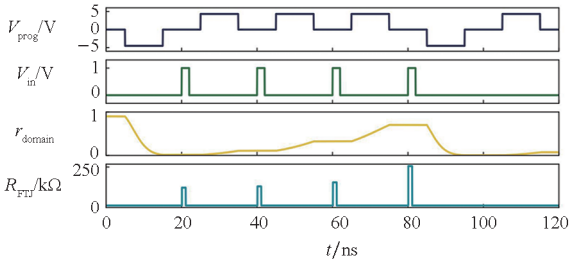


图 10 对概率调节突触进行的概率存储仿真

Fig. 10 Probabilistic storage simulation of the tunable probability synapse

首先,对编程模式,在 5 ~ 15 ns 之间,一个负向 V_{prog}脉冲将 FTJ 的铁电层设定为完全向上极化状态。随后,在 25 ~ 35 ns 期间,正向 V_{prog}脉冲激发畴壁成核和拓展,这一过程通过畴极化方向的变化体现。紧接着,在 45 ~ 55 ns 及 65 ~ 75 ns这两个时间段内,两次正向脉冲对 FTJ 进行编程,促使畴连续增长。之后,在 85 ~ 95 ns,一个负向 V_{prog}脉冲重置畴壁方向。而在 105 ~ 115 ns,正向脉冲再次引发畴的生长。整个过程中,畴在编程脉冲的控制下双向生长,展现了 FTJ 的忆阻特性。

接下来描述写驱动模式。在编程间隙,即分别在 20 ns、40 ns、60 ns 和 80 ns 时刻,输入写驱动脉冲 V_{in},较小的幅值并不会改变畴的极化状态,表明了忆阻器件的稳定性。不同极化畴状态宏观上表现为不同的电阻值,使得每次读取脉冲都能够读取到 FTJ 的不同阻值状态。

3.3.2 反馈写入

图 11 展示了对图 8 电路进行瞬态仿真的结果。图中从上至下为 SA 使能信号 Rd、写驱动使能信号 Wr、MTJ 状态以及图 7 中 Out 和$\overline{\mathrm{Out}}$处的读出电压 V_{Out}和 $V_{\overline{\mathrm{Out}}}$。每个随机比特生成周期包含读取和写入($t_{\mathrm{wr}}$)两个阶段;读取阶段进一步细分为预充电($t_{\mathrm{pre}}$)和读出($t_{\mathrm{rd}}$)。

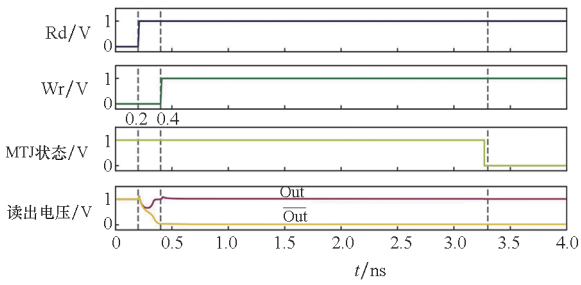


图 11 TRNG 执行反馈写入仿真

Fig. 11 Simulation of feedback write operation in TRNG

在读取阶段,Wr 和 Rd 信号均被设定为低电平。SA 完成预充电后,Rd 信号会被拉高。读取阶段结束,当前的 MTJ 状态便能在点 Out 处读

出。在写入阶段 Wr 信号被拉高,以设定概率向 MTJ 写入独处的相反值,在本节实验中设定 FTJ 阻值使 MTJ 在两个方向都以 50% 的概率翻转。可以看到,MTJ 在设定周期内发生翻转。瞬态仿真表明,对于翻转概率为 50% 的 TRNG,t_{pre} 和 t_{rd} 均为0.2 ns,t_{wr}设定为 2.9 ns 可正常工作,总的随机比特生成周期低至 3.3 ns。

3.3.3　MC 模拟与随机性检验

NIST 测试(SP 800-22 rev. 1a)[27]用于检验二进制比特序列是否满足统计随机性,常用来检测随机数生成器的随机数生成质量。通过 SPICE 级的 MC 仿真,在考虑 MTJ 随机性和器件工艺偏差的情况下,为 NIST 测试生成了 1×10^6 个随机比特。测试结果如表2 所示。可以看出, NIST 测试中每一项检测 P 值均超过阈值0.000 1,通过了表中所有的测试模块,这表明所设计的可调概率 TRNG 生成的比特流具有良好的统计随机性。

表 2　随机比特流 NIST 测试结果

Tab. 2　NIST test results for the random bit stream

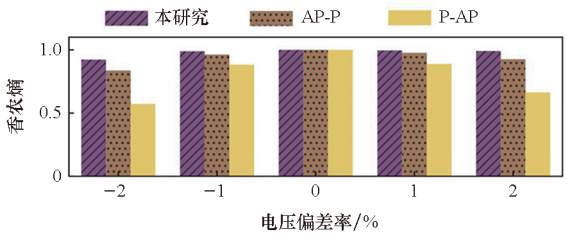
测试		P 值	通过率	Pass/Fail
Frequency		0.534 1	10/10	Pass
BlockFrequency		0.350 4	10/10	Pass
Cumu-lative-Sums	Forward	0.739 9	10/10	Pass
	Reverse	0.911 4	10/10	Pass
Runs		0.534 1	10/10	Pass
LongestRun		0.739 9	10/10	Pass
Rank		0.122 3	9/10	Pass
FFT		0.213 3	10/10	Pass
NonOverlapping-Template		—	145 9/148 0	Pass
Overlapping-Template		0.534 1	10/10	Pass
ApproximateEntropy		0.066 8	10/10	Pass
Serial	Forward	0.534 1	9/10	Pass
	Reverse	0.350 4	9/10	Pass
LinearComplexity		0.911 4	10/10	Pass

3.3.4　偏差容忍

器件工艺偏差已经由 MC 仿真引入,因此这里进一步考虑电压和温度偏差。本节 TRNG 翻转概率也固定在 50%。输出比特序列统计随机性引入输出熵量化。对于任意离散随机变量 X,其香农熵 $H_{Shannon}$由式(10)定义。对于每个比特,它有两个状态“0”和“1”,当其分布概率各为 50% 时,香农熵的最大值为“1”。同样地,对于这样的 X,其最小熵 H_{Min}由式(10)定义,描述最坏情况下的序列随机性。它的范围也在$[0,\log_2 m]$之间,其中 m 是 X 可能出现的状态数目。

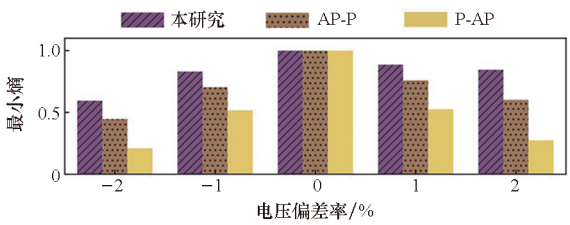
$$\begin{cases} H_{Shannon}(X) = -\sum_x P(x)\log_2 P(x) \\ H_{Min}(X) = \min[-\log_2 P(x)] \end{cases} \tag{10}$$

实验组采用了图 6 的控制逻辑,而对照组则使用单向概率写入,采用图 5 的控制逻辑,通过从 AP 到 P 和从 P 到 AP 的翻转过程控制随机比特生成。调节施加电流的大小,使得 MTJ 的翻转概率偏离 50%,电压偏差下不同配置 TRNG 的香农熵和最小熵比较如图 12 所示。对于受双向概率写入控制的 TRNG 电路,两个方向的翻转概率变化保持一致。可以看出,在电压偏差下对于受双向概率写入控制的 MTJ 单元而言,其香农熵和最小熵均能维持更接近“1”的水平。而受单向电流控制的随机翻转 TRNG 电路在电压偏差下香农熵和最小熵均显著下降。



(a) 香农熵对比

(a) Shannon entropy comparison



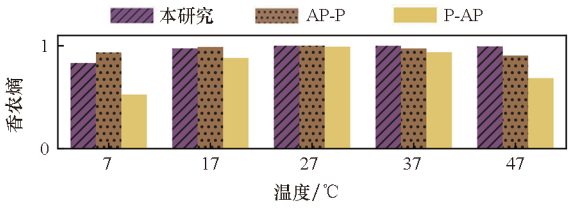
(b) 最小熵对比

(b) Min-entropy comparison

图 12　电压偏差下不同配置 TRNG 的香农熵和最小熵比较

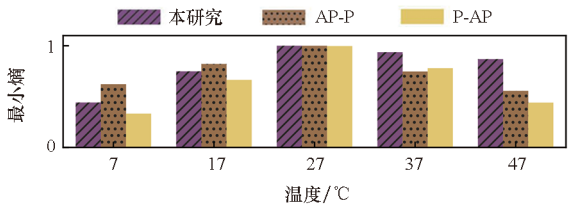
Fig. 12　Comparison of Shannon and minimum entropy for TRNG control logics under voltage variation

接下来考虑环境温度偏差,结果如图 13 所示,实验测量了从 7 ~ 47 ℃的温度偏差引起的输出熵的变化。可以看出,在温度升高时,受双向概率写入控制的 TRNG 电路具有更高的最小熵,但温度降低时,MTJ 在从 P 到 AP 的翻转过程的输出概率受到更大影响。因此,受双向概率写入控制的 TRNG 电路抵御高温偏差的能力强,这表明硬件电路适合集成到计算机系统中作为硬件加速器。



（a）香农熵对比

（a）Shannon entropy comparison



（b）最小熵对比

（b）Min-entropy comparison

图 13　温度偏差下不同配置 TRNG 香农熵和最小熵比较

Fig. 13　Comparison of Shannon and minimum entropy for TRNG control logics under temperature variation

3.4　系统仿真

3.4.1　问题描述

考虑一个粒子角通量密度问题，粒子状态只依赖于方向 Ω 而不依赖于空间，即方程(1)只考虑散射和吸收项的形式。假设一个粒子，它有两个方向状态：$\Omega=1$ 和 $\Omega=-1$，粒子可以发生散射，即按泊松过程均匀地选择一个新的方向，也可以按照第二泊松过程被吸收，粒子发生散射和吸收事件的速率分别由常数 Σ_a 和 Σ_s 控制。设定 $v=1$，散射后，粒子从状态 i 变为状态 j 的概率 $P_{i,j}=1/2$。粒子群的角通量密度服从玻耳兹曼输运方程：

$$\begin{cases}\dfrac{\partial}{\partial t}\Phi(t,\Omega) = -(\Sigma_a+\Sigma_s-S_\Sigma)\Phi(t,\Omega)+\\ \quad S_\Sigma\int[\Phi(t,\Omega+\omega)-\Phi(t,\Omega)]p^*(\omega+\Omega\,|\,\Omega)\,\mathrm{d}\omega\\ \Phi(0,\Omega)=\phi(\Omega)=\begin{cases}5 & \Omega=1\\ 3 & \Omega=-1\end{cases}\end{cases} \tag{11}$$

该玻耳兹曼输运方程的解析解为：

$$\Phi(t,\Omega)=\begin{cases}\dfrac{5}{2}\left[\mathrm{e}^{-\Sigma_a t}+\mathrm{e}^{-(\Sigma_a+\Sigma_s)t}\right]+\\ \quad\dfrac{2}{3}\left[\mathrm{e}^{-\Sigma_a t}-\mathrm{e}^{-(\Sigma_a+\Sigma_s)t}\right] & \Omega=1\\ \dfrac{5}{2}\left[\mathrm{e}^{-\Sigma_a t}+\mathrm{e}^{-(\Sigma_a-\Sigma_s)t}\right]+\dfrac{2}{3}\left[\mathrm{e}^{-\Sigma_a t}+\right.\\ \quad\left.\mathrm{e}^{-(\Sigma_a+\Sigma_s)t}\right] & \Omega=-1\end{cases} \tag{12}$$

接下来通过概率表示构造马尔可夫过程拟合的解析解。如 1.1 节所描述的方法，将玻耳兹曼方程描述为随机过程的期望，如式(4)的形式。粒子群的角通量密度的概率表示为：

$$\begin{cases}\Phi(t,\Omega)=E\{\mathrm{e}^{-\Sigma_a t}\phi[X(t)]\,|X(0)=\Omega\}\\ \mathrm{d}X(t)=\omega_{X(t)}\,\mathrm{d}P(t)\end{cases} \tag{13}$$

随机过程 $X(t)$ 表示粒子在 t 时刻的方向，粒子从初态 $\Omega=1$ 或 $\Omega=-1$ 开始，在泊松过程 $P(t)$ 发生时触发方向变化，吸收过程则由式(13)中的指数项表示。构造并跟踪表示泊松过程 $P(t)$ 的马尔可夫链，假设泊松事件发生的概率为 q_1。不发生的概率为 $1-q_1$。$P_{i,j}$ 定义与式(5)一致，则马尔可夫转移矩阵为：

$$\boldsymbol{C}=\begin{bmatrix}p_{1,1}q_1+(1-q_1) & p_{1,2}q_1\\ p_{2,1}q_1 & p_{2,2}+(1-q_1)\end{bmatrix} \tag{14}$$

对于所设计的可调概率 TRNG，可以设定其转移概率，使其中 MTJ 的状态转移矩阵与式(14)保持一致，由此即构建了通过所设计的概率可调 TRNG 构造马尔可夫链求解粒子输运问题的方法，Smith 等[9]对这个问题给出更详细的推导。

进一步设定散射截面的 $\Sigma_s=25.926$，吸收截面的 $\Sigma_a=0.5$，$\Delta t=0.01$。由于：

$$q_1=\Sigma_s\Delta t\mathrm{e}^{-\Sigma_s\Delta t} \tag{15}$$

则马尔可夫转移矩阵计算为：

$$\boldsymbol{C}=\begin{bmatrix}0.9 & 0.1\\ 0.1 & 0.9\end{bmatrix} \tag{16}$$

编程概率调节突触的阻值使得 P^1_{out} 与 P_2 均为 0.9，即完成构建模拟马尔可夫链的 TRNG。系统模拟器以 SPICE 仿真获得的随机序列作为粒子散射历史，在多核 CPU 上完成对跟踪历史的合并计算，通过式(13)计算期望得到系统中角通量密度随时间的变化。

3.4.2　效果和性能

通过在 TRNG 上执行大量随机比特生成跟踪粒子散射事件。考虑 $t\in[0\ \mathrm{s},1\ \mathrm{s}]$，设定跟踪粒子数 $W=500$，以平衡仿真开销和求解精度。这里进行了三组系统仿真。

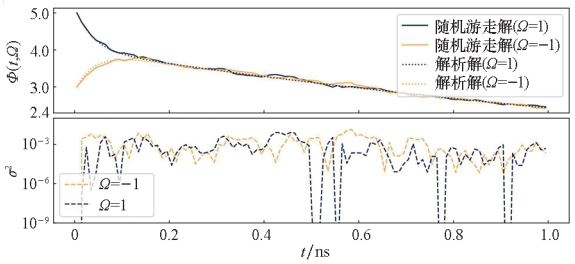
1）软件方法：通过软件生成随机序列，为模拟器提供粒子散射历史。

2）包含工艺偏差的模拟硬件方法：通过 SPICE 仿真生成随机序列，对每个粒子的跟踪从中取出长度为 100 的子序列作为粒子散射历史。工艺偏差由电路的 MC 仿真引入。

3）包含工艺和电压偏差的模拟硬件方法：相比上一种方法，在 SPICE 仿真中对 MTJ 翻转驱动电压，即图 8 中的 V_{out} 引入 $3\sigma=10\%$ 的正态分布

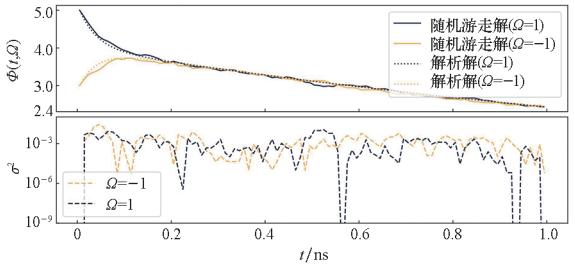
偏差,为模拟器提供粒子散射历史。

图 14 展示了三种方法执行随机游走获得的方程(11)的解与解析解的对比。



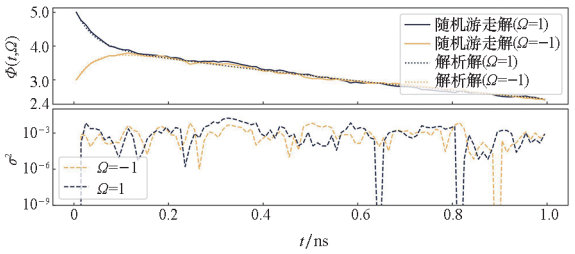
(a) 模拟器通过软件执行随机数生成解方程

(a) Simulator executes random number generation equations through software



(b) 模拟器读取模拟硬件生成随机比特序列解方程

(b) Simulator reads the random bit stream generated by the simulated hardware to solve equations



(c) 模拟器读取模拟硬件考虑电压偏差下生成随机比特序列解方程

(c) Simulator reads the random bit stream generated by the simulated hardware under voltage variation to solve equations

图 14　模拟器通过不同跟踪方法求得的输运方程解的对比

Fig. 14　Comparison of transport equation solutions using different tracking methods in the simulator

图 14(a)表示通过软件生成随机数求得的解。横坐标为时间,上半部分纵坐标为粒子角通量密度 $\Phi(t,\Omega)$,解析解由虚线表示,分别对应 $\Omega=1$ 和 $\Omega=-1$ 两种不同的初始情况。与解析解相同颜色的实线表示对应初始情况下通过随机游走近似的解。可以看到,软件生成随机数获得的游走历史求得的数值解能够很好地拟合解析解,它们之间的差异由图中下半部分表示,平方误差计算为:

$$\sigma^2 = |C_{\mathrm{rm}}(t) - C_{\mathrm{an}}(t)|^2 \tag{17}$$

式中,$C_{\mathrm{rm}}(t)$ 为时间 t 时的随机游走解,$C_{\mathrm{an}}(t)$ 为时间 t_i 时的解析解。在当前模拟粒子规模下,软件方法的平方误差可以控制在 1×10^{-2} 以下。

图 14(b)展示的是包含器件工艺偏差的模拟硬件方法,表示粒子散射历史的随机比特序列来源于 SPICE 仿真,器件工艺偏差由 MC 仿真引入。可见所设计的概率可调 TRNG 能够执行对粒子输运问题的求解,解的平方误差也控制在 1×10^{-2} 以下,与软件方法基本一致,说明所设计的电路在求解粒子输运问题时与理论算法有一致的效果。

为进一步评估电路在电压偏差下解方程的能力,图 14(c)展示 $3\sigma=10\%$ 的电压偏差下模拟硬件求解粒子输运的结果,引入的电压偏差被模拟随机游走所需的大量随机比特生成操作隐藏,因此没有造成求解精度的显著下降。另外,虽然 MTJ 翻转概率随电压变化并不是线性的,但是在引起 MTJ 概率翻转的较小的电压变化范围内,概率变化是近似线性的,因此电压的正态偏差不会使概率的期望显著偏离,这也体现了所设计的 TRNG 抵抗偏差的能力。

性能方面。考虑一次随机比特生成时间,所设计概率可调 TRNG 跟踪操作的执行时间由设计周期决定,包含 0.4 ns 读取操作和 1 ns 概率写操作。在通用 CPU 上通过软件生成随机序列,随机数生成方法分别采用 random 库的 random,os 库的 urandom 以及 secrets 库的 randbelow 三种方法,平均时间分别为 0.24 μs、0.49 μs 和 1.44 μs。因此 TRNG 相比 CPU 上的软件实现,性能提升约 171 ~ 1 028 倍。需要说明的是,具有 3 584 CUDA 核的 GPU 求解输运问题相较 CPU 仅获得 55.6 倍的加速比[28],因而 CPU 作为比较基准,代表具有最高串行性能的通用处理器。相较非冯·诺依曼架构,神经形态处理器 TrueNorth 的单次输运跟踪时间为 484 μs[8],远低于具有物理随机性的自旋硬件方法。

这个输运问题作为粒子输运 MC 仿真的概率计算原语,构建了从器件特性到算法求解的映射。为了充分发挥 MC 方法在高维问题的优势,多个 TRNG 单元可协同工作以跟踪复合概率事件,为构建高维概率计算阵列提供了硬件基础。以求解随机微分方程的随机游走模型[9]为例,一维粒子运动跟踪可通过分别跟踪泊松事件触发和随机游走方向的两组 TRNG 实现,更多维度则由独立变量的张量积获得。

4 自旋 TRNG 关键参数对比与分析

表3对比了本文 TRNG 与现有自旋 TRNG 的吞吐率、功耗和面积。通过 Cadence Virtuoso 仿真和 GPDK 45 nm 布局,在 MTJ 翻转概率为50%的条件下,本文设计随机比特吞吐率达到303 Mbit/s、功耗为 2.99 pJ/bit、单元面积为 9.79 μm^2。在现有基于 STT-MTJ 的 TRNG 设计中,本文提出的方案凭借铁电概率突触实现了独特的概率可调特性,同时通过简化控制逻辑和消除重置阶段获得了最优的随机比特生成速度。尽管受限于45 nm 工艺节点,该设计在面积和功耗方面仍保持可接受的性能水平。Zhang 等[29]提出的基于 SOT-MTJ 的可调概率 TRNG 设计,其随机写脉冲为300 ps,未给出读取时间和外围电路的面积、功耗情况,且缺乏对 PVT 容忍的相关评估。虽然 SOT-MTJ 同样具有设计概率可调 TRNG 的潜力,但作为三端器件,电路设计具有更高的复杂性,不利于进一步设计粒子输运加速器,这也是考虑使用 STT-MTJ 器件进行设计的主要原因。未来的性能改进将受益于与先进 CMOS 工艺的集成以及具有更快切换响应的自旋器件工艺等。

表3 与自旋 TRNG 的对比

Tab. 3 Comparison with spin-based TRNGs

文献	MTJ	概率可调	吞吐率/(Mbit/s)	功耗/(pJ/bit)	面积/μm^2
[15]	STT	N	7.7 ~ 15.1	5.7 ~ 13.4	50.6 ~ 200.6
[14]	STT	N	50	1.1	219
[17]	STT	N	66.7 ~ 177.8	0.6 ~ 0.8	3.8 ~ 7.6
[18]	STT	N	66.7	0.8	3.84
[20]	STT	N	303	2.65 ~ 5.3	14.5 ~ 24.29
[29]	SOT	Y	< 1 000	—	—
本文	STT	Y	303	2.99	9.79

5 结论

本文提出了一种概率可调 TRNG。该电路利用自旋器件的物理随机性作为熵源,通过铁电器件的忆阻特性实现输出概率可调,并结合电路逻辑设计缩短随机比特生成周期,展现出对 PVT 偏差的容忍性。相较于产生均匀分布的 TRNG,所提出的设计避免了后处理带来的运算和访存开销。特别在加速粒子输运问题求解中,相比在通用 CPU 上的执行,所提出硬件设计获得 171 ~ 1 028倍的性能加速,展示了概率可调 TRNG 加速粒子输运计算的显著优势。这项工作揭示了自旋电子器件在推动随机概率计算加速系统设计方面的巨大潜力。未来的工作将集中提高设计的通用性,包括基于当前架构,通过增加 TRNG 单元数量并按维度需求进行拓扑连接,配合相应的概率参数映射策略,可直接扩展至高维问题求解。

参考文献(References)

[1] WHITE R D, ROBSON R E, DUJKO S, et al. Recent advances in the application of Boltzmann equation and fluid equation methods to charged particle transport in non-equilibrium plasmas[J]. Journal of Physics D: Applied Physics, 2009, 42(19): 194001.

[2] 邓力, 李刚, 张宝印, 等. 蒙特卡罗粒子输运方法及应用研究[J]. 核动力工程, 2025, 46(3): 1 - 17.
DENG L, LI G, ZHANG B Y, et al. Study on Monte Carlo particle transport method and application[J]. Nuclear Power Engineering, 2025, 46(3): 1 - 17. (in Chinese)

[3] MA D H, YANG B, ZHANG Q Y, et al. Evaluation of single-node performance of parallel algorithms for multigroup Monte Carlo particle transport methods[J]. Frontiers in Energy Research, 2021, 9: 705823.

[4] TRAMM J, ALLEN B, YOSHII K, et al. Efficient algorithms for Monte Carlo particle transport on AI accelerator hardware[J]. Computer Physics Communications, 2024, 298: 109072.

[5] 张建民, 刘津津, 许炜康, 等. 符合粒子输运模拟的专用加速器体系结构[J]. 国防科技大学学报, 2025, 47(2): 155 - 164.
ZHANG J M, LIU J J, XU W K, et al. Specific accelerator architecture conforming to particle transport simulation[J]. Journal of National University of Defense Technology, 2025, 47(2): 155 - 164. (in Chinese)

[6] FU S Q, LI T J, ZHANG J M, et al. MiniMCTAD: minimalist Monte Carlo transport architecture design[C]// Proceedings of IEEE International Conference on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking, 2021: 1 - 10.

[7] SEVERA W, LEHOUCQ R, PAREKH O, et al. Spiking neural algorithms for Markov process random walk[C]// Proceedings of International Joint Conference on Neural Networks (IJCNN), 2018: 1 - 8.

[8] SMITH J D, SEVERA W, HILL A J, et al. Solving a steady-state PDE using spiking networks and neuromorphic hardware[C]//Proceedings of the International Conference on Neuromorphic Systems, 2020: 1 - 8.

[9] SMITH J D, HILL A J, REEDER L E, et al. Neuromorphic scaling advantages for energy-efficient random walk computations[J]. Nature Electronics, 2022, 5(2): 102 - 112.

[10] MISRA S, BLAND L C, CARDWELL S G, et al. Probabilistic neural computing with stochastic devices[J]. Advanced Materials, 2023, 35(37): 2370264.

[11] 王亚迪, 郭杭闻, 沈健. 基于磁隧道结的概率计算研究进展[J/OL]. 中国科学: 物理学 力学 天文学, 2025: 1 - 17 (2025 - 06 - 11)[2025 - 06 - 24]. https://kns.cnki.net/kcms/detail/11.5848.N.20250611.1122.002.html.

WANG Y D, GUO H W, SHEN J. Probabilistic computing based on magnetic tunnel junction[J/OL]. Scientia Sinica (Physica, Mechanica & Astronomica), 2025: 1 - 17 (2025 - 06 - 11)[2025 - 06 - 24]. https://kns.cnki.net/kcms/detail/11.5848.N.20250611.1122.002.html. (in Chinese)

[12] GALLAGHER W J, CHIEN E, CHIANG T W, et al. 22 nm STT-MRAM for reflow and automotive uses with high yield, reliability, and magnetic immunity and with performance and shielding options[C]//Proceedings of the IEEE International Electron Devices Meeting (IEDM), 2019: 2.7.1 - 2.7.4.

[13] SATO N, ALLEN G A, BENSON W P, et al. CMOS compatible process integration of SOT-MRAM with heavy-metal bi-layer bottom electrode and 10 ns field-free SOT switching with STT assist [C]//Proceedings of IEEE Symposium on VLSI Technology, 2020.

[14] AMIRANY A, JAFARI K, MOAIYERI M H. True random number generator for reliable hardware security modules based on a neuromorphic variation-tolerant spintronic structure[J]. IEEE Transactions on Nanotechnology, 2020, 19: 784 - 791.

[15] PERACH B, KVATINSKY S. An asynchronous and low-power true random number generator using STT-MTJ[J]. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2019, 27(11): 2473 - 2484.

[16] VATAJELU E I, DI NATALE G. High-entropy STT-MTJ-based TRNG[J]. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2019, 27(2): 491 - 495.

[17] QU Y Z, HAN J, COCKBURN B F, et al. A true random number generator based on parallel STT-MTJs [C]// Proceedings of the Design, Automation & Test in Europe Conference & Exhibition (DATE), 2017: 606 - 609.

[18] QU Y Z, COCKBURN B F, HUANG Z, et al. Variation-resilient true random number generators based on multiple STT-MTJs[J]. IEEE Transactions on Nanotechnology, 2018, 17(6): 1270 - 1281.

[19] MORSALI M, MOAIYERI M H, RAJAEI R. A process variation resilient spintronic true random number generator for highly reliable hardware security applications [J]. Microelectronics Journal, 2022, 129: 105606.

[20] FU S Q, LI T J, ZHANG C Y, et al. RHS-TRNG: a resilient high-speed true random number generator based on STT-MTJ device [J]. IEEE Transactions on Very Large Scale Integration (VLSI) Systems, 2023, 31(10): 1578 - 1591.

[21] MASUDA N, PORTER M A, LAMBIOTTE R. Random walks and diffusion on networks[J]. Physics Reports, 2017, 716: 1 - 58.

[22] HU Y Z. Some recent progress on stochastic heat equations[J]. Acta Mathematica Scientia, 2019, 39(3): 874 - 914.

[23] WU L Z, RAO S, TAOUIL M, et al. MFA-MTJ model: magnetic-field-aware compact model of pMTJ for robust STT-MRAM design[J]. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2022, 41(11): 4991 - 5004.

[24] FANG H, WANG J, NIE F, et al. Giantelectroresistance in ferroelectric tunnel junctions via high-throughput designs: toward high-performance neuromorphic computing[J]. ACS Applied Materials & Interfaces, 2024, 16(1): 1015 - 1024.

[25] ZHAO W S, CHAPPERT C, JAVERLIAC V, et al. High speed, high stability and low power sensing amplifier for MTJ/CMOS hybrid logic circuits [J]. IEEE Transactions on Magnetics, 2009, 45(10): 3784 - 3787.

[26] WANG Z H, ZHAO W S, KANG W, et al. Compact modelling of ferroelectric tunnel memristor and its use for neuromorphic simulation[J]. Applied Physics Letters, 2014, 104(5): 053505.

[27] RUKHIN A, SOTO J, NECHVATAL J, et al. A statistical test suite for random and pseudorandom number generators for cryptographic applications [R]. Gaithersburg: National Institute of Standards & Technology, 2010.

[28] SONG P T, ZHANG Z J, LIANG L, et al. Implementation and performance analysis of the massively parallel method of characteristics based on GPU[J]. Annals of Nuclear Energy, 2019, 131: 257 - 272.

[29] ZHANG R, LI X H, ZHAO M K, et al. Probability-distribution-configurable true random number generators based on spin-orbit torque magnetic tunnel junctions[J]. Advanced Science, 2024, 11(23): 2402182.