DOI:10.16136/j.joel.2023.05.0314

面向混沌激光熵源的安全随机数提取

李瑞敏[1],李 璞[1,2*],蔡 强[1],王冰洁[1],马 荔[3],徐兵杰[3]

(1.太原理工大学 新型传感器与智能控制教育部重点实验室,山西 太原 030024;2.广东工业大学 广东省信息光子技术重点实验室,广东 广州 410006;3.西南通信研究所 保密通信实验室,四川 成都 610041)

摘要:基于混沌激光的随机数发生器可以实现大量高速的真随机数产生。然而,受随机数产生过程中引入的额外技术噪声影响,原始随机数无法提供真随机性。为了从信息论的角度实现安全随机数的产生,必须定量评估熵源产生的随机性。本文中通过实验搭建基于混沌激光的随机数产生装置,使用条件最小熵评估原始随机数中可提取的真随机性。同时,通过更换不同的器件讨论关键参数对原始随机性的影响。在原始随机数的后处理提取过程中使用信息论可证安全的Toeplitz提取器,最终实现了安全随机数的产生。

关键词:光通信;安全随机数;条件最小熵;混沌激光;随机性;信息论

中图分类号:TN249 **文献标识码**:A **文章编号**:1005-0086(2023)05-0536-07

Secure random number extraction for chaotic laser entropy source

LI Ruimin[1], LI Pu[1,2*], CAI Qiang[1], WANG Bingjie[1], MA Li[3], XU Bingjie[3]

(1. Key Laboratory of Advanced Transducers and Intelligent Control System, Ministry of Education, Taiyuan University of Technology, Taiyuan, Shanxi 030024, China; 2. Guangdong Provincial Key Laboratory of Information Photonics Technology, Guangzhou, Guangdong 510006, China; 3. Science and Technology on Communication Security Laboratory, Institute of Southwestern Communication, Chengdu, Sichuan 610041, China)

Abstract: The random number generator based on chaotic laser can generate a large number of true random numbers at high speed. However, raw random numbers cannot provide true randomness due to the additional technical noise introduced in the random number generation process. To achieve secure random number generation from information-theoretic standpoint, the randomness generated by entropy source must be quantitatively evaluated. In this paper, a chaotic laser-based random number generation device is experimentally constructed and the true randomness that can be extracted from raw random numbers is evaluated by conditional minimum entropy. Simultaneously, the effect of key parameters on the original randomness is discussed by replacing different devices. An information-theoretically falsifiable secure Toeplitz extractor is applied in the post-processing extraction process. Finally, the generation of secure random numbers is implemented.

Key words: optical communication; secure random numbers; conditional minimum entropy; chaotic laser; randomness; information-theoretic

0 引 言

随机数是许多信息和通信技术的一个重要组成部分[1,2]。尤其在信息安全领域,随机数作为密钥对明文信息进行加密,是保障信息安全的重要基石。对于不考虑随机数安全性和唯一性的应用,一个具有均匀分布的数字序列就基本足够了。这样的序列可以使用通过确定性算法工作的伪随机数生成器(pseudo random number generator, PRNG)来生成,并且可以达到高度无偏。但由于

* E-mail:lipu8603@126.com

收稿日期:2022-04-30 修订日期:2022-05-26

基金项目:国家自然科学基金(62175177,U19A2076,61927811,61731014)、山西省自然科学基金(201901D211116,201901D211077)、四川省科技计划资助项目(2022YFG0330)、广东省创新创业团队计划和保密通信重点实验室稳定计划支持项目(2022)资助项目

它本身是基于固定算法的,不具备真正的随机性,一旦算法和种子被攻击者确定,所有安全性都将丢失。因此,它们不能用于需要信息安全的应用领域。相对于 PRNG,利用随机物理过程[3-6]作为熵源的物理随机数发生器拥有真正的随机性。

混沌激光物理熵源由于其高带宽和快响应度等优势受到了广泛的关注和研究[7-10]。现有的混沌光源产生方案包括如下几种:光反馈混沌[11]、光注入混沌[12]、互注入混沌[13]以及白混沌[14]等等。其中,白混沌是一种宽带混沌信号,具有功率谱平坦宽带、幅值分布对称且无时延特征等优点。近年来,利用白混沌作为物理熵源的随机数产生方案已被实现[15,16]。

然而,信号传输过程中引入的经典技术噪声造成了可被攻击的不安全因素,使熵源输出的随机数丧失了原有的随机性。所以需要对原始信号产生的随机数进行随机性评估,产生符合信息安全标准的随机数。现有的对随机数进行安全评估的方法大多使用 NIST(the U. S. National Institute of Standards and Technology)测试[17]来进行,仅通过判断随机数是否通过测试来验证安全,还达不到信息安全的标准。要达到该标准,需从信息论的角度来分析信号传输过程中的不安全因素,定量评估不安全条件下可提取的随机性,并使用信息论可证安全的后处理方法来进行随机数提取[18]。

本文通过实验搭建了一个基于宽带白混沌激光的随机数产生装置,通过条件最小熵定量评估了不安全因素存在条件下原始数据可提取的随机性含量。并且分析了经典技术噪声偏移 e 和混沌信噪比(chaos-classical noise ratio, *CCNR*)对于最终可提取随机性含量的影响。最后通过使用信息论可证安全的 Toeplitz 提取器对原始数据进行后处理提取,实现了安全随机数的产生。

1　白混沌熵源产生

实验装置如图 1 所示,把两路光反馈混沌信号进行拍频,外差测量获得带宽展宽的白混沌光信号。实验装置包含分布式反馈激光器(DFB1、DFB2)、偏振控制器(PC1、PC2)、光耦合器(OC1、OC2、OC3)、可调光衰减器(VOA1、VOA2)、光纤反射镜(FM1、FM2)、光隔离器(ISO1、ISO2)和光纤放大器(EDFA1、EDFA2)。首先,激光通过偏振控制器 PC 进入光耦合器 OC,被分为两路信号。两路信号分别经过光衰减器 VOA 和光纤反射镜 FM 进行反射,反射光再经过光衰减器 VOA 和光耦合器 OC 反馈回激光器中形成扰动,产生混沌激光,产生的混沌激光再由光耦合器的另一路输出。最后,光反馈产生的混沌激光经过光隔离器 ISO,被光纤放大器 EDFA 放大。两路放大的光信号进入 3 dB 光耦合器 OC3 中拍频,由光电平衡探测器 BPD 进行外差探测,获取最终的白混沌信号。

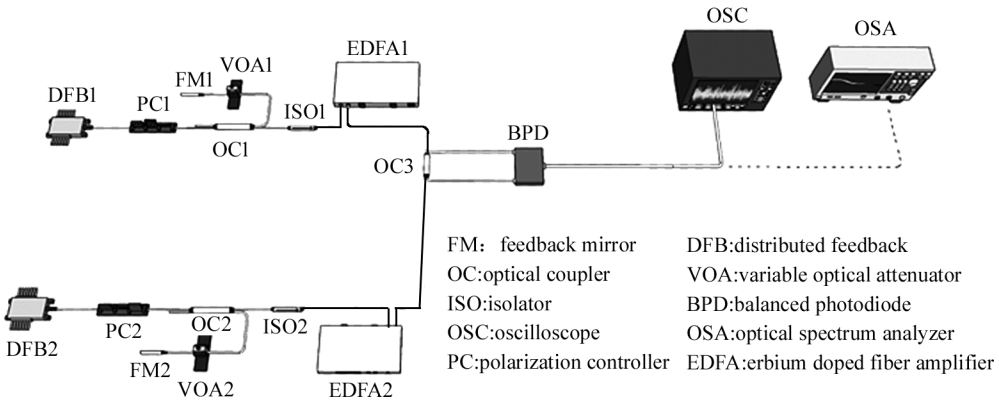

图 1　白混沌产生的实验装置

Fig. 1　Experimental setup of white chaos generation

实验中,分布式反馈激光器 DFB1 和 DFB2(Eblana, EP1550-DM-BO5-FM)的阈值电流分别为 12.05 mA 和 11.4 mA,通过电流源(ILX Lightwave, LDX-3412)调节其工作电流为 21.7 mA,工作温度在 24.9 ℃。两路混沌光拍频,经过光电平衡探测器(u^2t, BPDV2150R)后产生外差信号,由示波器(LeCroy, LabMaster10-36Zi,36 GHz,80 GS/s)和频谱仪(Agilent,N9030A)进行时序和频谱测量。

2　白混沌熵源量化

对于理想情况下的宽带白混沌信号,当不考虑传输过程中的额外噪声时,其产生的变量应服从一个以 0 为中心,方差为 σ_C^2 的高斯概率密度函数。而在实际情况下,这些变量不能在有独立于熵源外的

噪声的情况下被测量。测量到的总输出信号 $M = C + E$，且此信号的概率密度分布 p_M 为纯混沌信号概率密度分布 p_C 与传输中电子学噪声概率密度分布 p_E 的卷积。假设，电子学噪声服从中心为 0、方差为 σ_E^2 的高斯分布，那么测量到的总的概率密度分布为：

$$p_M(m) = \frac{1}{\sqrt{2\pi}\sigma_M}\exp\left(-\frac{m^2}{2\sigma_M^2}\right), \tag{1}$$

式中，电压值 m 所在总信号 M 的方差 $\sigma_M^2 = \sigma_C^2 + \sigma_E^2$。混沌信号与探测中电子学信号的方差定义了 $CCNR$，即 $CCNR = 10 \cdot \lg(\sigma_C^2/\sigma_E^2)$。

由于熵源输出的混沌信号 M 为连续信号，所以需要使用模数转化器（analog-to-digital converter，ADC）对其进行模数转换。对于分辨率为 n 的 ADC 来说，其动态范围为$[-R + \delta/2, R - 3\delta/2]$。当信号被测量后，采样的连续信号被离散化，一般令中心帧对应 0 电压值，此时整个采样范围被分为帧宽 $\delta = R/2^{n-1}$ 的 2^n 个帧。由此产生的离散化信号 M_{dis} 的概率分布如图 2 所示。

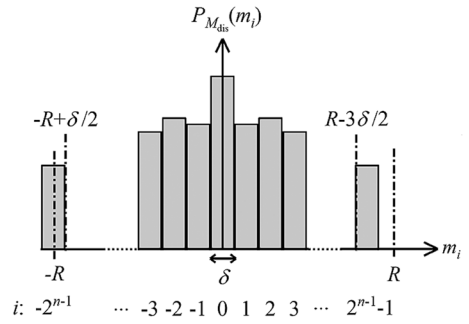

图 2　n 位 ADC 的模型

Fig. 2　Model of the n-bit ADC

图中，$m_i = \delta \times i$，i 为范围$\{-2^{n-1}, \cdots, 2^{n-1}-1\}$中的整数。两个边缘帧 $i_{\min}$ 和 $i_{\max}$ 分别代表有限输入范围 ADC 的第一个和最后一个帧的饱和累加概率。分离后的每一帧的统计概率可由下式计算：

$$P_{M_{\text{dis}}}(m_i) = \begin{cases} \int_{-\infty}^{-R+\delta/2} p_M(m)\mathrm{d}m, & i = i_{\min} \\ \int_{m_i-\delta/2}^{m_i+\delta/2} p_M(m)\mathrm{d}m, & i_{\min} < i < i_{\max} \\ \int_{R-3\delta/2}^{\infty} p_M(m)\mathrm{d}m, & i = i_{\max} \end{cases}。 \tag{2}$$

图 3 表示了不同动态参数 R 下的离散概率分布 $P_{M_{\text{dis}}}(m_i)$。从图中可看出，在模拟数字转换过程中，需要考虑 ADC 的采样范围的限制。当 ADC 采样范围相对于探测系统输出的光电信号幅值范围过小时（$R = 2$），超出采样范围的测量结果将会被赋以最低或最高的帧，此时将导致转换生成的随机序列中包含了大量的 0 或者 1 比特串。相反，如果 ADC 采样范围相对光信号过大（$R=8$），将导致过多的编码未被调用。

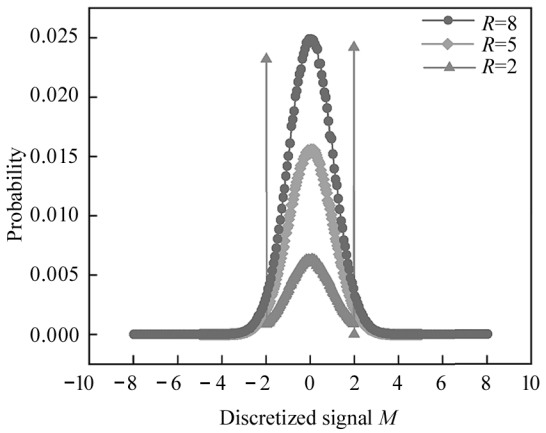

图 3　不同动态参数 R 设置下的离散概率分布 $P_{M_{\text{dis}}}(m_i)$

Fig. 3　Discretized distribution probabilities $P_{M_{\text{dis}}}(m_i)$ with different dynamical parameter R

以上两种情况都将导致产生的随机数据中某些比特组合的大量出现，使其丧失随机性。所以在实际的随机数产生过程中必须合理调整模拟信号的幅值范围并选择合适的 ADC 动态范围，以便尽可能正确地采用完整的 n 位采样位数采样。理论上，当中间帧的概率与两个边缘帧中概率的较大值相等时，ADC 的动态参数 R 为最优值。

考虑到作为第三方的技术噪声影响，理想混沌信号会在实际中产生偏移量 e，综合考虑实际情况后，信号经过模数转换离散化后的概率分布重置为：

$$P_{M_{\text{dis}}|E}(m_i \mid e) = \begin{cases} \int_{-\infty}^{-R+\delta/2} p_{M|E}(m \mid e)\mathrm{d}m, & i = i_{\min} \\ \int_{m_i-\delta/2}^{m_i+\delta/2} p_{M|E}(m \mid e)\mathrm{d}m, & i_{\min} < i < i_{\max} \\ \int_{R-3\delta/2}^{\infty} p_{M|E}(m \mid e)\mathrm{d}m, & i = i_{\max} \end{cases}。 \tag{3}$$

假设探测过程中引入的经典噪声可以在任意精度下被完全了解，则最大的条件概率为：

$$\max_{m_i \in M_{\text{dis}}} P_{M_{\text{dis}}|E}(m_i \mid e) =$$

$$\max\begin{cases}\frac{1}{2}\left[1-\mathrm{erf}\left(\frac{e+R-\delta/2}{\sqrt{2}}\right)\right]\\ \mathrm{erf}\left(\frac{\delta}{2\sqrt{2}}\right)\\ \frac{1}{2}\left[\mathrm{erf}\left(\frac{e-R+3\delta/2}{\sqrt{2}}\right)+1\right]\end{cases}。\quad(4)$$

最终,条件最小熵的表达式为:

$$H_{\min}(M_{\mathrm{dis}}\mid E)=-\log_2\left(\max\left\{\frac{1}{2}\left[\mathrm{erf}\left(\frac{e_{\max}-R+3\delta/2}{\sqrt{2}}\right)+1\right],\ \mathrm{erf}\left(\frac{\delta}{2\sqrt{2}}\right)\right\}\right),\quad(5)$$

并且该条件最小熵公式可以通过式(6)选择 R 的范围来优化:

$$\frac{1}{2}\left[\mathrm{erf}\left(\frac{e_{\max}-R+3\delta/2}{\sqrt{2}}\right)+1\right]=\mathrm{erf}\left(\frac{\delta}{2\sqrt{2}}\right)。\quad(6)$$

此时的最小熵为最坏条件下的最小熵,它等于窃听者可获取信息的最大概率,在此分析方法下,条件最小熵是信息论上可证明的评估工具,并且严格地评估了总信号中混沌信号的随机性,给出了该分布中随机性的最大含量。

3　白混沌熵源条件最小熵评估

图 4 为白混沌和噪声信号相关的测量结果。其中,图 4(a)为宽带白混沌信号和噪声的时序图,插图为噪声的放大图。经典技术噪声的数据通过关掉光路中的 DFB 和 EDFA 后在示波器上获得。图 4(b)为两者所对应的概率分布图,其中红线为得到的白混沌的高斯拟合曲线。从统计结果可以看出,其振幅呈类高斯分布。这样的特性能保证无偏和高质量的随即比特的产生。通过计算,可知该实验条件下对应的具体参数 σ_M 为 138 mV,σ_E 为 1.69 mV。图 4(c)为频谱图,可输出的 *CCNR* 可近似为频谱上混沌信号与噪声信号的平均间隙 20.1 dB。

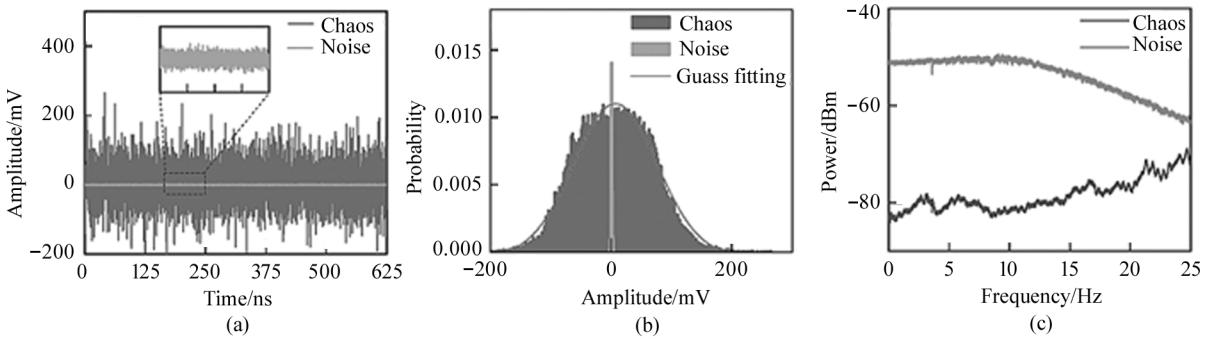

图 4　白混沌和噪声的测量结果:(a) 时序图; (b) 概率分布图; (c) 频谱图

Fig. 4　Measurement results of white chaos and noise: (a) Temporal waveforms; (b) Probability distribution; (c) Spectral power density

对于该实验数据下偏差 $e=7\sigma_E$,通过式(5)与(6)计算可得条件最小熵,如图 5 所示。通过图 5 的标记点可知,此数据可获得的条件最小熵为 6.7,即现有的 8 位 ADC 量化出的随机码中,每 8 位中有 6.7 位的真随机性含量。

通过实验中更换不同的器件,可以获得在置信水平为 99%的情况下,经典技术噪声产生的最小偏移范围为 $-3\sigma_E\leqslant e\leqslant 3\sigma_E$。同时,由于实验中器件不可避免的直流分量 Δ,则总的偏移量增大,总的偏移范围达到 $3\sigma_E\leqslant|e+\Delta|\leqslant 20\sigma_E$,在此条件下,使用公式(5)计算,可获得不同的 *CCNR* 对条件最小熵和标准化条件最小熵的影响,如图 6 所示。阴影部分表示偏移范围内条件最小熵的取值范围,实线为 $3\sigma_E$ 偏移量,虚线为 $20\sigma_E$ 偏移量。

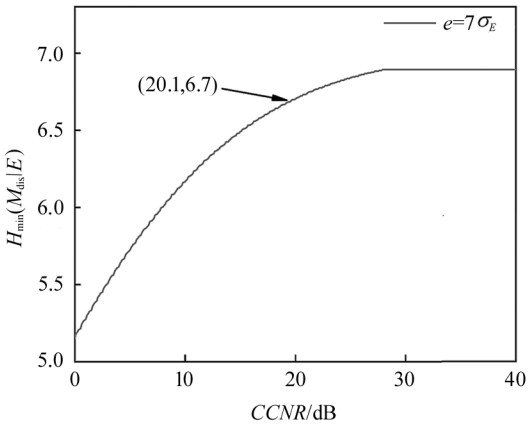

图 5　7 倍标准差 σ_E 下的条件最小熵 $H_{\min}(M_{\mathrm{dis}}\mid E)$

Fig. 5　The conditional minimum entropy $H_{\min}(M_{\mathrm{dis}}\mid E)$ with $e=7\sigma_E$

从图6(a)中可以看出，高*CCNR*状态下，噪声的贡献不会对可提取比特造成太大影响。随着噪声越来越接近混沌，需要丢弃更多的比特，但仍然可以提取相当数量的安全随机比特。偏移量越大，可提取的随机比特越少。而且从图中可以看出，即使*CCNR*低于0，即噪声信号大于混沌光信号，原则上仍然可以得到一定与噪声无关的非零数量的随机比特。从图6(b)中可知，每比特可提取的随机性随着分辨率n的增加而增加，并且在最大的噪声偏移量下，条件最小熵在每位中的提取比例也可达到87.6%。

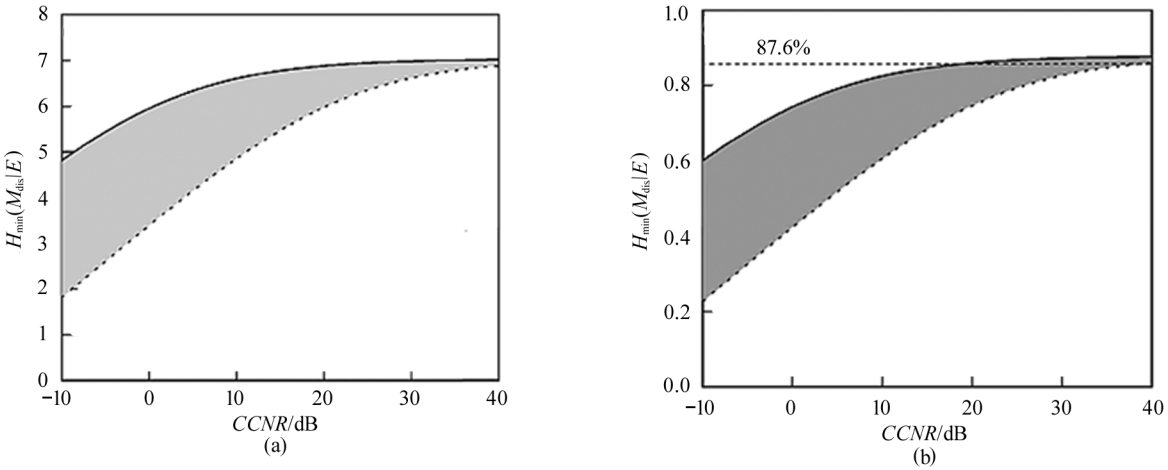

图6　不同信噪比*CCNR*的随机性评估：(a)条件最小熵$H_{min}(M_{dis}|E)$；(b)标准化条件最小熵$H_{min}(M_{dis}|E)/n$

Fig. 6　Randomness evaluation with different *CCNR*: (a) The conditional minimum entropy $H_{min}(M_{dis}|E)$; (b) The normalize conditional minimum entropy $H_{min}(M_{dis}|E)/n$

4　信息论可证安全随机数提取

对于宽带白混沌激光来说，输出信号的随机性并不理想化，其统计分布是有一定偏差的。为了产生理想的随机性，需要对原始输出进行后处理，以产生更短但几乎均匀分布的随机字符串。传统的后处理方法在实践中虽然简单，但不适用于此类原始比特中存在不可忽略的自相关的情况。要实现信息论可证安全的随机数产生，需要采用信息论可证明的Toeplitz后处理方法进行随机性提取。

本文采用Toeplitz强提取器从原始采样后的随机序列中提取真随机数。通过将原始序列n与Toeplitz矩阵(n×m矩阵)相乘来消除不安全因素造成的影响，从中提取真随机比特串m，同时将非均匀分布的原始随机比特串转化为均匀分布的随机比特串。通过条件最小熵计算的随机性含量构造Toeplitz矩阵，由于该矩阵的特殊性，只需要(n + m−1)个随机种子就可建立。因为Toeplitz提取器是强提取器，所以随机种子可以重复使用，其中n与m需满足$n/m \leqslant P$，$P = H_{min}(M_{dis}|E)/n$，n为ADC的分辨率。在具体的实验条件下，选择输入比特长度n为256，输出比特长度为256×6.7/8=214.4，为了缩小与理想随机序列的可分辨距离，输出长度m设置为114。根据剩余哈希引理，可计算出信息论安全参数$\varepsilon = 2^{-50}$，即提取后的随机序列和理想的均匀序列以2^{-50}为界不可区分。最后本文用256×114的Toeplitz矩阵实现了随机数提取，实现了信息论上可证安全的随机数产生。

为了对最终的随机数进行随机性检验，本文对原始随机序列与经过Toeplitz后处理的随机序列进行了正负自相关系数的对比，如图7所示。图7(a)为原始数据的自相关图，图7(b)为经过Toeplitz后处理的数据的自相关图。可以看出，经过Toeplitz后处理，自相关系数较处理前变小，自相关性更小，符合理想真随机数的特征。

5　结　论

本文通过条件最小熵对混沌激光随机数发生器产生的随机性进行了定量的评估。以自相关性好、幅值分布对称的白混沌作为混沌激光熵源，搭建了随机数产生装置。通过实验，获得了探测过程中噪声对理想混沌信号的偏移，计算了熵源的最小随机性含量。在此基础上分析并讨论了经典技术噪声偏移e和*CCNR*对输出随机性的影响。当*ADC*分辨位数n一定，原始数据中的熵含量会随着*CCNR*的

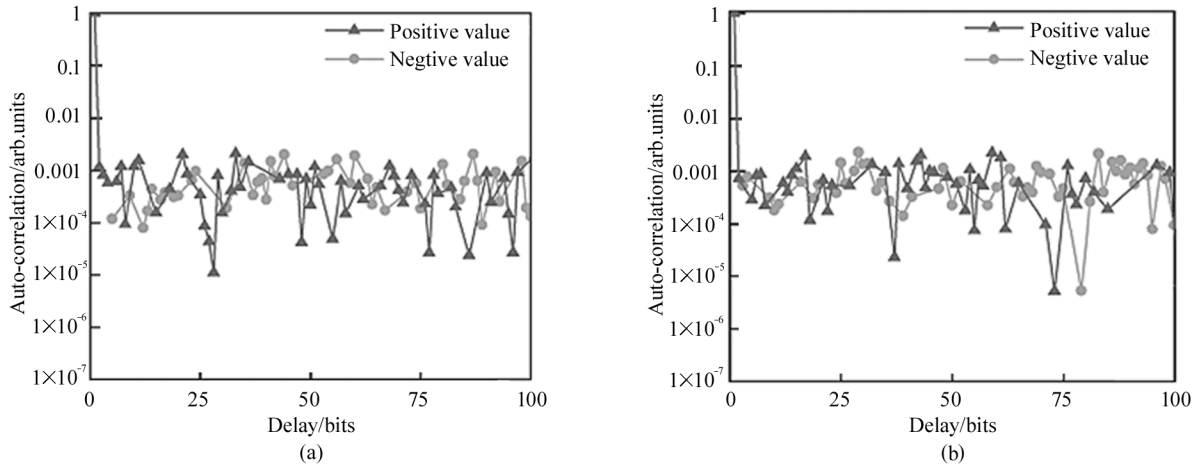

图 7　正负自相关系数对比图:(a) 原始数据的自相关; (b) 经 Toeplitz 后处理的数据的自相关

Fig. 7　Comparison of positive and negative auto-correlation coefficients: (a) Auto-correlation of the raw data; (b) Auto-correlation of the data after Toeplitz post-processing

增大而增长。只有 *CCNR* 比较高,才能够获得足够的真随机性。因此,在基于混沌激光的随机数发生器中,选择更大 *CCNR* 的混沌光信号可以确保随机性含量更高的真随机数产生,从而增加安全性。在条件最小熵评估后,可获得 8 位原始随机数中的熵含量为6.7位,通过Toeplitz后处理,实现了信息论可证安全的随机数产生。

参考文献:

[1] YU F, LI L, TANG Q, et al. A survey on true random number generators based on chaos[J]. Discrete Dynamics in Nature and Society, 2019, 2019(1): 1-10.

[2] DATCU O, MACOVEI C, HOBINCU R. Chaos based cryptographic pseudo-random number generator template with dynamic state change[J]. Applied Sciences, 2020, 10(2): 451.

[3] JIN J, LUO M, GONG Y H. Design and implementation of a true random number generator based on MOSFET thermal noise[J]. Microelectronics and Computer, 2015, 32(10): 7-11.
金杰,罗敏,宫月红.一种基于热噪声的真随机数发生器的设计与实现[J].微电子学与计算机,2015,32(10):7-11.

[4] STOJANOVSKI T, PIHL J, KOCAREV L. Chaos-based random number generators. Part II: practical realization[J]. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 2001, 48(3): 382-385.

[5] DENG H, JIN R H, CHEN J, et al. Oscillator-based high performance truly random number generator[J]. Research and Progress of SSE, 2007, 27(3): 391-396.
邓焕,金荣华,陈俊,等.基于振荡器的高性能真随机数发生器[J].固体电子学研究与进展,2007,27(3):391-396.

[6] YAN Q R, ZHAO B S, LIU Y A, et al. Optical quantum random number generator based on the time randomness of single-photon pulse[J]. Acta Optica Sinica, 2012, 32(3): 287-294.
鄢秋荣,赵宝升,刘永安,等.基于单光子脉冲时间随机性的光量子随机源[J].光学学报,2012,32(3):287-294.

[7] NIE Y Q, HUANG L, LIU Y, et al. The generation of 68 Gbps quantum random number by measuring laser phase fluctuations[J]. Review of Scientific Instruments, 2015, 86(6): 063105.

[8] LIU J, YANG J, LI Z, et al. 117 Gbits/s quantum random number generation with simple structure[J]. IEEE Photonics Technology Letters, 2016, 29(3): 283-286.

[9] YANG J, LIU J, SU Q, et al. 5.4 Gbps real time quantum random number generator with simple implementation [J]. Optics Express, 2016, 24(24): 27475-27481.

[10] LI L, WANG A, LI P, et al. Random bit generator using delayed self-difference of filtered amplified spontaneous emission[J]. IEEE Photonics Journal, 2014, 6(1): 1-9.

[11] UCHIDA A, AMANO K, INOUE M, et al. Fast physical ran-

dom bit generation with chaotic semiconductor lasers[J]. Nature Photonics, 2008, 2(12): 728-732.

[12] LI X Z, CHAN S C. Random bit generation using an optically injected semiconductor laser in chaos with oversampling[J]. Optics Letters, 2012, 37(11): 2163-2165.

[13] 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[J]. Optics Express, 2015, 23(26): 33130-33141.

[14] WANG A, WANG B, LI L, et al. Optical heterodyne generation of high-dimensional and broadband white chaos[J]. IEEE Journal of Selected Topics in Quantum Electronics, 2015, 21(6): 531-540.

[15] WANG A, WANG L, LI P, et al. Minimal-post-processing 320-Gbps true random bit generation using physical white chaos[J]. Optics Express, 2017, 25(4): 3153-3164.

[16] WANG L, GUO Y, LI P, et al. White-chaos radar with enhanced range resolution and anti-jamming capability[J]. IEEE Photonics Technology Letters, 2017, 29(20): 1723-1726.

[17] RUKHIN A. NIST statistical tests suite[EB/OL]. (2014-07-01)[2020-05-24]. http://www.scitech.people.com.cn/GB/1057/4017988.html.

[18] MA X, XU F, XU H, et al. Postprocessing for quantum random-number generators: Entropy evaluation and randomness extraction[J]. Physical Review A, 2013, 87(6): 062327.

作者简介：

李 璞 (1986—)，男，博士，教授，博士生导师，主要从事超快物理随机数产生及其应用方面的研究.