Two approaches for ultrafast random bit generation based on the chaotic dynamics of a semiconductor laser

Nianqiang Li, Byungchil Kim, V. Chizhevsky, A. Locquet, M. Bloch, D. Citrin, Wei Pan

Two approaches for ultrafast random bit generation based on the chaotic dynamics of a semiconductor laser

Nianqiang Li,[3,1] Byungchil Kim,[1,4] V. N. Chizhevsky,[2,*] A. Locquet,[4,1] M. Bloch,[1,4] D. S. Citrin,[1,4,5] and Wei Pan[3]

[1]*Georgia Institute of Technology, School of Electrical and Computer Engineering, Atlanta, Georgia 30332-0250 USA*

[2]*B. I. Stepanov Institute of Physics, National Academy of Science of Belarus, 220072 Minsk, Belarus*

[3]*Centers for Information Photonics and Communications, Southwest Jiaotong University, Chengdu, 610031 China*

[4]*UMI 2958 Georgia Tech-CNRS, Georgia Tech Lorraine, 2 Rue Marconi F-57070, Metz, France*

[5]*david.citrin@ece.gatech.edu*

[*]*vnc@dragon.bas-net.by*

Abstract: This paper reports the experimental investigation of two different approaches to random bit generation based on the chaotic dynamics of a semiconductor laser with optical feedback. By computing high-order finite differences of the chaotic laser intensity time series, we obtain time series with symmetric statistical distributions that are more conducive to ultrafast random bit generation. The first approach is guided by information-theoretic considerations and could potentially reach random bit generation rates as high as 160 Gb/s by extracting 4 bits per sample. The second approach is based on pragmatic considerations and could lead to rates of 2.2 Tb/s by extracting 55 bits per sample. The randomness of the bit sequences obtained from the two approaches is tested against three standard randomness tests (ENT, Diehard, and NIST tests), as well as by calculating the statistical bias and the serial correlation coefficients on longer sequences of random bits than those used in the standard tests.

OCIS codes: (140.1540) Chaos; (190.3100) Instabilities and chaos; (140.5960) Semiconductor lasers; (060.0060) Fiber optics and optical communications.

References and links

1. C. Masoller and N. B. Abraham, "Stability and dynamical properties of the coexisting attractors of an external cavity semiconductor laser," Phys. Rev. A **57**(2), 1313–1322 (1998).
2. J. Mørk, J. Mark, and B. Tromborg, "Route to chaos and competition between relaxation oscillations for a semiconductor laser with optical feedback," Phys. Rev. Lett. **65**(16), 1999–2002 (1990).
3. J. Ohtsubo, *Semiconductor Laser: Stability, Instability and Chaos* (Springer, 2008).
4. M. C. Soriano, J. Garcia-Ojalvo, C. R. Mirasso, and I. Fischer, "Complex photonics: Dynamics and applications of delay-coupled semiconductor lasers," Rev. Mod. Phys. **85**(1), 421–470 (2013).
5. A. Argyris, D. Syvridis, L. Larger, V. Annovazzi-Lodi, P. Colet, I. Fischer, J. García-Ojalvo, C. R. Mirasso, L. Pesquera, and K. A. Shore, "Chaos-based communications at high bit rates using commercial fibre-optic links," Nature **438**(7066), 343–346 (2005).
6. R. M. Nguimdo, P. Colet, L. Larger, and L. Pesquera, "Digital key for chaos communication performing time delay concealment," Phys. Rev. Lett. **107**(3), 034103 (2011).
7. F. Y. Lin and J. M. Liu, "Chaotic lidar," IEEE J. Sel. Top. Quantum Electron. **10**(5), 991–997 (2004).
8. L. Larger, M. C. Soriano, D. Brunner, L. Appeltant, J. M. Gutierrez, L. Pesquera, C. R. Mirasso, and I. Fischer, "Photonic information processing beyond Turing: an optoelectronic implementation of reservoir computing," Opt. Express **20**(3), 3241–3249 (2012).
9. A. Uchida, K. Amano, M. Inoue, K. Hirano, S. Naito, H. Someya, I. Oowada, T. Kurashige, M. Shiki, S. Yoshimori, K. Yoshimura, and P. Davis, "Fast physical random bit generation with chaotic semiconductor lasers," Nat. Photonics **2**(12), 728–732 (2008).
10. I. Reidler, Y. Aviad, M. Rosenbluh, and I. Kanter, "Ultrahigh-speed random number generation based on a chaotic semiconductor laser," Phys. Rev. Lett. **103**(2), 024102 (2009).

#199946 - $15.00 USD Received 22 Oct 2013; revised 14 Feb 2014; accepted 4 Mar 2014; published 14 Mar 2014

11. I. Kanter, Y. Aviad, I. Reidler, E. Cohen, and M. Rosenbluh, "An optical ultrafast random number generator," Nat. Photonics **4**(1), 58–61 (2010).
12. D. Knuth, *The Art of Computer Programming*, 3rd ed. (Addison Wesley Longman, 1998), Vol. 2.
13. N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, "Quantum cryptography," Rev. Mod. Phys. **74**(1), 145–195 (2002).
14. T. Honjo, A. Uchida, K. Amano, K. Hirano, H. Someya, H. Okumura, K. Yoshimura, P. Davis, and Y. Tokura, "Differential-phase-shift quantum key distribution experiment using fast physical random bit generator with chaotic semiconductor lasers," Opt. Express **17**(11), 9053–9061 (2009).
15. T. Durt, C. Belmont, L. P. Lamoureux, K. Panajotov, F. Van den Berghe, and H. Thienpont, "Fast quantum-optical random-number generators," Phys. Rev. A **87**(2), 022339 (2013).
16. C. R. S. Williams, J. C. Salevan, X. W. Li, R. Roy, and T. E. Murphy, "Fast physical random number generator using amplified spontaneous emission," Opt. Express **18**(23), 23584–23597 (2010).
17. H. Guo, W. Tang, Y. Liu, and W. Wei, "Truly random number generation based on measurement of phase noise of a laser," Phys. Rev. E Stat. Nonlin. Soft Matter Phys. **81**(5), 051137 (2010).
18. B. Qi, Y.-M. Chi, H.-K. Lo, and L. Qian, "High-speed quantum random number generation by measuring phase noise of a single-mode laser," Opt. Lett. **35**(3), 312–314 (2010).
19. X. Li, A. B. Cohen, T. E. Murphy, and R. Roy, "Scalable parallel physical random number generator based on a superluminescent LED," Opt. Lett. **36**(6), 1020–1022 (2011).
20. T. Stojanovski, J. Pihl, and L. Kocarev, "Chaos-based random number generators–Part II: practical realization," IEEE Trans. Circ. Syst. I Fundam. Theory Appl. **48**(3), 382–385 (2001).
21. J. Walker, Hotbits: Genuine Random Numbers, Generated by Radioactive Decay, http://www.fourmilab.ch/hotbits.
22. W. T. Holman, J. A. Connelly, and A. B. Dowlatabadi, "An integrated analog/digital random noise source," IEEE Trans. Circuits Syst. I **44**(6), 521–528 (1997).
23. J. T. Gleeson, "Truly random number generator based on turbulent electroconvection," Appl. Phys. Lett. **81**(11), 1949 (2002).
24. D. P. Rosin, D. Rontani, and D. J. Gauthier, "Ultrafast physical generation of random numbers using hybrid Boolean networks," Phys. Rev. E Stat. Nonlin. Soft Matter Phys. **87**(4), 040902 (2013).
25. X. Ma, F. Xu, H. Xu, X. Tan, B. Qi, and H. Lo, "Postprocessing for quantum random-numbers generators: Entropy evaluation and randomness extraction," Phys. Rev. A **87**(6), 062327 (2013).
26. V. N. Chizhevsky, "Fast generation of random bits based on polarization noises in a semiconductor vertical-cavity laser," Opt. Spectrosc. **111**(5), 689–694 (2011).
27. T. Harayama, S. Sunada, K. Yoshimura, P. Davis, K. Tsuzuki, and A. Uchida, "Fast nondeterministic random-bit generation using on-chip chaos lasers," Phys. Rev. A **83**(3), 031803 (2011).
28. A. Argyris, S. Deligiannidis, E. Pikasis, A. Bogris, and D. Syvridis, "Implementation of 140 Gb/s true random bit generator based on a chaotic photonic integrated circuit," Opt. Express **18**(18), 18763–18768 (2010).
29. J. Z. Zhang, Y. C. Wang, M. Liu, L. G. Xue, P. Li, A. B. Wang, and M. J. Zhang, "A robust random number generator based on differential comparison of chaotic laser signals," Opt. Express **20**(7), 7496–7506 (2012).
30. P. Li, Y. C. Wang, and J. Z. Zhang, "All-optical fast random number generator," Opt. Express **18**(19), 20360–20369 (2010).
31. R. M. Nguimdo, G. Verschaffelt, J. Danckaert, X. Leijtens, J. Bolk, and G. Van der Sande, "Fast random bits generation based on a single chaotic semiconductor ring laser," Opt. Express **20**(27), 28603–28613 (2012).
32. N. Oliver, M. C. Soriano, D. W. Sukow, and I. Fischer, "Dynamics of a semiconductor laser with polarization-rotated feedback and its utilization for random bit generation," Opt. Lett. **36**(23), 4632–4634 (2011).
33. N. Oliver, M. C. Soriano, D. W. Sukow, and I. Fischer, "Fast random bit generation using a chaotic laser: approaching the information theoretic limit," IEEE J. Quantum Electron. **49**(11), 910–918 (2013).
34. T. Harayama, S. Sunada, K. Yoshimura, J. Muramatsu, K. Arai, A. Uchida, and P. Davis, "Theory of fast nondeterministic physical random-bit generation with chaotic lasers," Phys. Rev. E Stat. Nonlin. Soft Matter Phys. **85**(4), 046215 (2012).
35. T. Mikami, K. Kanno, K. Aoyama, A. Uchida, T. Ikeguchi, T. Harayama, S. Sunada, K.-i. Arai, K. Yoshimura, and P. Davis, "Estimation of entropy rate in a fast physical random-bit generator using a chaotic semiconductor laser with intrinsic noise," Phys. Rev. E Stat. Nonlin. Soft Matter Phys. **85**(1 Pt 2), 016211 (2012).
36. X. Fang, B. Wetzel, J. Merolla, J. M. Dudley, L. Larger, C. Guyeux, and J. M. Bahi, "Noise and chaos contributions in fast random bit sequence generated from broadband optoelectronic entropy sources," IEEE Trans. Circuits Syst. I **99**, 1–14 (2013).
37. K. Hirano, T. Yamazaki, S. Morikatsu, H. Okumura, H. Aida, A. Uchida, S. Yoshimori, K. Yoshimura, T. Harayama, and P. Davis, "Fast random bit generation with bandwidth-enhanced chaos in semiconductor lasers," Opt. Express **18**(6), 5512–5524 (2010).
38. J. G. Wu, X. Tang, Z. M. Wu, G. Q. Xia, and G. Y. Feng, "Parallel generation of 10 Gbits/s physical random number streams using chaotic semiconductor lasers," Laser Phys. **22**(10), 1476–1480 (2012).
39. X. Z. Li and S. C. Chan, "Random bit generation using an optically injected semiconductor laser in chaos with oversampling," Opt. Lett. **37**(11), 2163–2165 (2012).
40. X. Z. Li and S. C. Chan, "Heterodyne random bit generation using an optically injected semiconductor laser in chaos," IEEE J. Quantum Electron. **49**(10), 829–838 (2013).

#199946 - $15.00 USD Received 22 Oct 2013; revised 14 Feb 2014; accepted 4 Mar 2014; published 14 Mar 2014

41. Y. Akizawa, T. Yamazaki, A. Uchida, T. Harayama, S. Sunada, K. Arai, K. Yoshimura, and P. Davis, "Fast random number generation with bandwidth-enhanced chaotic semiconductor lasers at $8\times 50\,\mathrm{Gb/s}$," IEEE Photon. Technol. Lett. **24**(12), 1042–1044 (2012).
42. http://people.seas.harvard.edu/~salil/pseudorandomness.
43. C. E. Shannon, "A mathematical theory of communication," Bell System Technical Journal **27**, 379–423 and 623–656 (1948).
44. M. A. Wayne and P. G. Kwiat, "Low-bias high-speed quantum random number generator via shaped optical pulses," Opt. Express **18**(9), 9351–9357 (2010).
45. K. Hirano, K. Amano, A. Uchida, S. Naito, M. Inoue, S. Yoshimori, K. Yoshimura, and P. Davis, "Characteristics of fast physical random bit generation using chaotic semiconductor lasers," IEEE J. Quantum Electron. **45**(11), 1367–1379 (2009).
46. T. Yamazaki and A. Uchida, "Performance of random number generators using noise-based superluminescent diode and chaos-based semiconductor lasers," IEEE J. Sel. Top. Quantum Electron. **19**(4), 0600309 (2013).
47. V. N. Chizhevsky, "Symmetrization of single-sided or nonsymmetrical distributions: The way to enhance a generation rate of random bits from a physical source of randomness," Phys. Rev. E Stat. Nonlin. Soft Matter Phys. **82**(5), 050101 (2010).
48. T. E. Murphy and R. Roy, "Chaotic lasers: The world's fastest dice," Nat. Photonics **2**(12), 714–715 (2008).
49. J. Z. Zhang, Y. C. Wang, L. G. Xue, J. Y. Hou, B. B. Zhang, A. B. Wang, and M. J. Zhang, "Delay line length selection in generating fast random numbers with a chaotic laser," Appl. Opt. **51**(11), 1709–1714 (2012).
50. A. B. Wang, P. Li, J. G. Zhang, J. Z. Zhang, L. Li, and Y. C. Wang, "4.5 Gbps high-speed real-time physical random bit generator," Opt. Express **21**(17), 20452–20462 (2013).
51. I. Kanter, Y. Aviad, I. Reidler, E. Cohen, and M. Rosenbluh, "Towards the generation of random bits at terahertz rates based on chaotic semiconductor laser," J. Phys. Conf. Ser. **233**, 012002 (2010).
52. D. Rontani, A. Locquet, M. Sciamanna, and D. S. Citrin, "Loss of time-delay signature in the chaotic output of a semiconductor laser with optical feedback," Opt. Lett. **32**(20), 2960–2962 (2007).
53. S. Priyadarshi, Y. Hong, I. Pierce, and K. A. Shore, "Experimental investigations of time-delay signature concealment in chaotic external-cavity VCSELs subject to variable optical polarization angle of feedback," IEEE J. Sel. Top. Quantum Electron. **19**(4), 1700707 (2013).
54. J. Walker, Ent-a pseudorandom sequence test program, http://www. fourmilab.ch/random.
55. G. Marsaglia, The diehard test suite (2003), http:// www. Csis. hku. hk/diehard.
56. A. Rukhin, J. Soto, J. Nechvatal, M. Smid, E. Barker, S. Leigh, M. Levenson, M. Vangel, D. Banks, A. Heckert, J. Dary, and S. Vo, "A statistical test suite for the validation of random number generators and pseudo random number generators for cryptographic applications", *http://csrc.nist.gov/groups/ST/toolkit/rng/documentation_software.html*.
57. T. Granlund, *et al*., GMP, the GNU multiple precision arithmetic library. *http://gmplib.org* (2013).

1. Introduction

Semiconductor lasers are very sensitive to perturbations from the outside environment [1, 2]. Even very weak optical feedback may significantly increase both the intensity noise and lasing linewidth, which is undesirable in most applications. Consequently, many conventional laser-diode systems are commonly prepared with optical isolators to impede feedback from surface reflections. Recently, however, in view of the technological importance of these devices, the rich nonlinear dynamics of semiconductor lasers with delayed feedback/injection have been widely investigated [3, 4]. Among them, a form of chaotic dynamics, termed *coherence collapse*, has been exploited for several applications, such as chaos-based communications [5, 6], chaotic lidar/radar [7], reservoir computing [8], and chaos-based random bit generation (RBG) [9–11]. Our focus here is on chaotic laser-diodes for RBG.

Generally speaking, there exist two approaches to generate random bits. The first is based on deterministic mathematical algorithms and is known as *pseudo* RBG [12]; unfortunately, the quality and generation rate may be inadequate, for applications such as cryptography and large-scale Monte Carlo numerical computations in which it is crucial to have nearly unpredictable bits [13, 14]. The second approach consists in extracting randomness from physical phenomena, such as radioactivity, noise, turbulence, and electrical chaos in nonlinear circuits, and is known as *physical* RBG [15–25]; the generation rate of most physical random number generators is typically on the order of a couple of Gb/s or lower [26]. Following the first demonstration of physical RBG based on broadband optical chaos [9], chaotic semiconductor lasers have gained interest as physical sources of randomness because of the high potential generation rate and the ease of implementation [10, 11, 27–33]. It should also be mentioned that there is debate about the role of noise in chaotic systems with regard to

#199946 - $15.00 USD Received 22 Oct 2013; revised 14 Feb 2014; accepted 4 Mar 2014; published 14 Mar 2014

RBG [34–36]. Progress has recently been made in improving the generation rate using high-order derivatives [11], coupled chaotic lasers [37–40], as well as advanced post-processing techniques, such as the bit-order-reversed method [41]. Experimental implementations have been demonstrated based on photonic integrated circuits [27, 28], as well as numerical implementation using various schemes, such as RBG based on all-optical components [30] and a single external-cavity semiconductor laser (ECSL) [34]. Higher generation rates have been obtained thanks to large sampling rates and to complex post-processing that artificially increase the allowed number of retained bits from each high-resolution analog-to-digital converted sample. As an example, Kanter *et al.* have demonstrated ultrafast chaos-based RBG with a rate of 300 Gb/s based on the use of the high-order derivatives using a sampling rate of 20 GHz and extracting 15 random bits per sample, although the raw data were digitized with 8-bit resolution [11]. Very recently, Oliver *et al.* have highlighted that the fundamental limits of maximum RBG rate imposed by information theory depend on the analog bandwidth of the chaotic laser and cannot be improved by simply increasing the sampling rate or by creating additional bits through post-processing [33]. In fact, in earlier works, the number of bits extracted from per sample from the chaotic laser intensity has been chosen heuristically. In [31, 33, 37], this number has been estimated by plotting histograms of the truncated values until a flat histogram, [i.e., probability density distribution (PDF)], is obtained. The heuristics were validated by testing the generated bit stream against existing statistical tests of randomness. Nevertheless, intuitively, to extract m almost-uniform and independent bits from a source, the source should initially contain at least "m bits of randomness" in it; this can be formally quantified with information-theoretic measures, such as the Shannon entropy and min-entropy [33, 42–44]. In particular, if one were to extract more than the number of quantization bits from each sample, the generator should be considered as a pseudo random number generator based on physical phenomena or called a physical-based pseudo random number generator, rather than a true physical random number generator.

The majority of existing schemes for fast chaos-based RBG in the literature require bitwise exclusive-or (XOR) operation to amplify inherent randomness to remove residual correlations, in addition to the selection of several least significant bits (LSBs) [9, 29, 31, 38–41, 45]. In this paper, we explore two approaches to fast RBG based on the processing of the chaotic laser intensity time series of an ECSL. In the first approach, the number of retained bits is conservatively selected according to the limits set by information theory. In the second approach, the number of extracted random bits is merely chosen to pass standard randomness tests. Accordingly, we name the first approach *physical RBG*, and the second *physical-based pseudo RBG*. Proving that our conservative first approach does indeed ensure information-theoretic RNG requires an in-depth analysis of our post-processing, which is beyond the scope of the present paper and will be the subject of a subsequent work. Some previous works show that the interplay of dynamical properties, acquisition conditions, and post-processing plays a critical role in the performance of RBG [33, 45, 46]. We apply the procedure described in [47] which is based on the calculation of the high-order finite differences (HFD) of the initial data, which results in symmetric statistical distributions. It must be noted that this method is similar to the independently developed high-order derivatives method presented in the pioneering [11]. The post-processing proposed in [47] and [11] reduces the dependence on the dynamical properties and acquisition conditions at the expense of additional numerical processing. For this reason, the strict adjustment of the feedback strength, the injection current, and the external-cavity length is not needed. Throughout the study, the post-processing employed only includes the calculation of HFD and the selection of LSBs. Moreover, unlike many existing reports of chaos-based RBG [9–11, 27–33, 37–41], the randomness in the present experimental study is not only verified by three standard randomness tests, for which the required size of the sequence of random bits is only of the order of 1 Gbit, but is also checked by calculating the statistical bias and correlation coefficient, for very long (10-50 Gbits) random bit sequences.

#199946 - $15.00 USD Received 22 Oct 2013; revised 14 Feb 2014; accepted 4 Mar 2014; published 14 Mar 2014

This paper is organized as follows. Section 2 briefly describes the experimental setup employed for ultrafast RBG, which is a semiconductor laser subject to time-delayed optical feedback. Section 3 presents the experimental results; we first focus on physical RBG and then investigate physical-based pseudo RBG. Section 4 provides concluding remarks.

2. Experimental setup

The schematic diagram of the experimental setup for RBG is shown in Fig. 1. The setup consists of a semiconductor laser with optical feedback, which generates intensity chaos and provides the physical source, and of a post-processing unit, which extracts randomness. This processing is performed offline as in most previous studies [31–33, 37, 38].

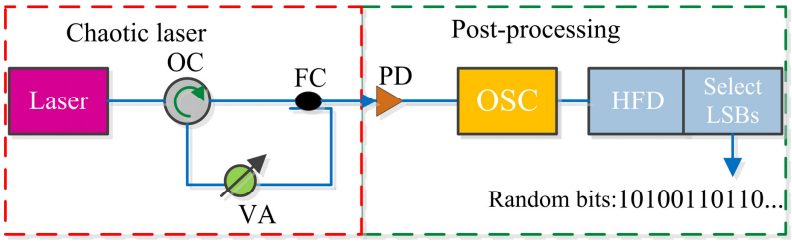

Fig. 1. Schematic diagram of ultrafast RBG based on optical chaos. Laser: a distributed feedback (DFB) laser diode; OC: optical circulator; VA, variable optical attenuator; FC, 85: 15 fiber coupler; PD, photodiode; OSC, 40 GHz real-time oscilloscope; HFD, high-order finite differences; LSBs, least significant bits.

2.1 Chaos generation

The generation of complex chaotic signals is realized by perturbing a semiconductor laser using an external-cavity to feed light back into the gain medium. The semiconductor laser used in our experiments is an intrinsically single-longitudinal mode DFB laser that operates at a nominal wavelength of $\lambda = 1550$ nm and has a threshold current of $I_{th} = \sim 10$ mA . The output light is fed back into the laser facet after it passes through a fiber ring cavity that consists of an optical circulator (OC), a variable optical attenuator (VA), and a fiber coupler (FC). The DFB laser is easily destabilized by adjusting the VA to change the feedback power. The laser is driven by an ultra-low-noise current source (ILX-Lightwave, LDX-3620B) and controlled by a thermoelectric controller (ILX-Lightwave, LDT-5412).

The laser was pumped well above threshold to ensure operation in the coherence collapse regime: the injection current was set to ~20 mA so that the chaotic fluctuations dominate the background noise. The external-cavity formed by the abovementioned components results in a large roundtrip delay time of 57.68 ns. The feedback strength, defined as the ratio of the optical feedback power fed back into the laser to the laser power in the absence of feedback was set to ~8%. As we will not focus on the dynamical properties, finding the optimum operating conditions for the RBG by a careful adjustment of the feedback strength or the injection current is beyond the scope of the present work. However, we stress that one can optimize the statistical properties of chaotic dynamics by carefully controlling the feedback level.

2.2 Post-processing

The dynamics of the laser intensity were detected at a photodiode (PD), and then acquired by a real-time oscilloscope (LeCroy WaveMaster 813zi, 13 GHz bandwidth). The experimental data was digitized by the 8-bit analog-to-digital converter (ADC) in the oscilloscope. As mentioned above, the post-processing employed for the chaos-based RBG consists of the calculation of the HFD and followed by the retention of the LSB. For the first approach, we point out here that we do not extract more bits than the limit set by information theory,

#199946 - $15.00 USD Received 22 Oct 2013; revised 14 Feb 2014; accepted 4 Mar 2014; published 14 Mar 2014

leading to physical RBG; for the second approach, we attempt to extract as many bits for each sample as possible, so as to realize ultra-high speed physical-based pseudo RBG.

In our experiments, unlike many previous optical RBG schemes summarized in the introduction, we choose a high sampling rate of 40 GHz. The chaos bandwidth obtained from such an ECSL is on the order of several GHz, and the bandwidth of the oscilloscope is 13 GHz. The detection bandwidth and sampling rate are sufficient to capture all relevant dynamics in the intensity time series. In addition, our two schemes of chaos-based RBG only involve limited post-processing. Moreover, as also explained in [10, 11], a single laser is necessary contrary to [9, 38–40, 45, 48], and contrary to [38, 49, 50], we do not need to use and adjust a delay line to generate a second, uncorrelated, bit stream from the measured stream. Thanks to the use of HFD, the rigorous, and experimentally difficult, threshold voltage adjustment, as well as the selection of the retained range of output intensities, described in [9, 45, 46] are also avoided.

3. Experimental results

3.1 Statistical properties of laser chaos

True random bit sequences lack any pattern in their appearance and, in principle, are completely unbiased and unpredictable. The quality of the generated random bits depends upon the statistical properties of the sources of randomness.

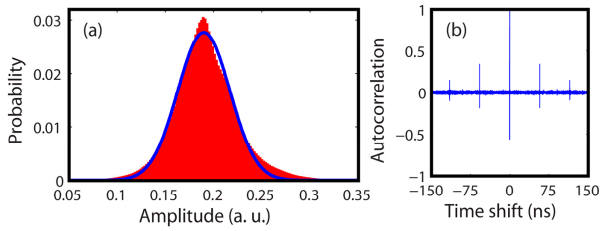

Fig. 2. Statistical properties of laser chaos. (a) PDF and (b) autocorrelation function of the chaotic waveforms. The blue curve in (a) denotes the fitted Gaussian.

Entropy sources with a close to uniform distribution are good candidates for high-quality RBG; however, the physical sources known to us do not have an ideally symmetric PDF, to say nothing of being uniformly distributed at each digitization level. For example, Fig. 2(a) presents the calculated PDF for the chaotic laser intensity generated by the experimental setup displayed in Fig. 1. The PDF is directly obtained from 3×10^7 digitized samples. Although the PDF resembles a Gaussian distribution, the asymmetry of the PDF is clearly identified by comparing it with the fitted Gaussian [blue curve in Fig. 2(a)], which is a common feature of chaotic semiconductor lasers. In our case, the coefficient of skewness and kurtosis are about 0.27 and 4, respectively, which substantially deviate from a Gaussian distribution. This means that keeping all LSBs of the 8-bit data cannot pass the statistical tests of randomness since the non-uniformity of the initial distribution influences the randomness of the bits generated. Moreover, neither are consecutive bits independent. The autocorrelation trace for the full 8-bit signal is shown in Fig. 2(b). It can be clearly seen that peaks appear at integer multiples of the delay time, *e.g.,* a pronounced peak with a correlation coefficient about 0.34 is located at a delay time of 57.68 ns, which corresponds to the roundtrip time in the external-cavity. Therefore, to extract random bits from the chaotic sources, post-processing techniques should be employed.

#199946 - $15.00 USD Received 22 Oct 2013; revised 14 Feb 2014; accepted 4 Mar 2014; published 14 Mar 2014

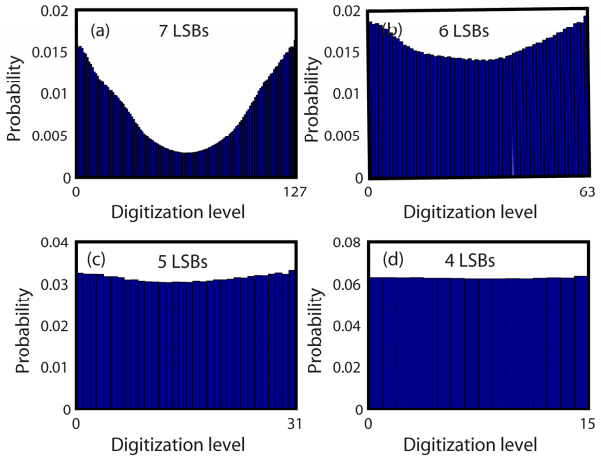

Fig. 3. PDFs over the range of digitization levels for (a) 7 LSBs, (b) 6 LSBs, (c) 5 LSBs, and (d) 4 LSBs retained from each 8-bit sample.

Selection of m LSBs is a common, simple post-processing procedure for improving the uniformity of the bit distributions and for destroying the residual correlations in the original dynamics [31–33, 37–40, 51]. Here, this technique is applied to our experimental signal. Figures 3(a)-3(d) show the distributions for $m = 7$ down to 4, respectively. In Fig. 3(a) only the most significant bit (MSB, $m = 7$) is excluded and the corresponding PDF significantly differs from a uniform distribution. When more MSBs are discarded the uniformity of the PDF is improved, as shown in Figs. 3 (b) and 3(c). Furthermore, if one selects only 4 LSBs ($m = 4$) for each 8-bit sample, the resulting PDF is close to uniform, as is illustrated in Fig. 3(d). In addition, the residual correlations are gradually eliminated as the number of discarded MSBs increases even though the original dynamics exhibit obvious short-term correlation. For this reason, the signature of the feedback delay time is not important for RBG when such post-processing is introduced. We will explore the residual correlation feature for our chaos-based RBG in the following sections.

Note that in the experiment, for the fixed injection current and feedback strength given above, random bits obtained by keeping the 4 LSBs of the 8-bit data cannot pass all the statistical tests of randomness. This is because there still exist significant biases or correlations in the generated random bits. To further distill the randomness, as mentioned above, additional post-processing is needed. In this study, we therefore only employ the selection of m LSBs as the second post-processing step after calculating the HFD of the experimental signal.

3.2 High-quality physical RBG

In order to achieve efficient generation of random bits, we adopt the *n*th-order finite differences (HFD) procedure described in [47], which allows one to extract random bits from any source of randomness with nonsymmetric distribution. A similar method for increasing the speed of chaos-based RBG was previously discussed in [10, 11]. To generate random bit sequences, we need not choose the optimal conditions regarding the chaotic dynamics and acquisition process. This is because the calculation of HFD reduces dependence on these properties. The working point of the laser is fixed as given in Sec. 2, unless otherwise stated. In the HFD method, the generation of random bit sequence consists of the following two steps. The first step involves collecting N integer samples a_k ($k = 1, 2, ...N$) with 8-bit resolution from the measured output intensities of a chaotic laser. They are transformed into

#199946 - $15.00 USD Received 22 Oct 2013; revised 14 Feb 2014; accepted 4 Mar 2014; published 14 Mar 2014

floating-point numbers with M-bit resolution ($M >> 8$), and then the nth-order differences of the floating-point data are calculated. It is worth noting that one cannot aim to extract all significant bits from each obtained sample since strong correlations appear in the bit stream after the calculation of HFD. Therefore, one should discard certain MSBs to eliminate these correlations. To this end, in the second step, one just retains m LSBs of each new variable in the floating-point representation and consecutively concatenates all the retained bits to obtain a long random bit sequence.

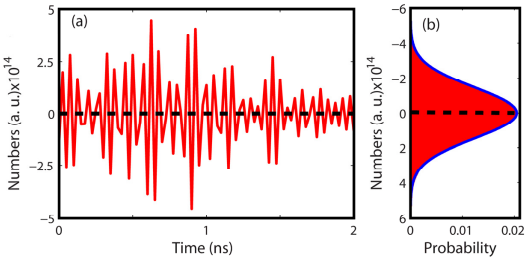

Fig. 4. (a) Characteristics of the calculated floating-point numbers based on 50th-order finite differences. (a) Time series and (b) its associated statistical distribution. The dashed lines stand for the mean value of the time series. The blue line in (b) represents a fitted perfect Gaussian distribution.

As a point of comparison, we evaluate the potential rate of randomness extraction in the first approach, by properly quantifying the amount of randomness in the obtained sequences $\{a_k\}$. While, the Shannon entropy and min-entropy may have operational information-theoretic meaning in the present context [42] we opt to evaluate the min-entropy, which is the most conservative way to measure then randomness that could be extracted from a discrete random variable [25, 44].

The min-entropy is defined as $H_\infty = -\log\{\max(P_i)\}$ with $\max(P_i)$ being the probability of the most likely event [42] and P_i is defined to be the probability of the ith value of the alphabet of the discrete variable. In our experiments, the real-time oscilloscope recorded 10^7 samples at a time of 8-bit data at 40 GHz sampling rate; We calculated the min-entropy values of 200 sets of 10^7 samples and then made an average over 200 values of min-entropy. For our experimental data, the final min-entropy is ~4.4 bits per sample, which indicates that, even the best method for extracting randomness cannot lead to more than 4.4 information-theoretically random bits from each raw 8-bit resolution sample. In addition, random bit sequences obtained by directly keeping 4 LSBs per sample cannot pass the entire statistical randomness tests because the generated random bit sequences may still contain certain biases and correlations, which can stem for example from long term correlations associated with the external-cavity length of the chaotic laser [52, 53] as illustrated in Fig. 2(b). (Thus, in the first approach, the technique of calculating nth-order finite differences is utilized.) Firstly, all the random integers a_k of 8-bit resolution obtained from our oscilloscope were transformed into floating-point numbers according to the procedure described in [26, 47]. Secondly, the HFD of these numbers were calculated. Specifically, the maximal number of bits that represents each sample is chosen to be 52, since numbers represented in the double format have a maximum precision of 52 bits. The maximal difference order $n_{\max}$ allowed for the experimental data can be determined by the condition (1) in [47], where the author demonstrated the violation of that condition leads to appearance of strong correlations in the generated bit stream. However, in our case, we aim to extract only 4 LSBs for each sample to avoid exceeding the min-entropy. The first step of our procedure is to calculate the 50th-order finite differences of a sequence of the floating-point numbers. It is

#199946 - $15.00 USD Received 22 Oct 2013; revised 14 Feb 2014; accepted 4 Mar 2014; published 14 Mar 2014

worth noting that the use of lower-order differences leads to 4 LSB-based random sequences that pass the standard statistical tests (as mentioned below) but not the three-standard-deviation criteria. This highlights the beneficial role of HFD post-processing in the extraction of randomness, as well as the limitations of the three tests cited above. Figure 4(a) presents the fluctuation evidence for such floating-point numbers. One can see that these numbers symmetrically fluctuate around their mean. As a result, a highly symmetric distribution is obtained with a coefficient of skewness of the order of $\approx 10^{-7}$, as shown in Fig. 4(b), where a perfect Gaussian distribution fit is also depicted, indicating that one can achieve RBG based on the data after this post-processing. Before extracting random bits, all new a_k after calculating the HFD were transformed into positive numbers by adding 2^{51} and mapped into the binary format. Finally, the random bits obtained by keeping 4 LSBs for each sample after the use of HFD have a distribution that is significantly closer to uniform than that of Fig. 3(d). Moreover, the serial correlation coefficients are also significantly closer to zero, as will be shown below. In that regard, we consider that the use of HFD combined with a restriction to the number of LSBs corresponding to the min-entropy may have the potential to generate information-theoretic random bits at a rate of 160 Gb/s ($= 4$ LSBs $\times 40$ GHz) .

Table 1. Results of NIST statistical tests for physical random bits. The results have been performed using 1000 samples of 1 Mbit data and a significance level $\alpha = 0.01$, for "Success", the P value (uniformity of p values) should be larger than 0.0001 and the proportion should be in the 0.99 ± 0.0094392 range [56]. For the tests that produce multiple P values and proportions, the worst case is shown.

Statistical test	P value	Proportion	Result
Frequency	0.450297	0.9890	Success
Block frequency	0.610070	0.9910	Success
Cumulative sums	0.211144	0.9890	Success
Runs	0.911413	0.9830	Success
Longest runs	0.406499	0.9900	Success
Rank	0.234373	0.9910	Success
FFT	0.701366	0.9900	Success
Nonoverlapping templates	0.302058	0.9850	Success
Overlapping templates	0.090388	0.9870	Success
Universal	0.812905	0.9850	Success
Approximate entropy	0.308561	0.9900	Success
Random excursions	0.137487	0.9888	Success
Random excursions variant	0.164071	0.9888	Success
Serial	0.285427	0.9930	Success
Linear complexity	0.738534	0.9850	Success

We verified the quality of the generated bit sequences utilizing three collections of standard statistical tests. These are the pseudorandom bit sequence test program (ENT) [54], the Diehard tests [55], and the National Institute of Standards and Technology test suite (NIST Special Publication 800-22) [56].

First, a sequence of 1 Gbit length was tested in ENT program. The ENT results are as follows: Entropy = 1.000 000 bits per bit (the optimum compression would reduce the bit file by 0 percent), χ^2 distribution is 0.39 (randomly would exceed this value 50% of the times),

#199946 - $15.00 USD Received 22 Oct 2013; revised 14 Feb 2014; accepted 4 Mar 2014; published 14 Mar 2014

arithmetic mean value of data bits is 0.5000, Monte Carlo value for π is 3.142211186, and serial correlation coefficient is 0.000015. Second, we carried out the Diehard tests according to the description in [55], which outlines that the Diehard battery consists of 17 tests, and p-values obtained in each test (269 in total for a sequence of 2.5×10^9 bits used in our case) should not be 0 or 1 up to 6 decimal places and are supposed to be uniform in [0, 1). The overall p-value (p_0) characterizes the uniformity of this distribution. We obtained $p_0 = 0.669760$ [KS] with minimal $p = 0.0079$ and maximal $p = 0.9984$ from the total number of 269 (KS denotes the Kolmogorov-Smirnov test). Finally, the NIST tests were performed using 1000 instances of 1Mbit sequences and are shown in Table 1.

The bit sequences generated by the chaotic dynamics of a semiconductor laser with optical feedback at the bit rate of160 Gb/s passed all three standard randomness tests of ENT, Diehard, and NIST.

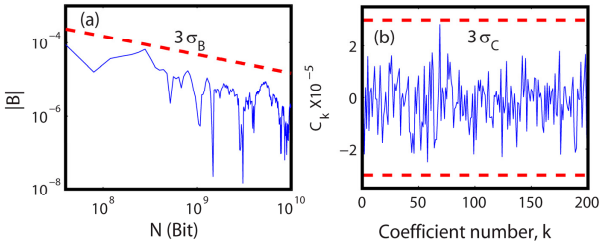

Fig. 5. The results of the statistical bias B and serial autocorrelation coefficient C_k for the physical RBG at a speed of 160 Gb/s. (a) B as a function of the number of generated bits N; (b) first 200 serial autocorrelation coefficients for the binary sequence of 10 Gbit length. The values of C_k were calculated by ensemble averaging over 1×10^4 sequences of 1 Mbit length each.

It is worth noting that the standard tests above operate on bit sequences that are limited to a couple of Gbits, such as 1 Gbit in ENT and NIST tests, as well as 2.5 Gbits in Diehard tests. It is important to check the statistical properties on longer sets that are more in line with the large bit rates obtained here. Therefore, to further evaluate the randomness of the long random-bit sequences, we calculate the statistical bias B and the serial autocorrelation coefficients C_k . B and C_k are defined as [47]

$$B = \langle b_i \rangle - 0.5\,, \tag{1}$$

$$C_k = \frac{\left\langle \left(b_i - \langle b_i \rangle\right)\left(b_{i+k} - \langle b_i \rangle\right)\right\rangle}{\left\langle \left(b_i - \langle b_i \rangle\right)^2 \right\rangle}, \tag{2}$$

where b_i ($i = 1, 2, \ldots$) takes the values "0" or "1", k is the delay in bits and the averaging $\langle \cdots \rangle$ is performed over the index i . Note that the statistical bias B and the serial autocorrelation coefficients C_k are random variables which vary from one sequence to another and depend on its length. For high-quality random bit sequence of length N , both the statistical bias B and the absolute value of the serial autocorrelation coefficients C_k should be smaller than the three-standard-deviations: $3\sigma_B = 1.5N^{-0.5}$ for B and $3\sigma_C = 3N^{-0.5}$ for C_k with a probability of 99.7% [12]. Figure 5(a) shows on a log-log scale the dependence of the statistical bias B on the sequence length N . It is apparent that the statistical bias always lies below the $3\sigma_B$ -criterion (dashed line). Figure 3(b) shows the first 200 autocorrelation coefficients for a

#199946 - $15.00 USD Received 22 Oct 2013; revised 14 Feb 2014; accepted 4 Mar 2014; published 14 Mar 2014

sequence with a length of 10 Gbits. One can see that, all values C_k are within the range that is bounded by dashed lines corresponding to the $3\sigma_C$ -criterion. Therefore, both criteria are well satisfied for the random bit sequence of 10 Gbits.

3.3 Ultrafast physical-based pseudo RBG

Let us now consider the randomness extraction in the second approach. Concretely, information theory shows us that if one samples at a rate R_s GHz and quantizes on M bits, the entropy rate of the resulting discrete-time stochastic process cannot exceed $M \cdot R_s$ b/s. As an example [11], reported a generation rate of 300 Gb/s with $R_s = 20$ GHz and $M = 8$ bits , which violates the above reasoning. Here, in this second approach, we aim to extract as many bits as possible from the chaotic dynamics of a single ECSL that pass the three standard tests and satisfy the three standard-deviation criteria. To this end, the calculation of HFD of the initial data was also implemented. In contrast to [26, 47], where all the random integers a_k were transformed into floating-point numbers of 52-bit resolution, in our implementation here the initial data of 8-bit resolution were transformed into a 64-bit integer data type ("int64"). Moreover, we found that the maximal difference order $n_{\max} = 58$ determined by condition (1) in [47] may be not the optimal one, and we obtained better performance for generating random bits when $n_{\max}$ was slightly increased to 62. Therefore, the 62th-order finite differences were used. The need for such a high order of difference may be caused by the sharpness and asymmetry of initial distribution, as shown in Fig. 2(a). All a_k were then transformed into positive numbers by adding 2^{62} and converted in binary.

Table 2. Results of NIST statistical tests for ultrafast physical-based pseudo random bits.

Statistical test	P value	Proportion	Result
Frequency	0.651693	0.9830	Success
Block frequency	0.275709	0.9900	Success
Cumulative sums	0.186566	0.9830	Success
Runs	0.079051	0.9850	Success
Longest runs	0.680755	0.9880	Success
Rank	0.587274	0.9930	Success
FFT	0.676615	0.9910	Success
Nonoverlapping templates	0.011709	0.9810	Success
Overlapping templates	0.074330	0.9850	Success
Universal	0.158133	0.9850	Success
Approximate entropy	0.552383	0.9870	Success
Random excursions	0.256333	0.9823	Success
Random excursions variant	0.184128	0.9871	Success
Serial	0.500279	0.9900	Success
Linear complexity	0.662091	0.9890	Success

Next, the randomness extraction was carried out. To reiterate, even though all numbers contain 64 bits, again, one cannot extract all significant bits from each of them because keeping the full bits results in appearance of strong correlations in the generated bit sequences. To eliminate these correlations several of the MSBs have to be discarded. One can

#199946 - $15.00 USD Received 22 Oct 2013; revised 14 Feb 2014; accepted 4 Mar 2014; published 14 Mar 2014

estimate the number of bits that need to be removed by plotting PDFs of the truncated values as is done in Fig. 3. Thanks to HFD, the removal of only a few bits leads to flat PDFs. Then one has to further check the ENT, Diehard, and NIST tests to determine the number of bits that should be retained. Based on extensive tests, we found that an extraction of 55 LSBs from each sample may pass all the statistical tests for randomness. We carried out the same procedure of the randomness tests as that for the physical RBG in the first approach. The ENT results are as follows: Entropy = 1.000 000 bits per bit (the optimum compression would reduce the bit file by 0 percent), χ^2 distribution is 0.27 (randomly would exceed this value 50% of the times), arithmetic mean value of data bits is 0.5000, Monte Carlo value for π is 3.141596402 and serial correlation coefficient is −0.000014. For the Diehard tests, it was obtained that $p_0 = 0.068578\,[\mathrm{KS}]$ with minimal $p = 0.0017$ and maximal $p = 0.9897$ from the total number of 269. Table 2 presents the NIST results.

For this second approach, the statistical bias B and the serial autocorrelation coefficients C_k of the random bit sequences were tested as well. As 55 LSBs were extracted from each sample, one can easily obtain a long bit-sequence from the chaos-based RBG. The bias was evaluated versus the length of the sequence N. It is clearly shown in Fig. 6(a) that the statistical bias B of the random bit sequence always keeps below the significance level of $3\sigma_B$ up to 50 Gbits. In the meantime, the first 200 serial autocorrelation coefficients C_k of 50 Gbits random bit stream is presented in Fig. 6(b). One can see that, all values C_k are kept within the range, which is bounded by dashed lines corresponding to the $3\sigma_C$ -criterion. These results further confirm the uncorrelatedness of the sequences, indicating that the chaos-based RBG can also be used to produce random bit sequences of ultra-long length at the same time the corresponding generation rate is increased up to the order of a couple of Tb/s.

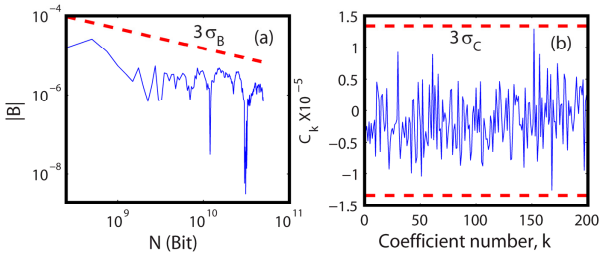

Fig. 6. The results of the statistical bias B and serial autocorrelation coefficient C_k for the physical-based pseudo RBG at a speed of 2.2 Tb/s. (a) B as a function of the number of generated bits N; (b) first 200 serial autocorrelation coefficients for the binary sequence of 50 Gbit length. The values of C_k were calculated by ensemble averaging over 5×10^4 sequences of 1 Mbit length each.

The results given above suggest that the generated bit sequences passed all the statistical tests for randomness, in the sense that the inclusion of 55 consecutive LSBs leads to RBG at a rate of 2.2 Tb/s ($= 55\ \mathrm{LSBs}\times 40\ \mathrm{GHz}$) , which is the fastest generation rate based on the chaotic dynamics of semiconductor lasers known to us. However, as it violates the limit set by information theory, we have to consider the corresponding generator as an ultrafast physical-based pseudo random number generator.

A further increase of the generation rate of random bits could possibly be achieved by a transformation of the initial data into a class of integers such as “int128” or “int256” (see, for instance, the GNU Multiple Precision Arithmetic Library [57]), which may result in a generation rate of 4 Tb/s or 8 Tb/s. Such types of integers can be implemented in hardware by

#199946 - $15.00 USD Received 22 Oct 2013; revised 14 Feb 2014; accepted 4 Mar 2014; published 14 Mar 2014

employing field programmable gate array (FPGA), but a discussion of the experimental realization of such approaches is beyond the scope of this paper.

Finally, we would like to make several remarks on our work. The two approaches discussed here do not require extensive post-processing. In the two, the post-processing only involves the calculation of HFD and the selection of m LSBs. Moreover, what we discuss requires neither the use of several lasers as in [38–40, 45] nor of delay lines as in [38, 49, 50]. Because of the post-processing chosen, we need not worry about the details of the acquisition nor of selecting a specific chaotic regime for the laser. Indeed, we checked (results not given here) that the randomness tests are passed for a wide range of operating parameters and for various acquisition conditions. That is, the two approaches are insensitive to these parameters, which is obviously another major advantage in addition to the ultrafast bit rate. Moreover, the two approaches introduced in the present study can also be carried out in a more compact photonic integrated circuit [27, 28].

4. Conclusion

In summary, we have reported two approaches to ultrafast RBG based on the chaotic dynamics of a semiconductor laser subject to optical feedback. The post-processing only includes the calculation of HFD and the selection of LSBs. In the first approach, we only have retained 4 LSBs for each sample according to the min-entropy and demonstrated a possibility of an high-quality physical random number generator with a generation rate of 160 Gb/s; in the second approach, starting with the use of a transformation of initial data obtained with 8-bit resolution into a 64-bit integer type, we have succeeded in extracting 55 LSBs and shown a feasibility of ultrafast physical-based pseudo RBG at a rate of 2.2 Tb/s. As mentioned above, post-processing requires additional computational resources. These might be based on parallel arrays of suitable ultrahigh-speed FPGAs or ASICs, though the actual implementation is beyond the scope of the present study. Our work not only highlights an approach to physical RBG that satisfies the bounds set by min-entropy, but also demonstrates the high potential for ultrafast physical-based pseudo RBG based on the chaotic lasers.

Acknowledgment

The authors would like to thank all reviewers for their helpful comments and suggestions on this manuscript. This work was partially supported by the National Natural Science Foundation of China (60976039, 61274042), the Basic Research Foundation of Sichuan Province (2011JY0030), and the funds for the Excellent Ph.D. Dissertation of the Southwest Jiaotong University (2011). BK, AL, and DSC acknowledge the support of the Conseil Régional de Lorraine and the CNRS through the PEPS OPTO-ALEA grant. DSC acknowledges the partial support of the US National Science Foundation through grant ECCS 0925713. NL would like to thank the China Scholarship Council for supporting him as a visiting Ph.D. student at the Georgia Institute of Technology.

#199946 - $15.00 USD Received 22 Oct 2013; revised 14 Feb 2014; accepted 4 Mar 2014; published 14 Mar 2014