# ARTICLE

# Quantum Monte Carlo Simulations with RANLUX Random Number Generator

Kenta HONGO<sup>1,2,\*</sup> and Ryo MAEZONO<sup>1</sup>

<sup>1</sup> School of Information Science, Japan Advanced Institute of Science and Technology, Asahidai 1-1, Nomi, Ishikawa, 923-1292, Japan

<sup>2</sup> Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, 02138, USA

We estimate the ground state energy of the hydrogen atom using the variational Monte Carlo (VMC) method with a simple model trial wavefunction and the RANLUX random number generator with five luxury levels (RANLUX-[0-4]). The resulting VMC energies are compared with the reference (exact) energy that is analytically evaluated with the same wavefunction as in the VMC simulations. The VMC energies with RANLUX-[1-4] are in good agreement with the reference energy within their statistical errors. On the other hand, RANLUX-0 with the lowest luxury level is found to give a systematic error of about 0.2 mhartree. In addition, we examine the time-step dependence of statistical quantities in VMC simulations. We find that the error bar of the VMC energy and the correlation time in VMC data have a minimum at a 50 % acceptance ratio in the Metropolis procedure.

KEYWORDS: computational chemistry, quantum chemistry, electronic structure, quantum Monte Carlo, random number generator

# I. Introduction

In our previous study of atomic and molecular systems, we have investigated several random number generators (RNGs) in quantum Monte Carlo (QMC) simulations.<sup>1)</sup> A multiple recursive generator with 8-th order recursion (MRG8)<sup>2)</sup> and the Mersenne twister generator (MT19937)<sup>3)</sup> were tested and compared with the RANLUX generator with five luxury levels (RANLUX-[0-4]).<sup>4)</sup> It was found that the OMC energies using the RANLUX generators consist with each other beyond the statistical error bar (standard error in the QMC energy estimate). In contrast, MRG and MT19937 give consistent results to within their error bars. This result may indicate that the statistical biases (systematic errors) in the QMC simulations strongly depend on the luxury levels in the RAN-LUX generators. In this study we made a detailed analysis of QMC results with RANLUX-[0-4] for the simplest example in quantum chemistry, i.e., the hydrogen atom. We estimated its ground state energy using the variational Monte Carlo (VMC) method with a simple model trial wavefunction. The resulting VMC energies were compared with the reference energy that is analytically evaluated with the same wavefunction as in the VMC simulations. In addition, we have first examined how the choice of time step in the VMC simulations affects statistical quantities such as the error bar of the VMC energy and correlation time.

In various fields Monte Carlo (MC) methods<sup>5–7)</sup> are used mainly for the following two purposes: (i) to simulate a system that has an intrinsic randomness and (ii) to solve a problem that is impossible or difficult to solve in either analytic or standard numerical ways. In quantum chemistry, MC is available in the context of the latter case to evaluate the electronic structure of atoms, molecules, and solids, which is known as (ab initio or first-principles) quantum Monte Carlo (QMC).<sup>8,9)</sup> QMC has attracted a lot of attention because of its highly accurate chemical description as well as moderate computational cost that scales as linear to cubic with respect to system size, compared to other conventional quantum chemistry methods.

Due to its statistical treatment, QMC evaluates any physical quantity accompanied with a statistical error bar (standard error). In quantum chemistry, interest is taken in the energy difference between the two systems, e.g., molecular binding energy that is the difference in the total energy between the molecule and its constituent atoms. In order to obtain numerically meaningful results using QMC, the total energy difference between the two systems should be at least larger than the corresponding error bar by an order of magnitude. Since a typical quantum chemistry calculation requires so-called chemical accuracy of about 1 kcal/mol (= 1.6 mhartree) for the energy difference, the statistical error in the QMC energy should be  $\sim 0.1$  mhartree.

A (pseudo) random number generator (RNG)<sup>5–7)</sup> plays an important role in any MC simulation, including QMC. It has been reported that 'good' RNGs in the context of statistical tests<sup>10,11)</sup> are not necessarily good for some MC simulations,<sup>12–20)</sup> i.e., the RNGs give a systematic error larger than the statistical error. In practice, the statistical biases due to RNGs depend on not only RNGs themselves, but also MC applications which use the RNGs. In our previous study, we have found that this is the case for QMC simulations of atomic and molecular systems.<sup>1)</sup>

The present paper is organized as follows: In Section II

<sup>\*</sup>Corresponding author, E-mail: kenta\_hongo@mac.com

<sup>© 2011</sup> Atomic Energy Society of Japan, All Rights Reserved.

we describe VMC algorithms focusing on where RNG is used and make a brief comment on RNGs, including the RANLUX generator. Numerical results and discussion are given in Section III. Section IV summarizes the present study.

# II. Variational Monte Carlo Method and Random Number Generator

This section describes the VMC algorithm related only to RNG, namely, the generation of random walks and the Metropolis accept/reject procedure. For a general and detailed description of QMC methods, it would be helpful to refer to textbooks<sup>8)</sup> and review papers.<sup>9)</sup> All the present calculations were performed using the CASINO code.<sup>22)</sup>

The total energy of a many-electron system is given as the expectation value of the Hamiltonian  $\hat{H}$  with respect to the many-electron wave function  $\Psi$ :

$$E[\Psi] = \frac{\int d\mathbf{R} E_L(\mathbf{R}) |\Psi(\mathbf{R})|^2}{\int d\mathbf{R} |\Psi(\mathbf{R})|^2},$$
(1)

where  $\mathbf{R} = (\mathbf{r}_1, \dots, \mathbf{r}_N)$  is a point in the configuration space of the *N*-electron system and  $E_L(\mathbf{R}) = \Psi^{-1}(\mathbf{R})\hat{H}(\mathbf{R})\Psi(\mathbf{R})$ is the local energy. (N = 1 for the hydrogen atom.) In VMC one adopts the trial wavefunction  $\Psi_T(\mathbf{R})$  and evaluates the expectation value with respect to  $\Psi_T(\mathbf{R})$  using the MC integration scheme, i.e., generating a sequence of configurations  $\{\mathbf{R}_m : m = 1, \dots, M\}$  in 3*N*-dimensional space distributed according to  $\Psi_T^2$  using the the Metropolis algorithm<sup>23</sup> and averaging the corresponding local energies:

$$E[\Psi] \approx \frac{1}{M} \sum_{m=1}^{M} E_L(\mathbf{R}_m) \equiv \langle E_L \rangle.$$
 (2)

The Metropolis transition probability density is a Gaussian distribution with variance  $\tau$ , where  $\tau$  is the VMC time step (dtvmc).

According to the central limit theorem, the variance of the MC estimate of  $E[\Psi]$  is  $\hat{\sigma}_{E_L}^2/M$ , where the variance of the function  $E_L$  (i.e.  $\hat{\sigma}_{E_L}^2$ ) is given by

$$\hat{\sigma}_{E_L}^2 = \frac{1}{M} \sum_{m=1}^M \left[ E_L(\mathbf{R}_m)^2 \right] - \langle E_L \rangle^2, \tag{3}$$

and hence an estimate of the size of the error bar (or standard error) on the MC estimate of  $E[\Psi]$  is  $\pm \hat{\sigma}_{E_L}/\sqrt{M}$ . Note that this relation holds if and only if  $\langle E_L \rangle$  is known exactly. Actually the denominator in the variance M is replaced with M-1, as shown in statistics texts, because  $\langle E_L \rangle$  is in reality measured as a MC sampling mean on computers.

In the actual determination of statistical errors of the VMC estimates, however, we should take into account (serial) correlations (or autocorrelations) between successive points  $\{\mathbf{R}_m\}$  generated in the Metropolis procedure. Using the autocorrelation function

$$C(k) = \sum_{m=1}^{M-k} \frac{(E_L(\mathbf{R}_m) - \langle E_L \rangle)(E_L(\mathbf{R}_{m+k}) - \langle E_L \rangle)}{(M-k) \,\hat{\sigma}_{E_L}^2}, \quad (4)$$

we may quantitatively evaluate such correlations in two ways: exponential and integrated correlation times. The exponential correlation time  $\tau_{\rm exp}$  is obtained from a relation  $C(k) \sim \exp(-k/\tau_{\rm exp})$  for  $k \gg 1$ , whereas the integrated one  $\tau_{\rm int}$  is defined as

$$\tau_{\rm int} = \sum_{k=1}^{\infty} C(k).$$
(5)

For large enough M, we obtain the true variance of  $E_L$ , i.e.,  $\sigma_{E_L}$ , from a relation

$$\sigma_{E_L}^2 = \hat{\sigma}_{E_L}^2 (1 + 2\tau_{\text{int}}),$$
 (6)

which means that the naive estimate of the size of the error bar always underestimates the true value. In other words, autocorrelations reduces the number of statistically uncorrelated samples from M to  $M/(1 + 2\tau_{\rm int})$ .

The method of data blocking is also available for estimating the error bar.<sup>26)</sup> In this method, we gather successive data points into blocks and average over the data in each block, and then calculate the variance of the set of block averages. Provided that the block length is much greater than the correlation time between data points, the block averages are statistically independent and an unbiased estimate of the mean is obtained. We used the blocking method to estimate the error bar of the VMC energy. In the present study we chose a block length of 1000, which is two orders of magnitude longer than the correlation times (see Section III.3) and hence long enough for the purpose of estimating the error bar correctly.

In the present study we adopt a trial wavefunction as

$$\Psi_T(\mathbf{r}) = \alpha e^{-\alpha r} + (1 - \alpha)e^{-(\alpha + 1)r}$$
(7)

with a parameter  $\alpha = 1.1$ .<sup>24)</sup> Although this form of the wavefunction contains the exact one ( $\alpha = 1$ ) for the hydrogen atom, the MC integration is meaningless since the local energy is constant. This model wavefunction detunes the exact one and hence the local energy  $E_L$  can be a function of the position in configuration space. In addition, this trial wavefunction automatically satisfies the electron-nucleus cusp condition, which prevents the local energy having a singularity at the electronnucleus coalescence and enables one to investigate systematic errors due to RNGs alone.

The RNG is used (i) to generate the trial move of configuration and (ii) to perform its accept/reject procedure in the Metropolis algorithm.<sup>23)</sup> In MC simulations the statistical biases due to RNGs originate from mainly the autocorrelations (or serial correlations) and biases due to non-uniformity of the sampling points.<sup>25)</sup> The former leads to an incorrect statistical error bar which is smaller than the exact one, while the latter introduces a systematic error in the mean value. For the former, a blocking procedure can be used to remove the correlation.<sup>26)</sup> The problem regarding the latter is considerably more difficult to resolve.<sup>25)</sup>

RANLUX has been a very popular RNG in the Monte Carlo community, especially for computationally demanding physical/chemical problems.<sup>4)</sup> Its algorithm was derived by applying chaos theory to an implementation of the subtract-with-borrow (SWB) generator, RCARYY.<sup>21,27)</sup> According to

chaos theory, some random numbers are removed to destroy the serial correlations in the original RCARRY sequence, i.e., given an integer p (p > 24), one first takes 24 successive RCARRY sequences and discards p - 24 numbers, then takes 24 sequences, and so on. In his paper Lüscher<sup>4</sup>) provided five p values, i.e., p = 24, 48, 97, 223, and 389, which correspond to the "luxury levels" of RANLUX ranging 0-4. Note that SWB has been proven to be equivalent to an efficient implementation of the linear congruential generator with a much longer period.<sup>28</sup> This is the reason why SWB fails in some randomness tests.<sup>29</sup>

We tested the convergence of the VMC energy with respect to the Monte Carlo steps ranging  $N_{\rm MC} = 10^9 - 10^{15}$  (see **Fig. 1**). The time step for the trial move (dtvmc) is determined such that the acceptance/rejection ratio  $\approx 0.5$ . In the present study, we obtained dtvmc = 0.5 from preliminary simulations. We also show similar results using RANLUX-0 and RANLUX-4 in order to understand how the choice of dtvmc affects the properties of a VMC calculation, i.e., the acceptance ratio, total energy, error bar, and correlation time.

In practical VMC simulations, a subsequence is generated by taking every k-th element from the original sequence, where k depends on the system, in order to weaken the serial autocorrelation. The subsequence can modify the original sequence. In this study, however, we did not consider such a treatment because we intend to focus on the intrinsic properties of the RANLUX generators. In order to consider intrinsic properties of a single RNG sequence, we did not perform a parallel VMC calculation, which may hide such properties.

#### **III. Results and Discussion**

#### 1. Sampling Bias

The VMC ground state energies of the hydrogen atom with the trial wavefunction in Eq. (7) were evaluated using different RANLUX generators. The reference energy of -497.3792mhartree was analytically calculated with the same wavefunction as in VMC. Figure 1 highlights the VMC energy  $E_{\rm VMC}$ convergence with respect to the Monte Carlo steps  $N_{\rm MC}$  for RANLUX-0 and RANLUX-4. Comparing the energies at  $N_{\rm MC} = 10^{13}$  and  $10^{15}$ , it is found that the VMC energies with RANLUX-0 and RANLUX-4 at  $N_{\rm MC} = 10^{13}$  converge to their own limits. Within one error bar ( $\sigma$ ), RANLUX-4 gives the same VMC energy at every  $N_{\rm MC}$ , while the VMC energies with RANLUX-0 differ at  $N_{\rm MC} = 10^{10}$  and  $10^{13}$ , though they coincide within  $2\sigma$ . Consequently, we may adopt the VMC energy at  $N_{\rm MC} = 10^{13}$  as the converged one.

**Table 1** reports the VMC energies and their error bars with RANLUX-[0-4] at  $N_{\rm MC} = 10^{13}$ . RANLUX-1 and RANLUX-2 give the same energies as the reference energy within their  $\sigma$ 's, respectively, while RANLUX-3 and RANLUX-4 within their  $2\sigma$ 's, respectively. On the other hand, RANLUX-0 give the VMC energy outside  $40\sigma$ . It is evident that RANLUX-0 gives a systematic error arising from a poor quality RNG. This result seems unsurprising since RANLUX-0 is theoretically equivalent to a notorious linear congruential generator, as mentioned in Section II.



**Fig. 1** VMC energy convergence with respect to the Monte Carlo steps. The energies are in units of mhartree and shifted by the reference energy of 497.3792 mhartree.

**Table 1** Energy differences between evaluated VMC and reference energies [ $\Delta E_{\rm VMC} = E_{\rm VMC} - (-497.3792)$  (mhartree)] and their error bars ( $\sigma$ ) with RANLUX-[0-4] at  $N_{\rm MC} = 10^{13}$ . All energies are in units of mhartree.

	$\Delta E_{\rm VMC}$	$\sigma$
RANLUX-0	0.2375	0.0056
RANLUX-1	-0.0033	0.0066
RANLUX-2	-0.0045	0.0056
RANLUX-3	0.0105	0.0065
RANLUX-4	0.0066	0.0062

## 2. Computational Costs

**Figure 2** shows the computational (CPU) times in the RANLUX routine to generate each RANLUX sequence as a function of p (see Section II). It is obvious from this figure that the CPU time is approximately proportional to p, which is consistent with a previous study.<sup>1)</sup> The fraction of time used by the RANLUX routine compared to the total VMC routine is about 1%. In the present case the difference in the total CPU time between luxury levels was negligible, but it would be more significant (an order of hours) for larger systems in actual QMC simulations, which take a couple of days.

#### 3. Correlation Time in QMC

**Table 2** gives a list of the integrated correlation times defined in Eq. (5) and their error bars. No significant difference of the correlation time between the RANLUX generators was found, similar to results reported in our previous study.<sup>1)</sup> It is likely that the autocorrelation does not explicitly relate to the systematic error.

### 4. Time-Step Dependence in VMC

**Figure 3** highlights the VMC time step (dtvmc) dependence of various quantities in a VMC simulation: (a) acceptance ratio in the Metropolis procedure, (b) VMC energy, (c)



Fig. 2 CPU time [sec] for each RANLUX, plotted as the dependence on *p*-parameter controlling the luxury level, e.g., p = 24 indicates RANLUX-0 and so on. See Section II.

Table 2Correlation times and their error bars with<br/>RANLUX-[0-4]. All times are given in units of Monte<br/>Carlo steps.

	Correlation time	Error bar
RANLUX-0	17.80	0.82
RANLUX-1	18.43	0.87
RANLUX-2	17.92	0.84
RANLUX-3	17.90	0.84
RANLUX-4	17.95	0.84

error bar in the evaluation of the VMC energy, and (d) correlation time in the evaluation of the VMC energy. The VMC calculations were performed with RANLUX-0 and RANLUX-4 at  $N_{\rm MC} = 10^{13}$ , i.e., the number of Monte Carlo steps is common to every VMC simulation with different values of dtvmc. As shown in Fig. 3(a), the acceptance ratio abruptly decreases with a larger dtvmc. dtvmc = 0.5 gives an acceptance ratio of  $\approx 50\%$ . Figure 3(b) shows that the VMC energies with RANLUX-0 strongly depend on the values of dtvmc, whereas those with RANLUX-4 have a relatively weak dependence. It is found from Figs. 3(c) and (d) that the error bars and correlation times are a minimum at dtvmc  $\approx 0.5$ , which corresponds to an acceptance ratio of 50%. This result implies that a 50 % acceptance ratio leads to the most efficient VMC simulation. Although in the Monte Carlo community the above results have been heuristically determined as approximately optimal time step choices, the theoretical reason for this has not been investigated and discussed, especially in the firstprinciples QMC community, so far as we are aware.

# **IV. Summary**

We estimated the ground state energy of the hydrogen atom using the variational Monte Carlo (VMC) method with a simple model trial wavefunction and the RANLUX random number generator with five luxury levels (RANLUX-[0-4]). The VMC energies with RANLUX-1 and RANLUX-2 were consistent with the exact energy within their  $\sigma$ 's, whereas those



Fig. 3 VMC time step (dtvmc) dependence of (a) acceptance ratio, (b) VMC energy, (c) error bar (error bar), and (d) correlation time for RANLUX-0 and RANLUX-4 at  $N_{\rm MC} = 10^{13}$ .

with RANLUX-3 and RANLUX-4 within their  $2\sigma$ 's. On the other hand, it was found that RANLUX-0 gives an evident systematic error of about 0.2 mhartree in the evaluation of

the VMC energy. We also examined the VMC time step dependence in VMC simulations. It was found that a 50 % acceptance ratio gives an optimum VMC time step in terms of VMC simulation efficiency. Future work on this subject will be to conduct the same benchmarks on parallel/vector supercomputer systems, and analyze the performance of various RNGs such as MRG8,<sup>2)</sup> MT19937,<sup>3)</sup> as well as physical random number generators.

#### Acknowledgment

The computation in this work has been performed using the computer facilities at JAIST and Harvard University. Financial support was provided by Precursory Research for Embryonic Science and Technology, Japan Science and Technology Agency (PRESTO-JST), and by a Grant-in-Aid for Scientific Research in Priority Areas Development of New Quantum Simulators and Quantum Design (No.17064016) of The Japanese Ministry of Education, Culture, Sports, Science, and Technology (KAKENHI-MEXT) for R.M. We also appreciate financial support from the ISM Cooperative Research Program (2010-ISM-CRP-2060). K.H. is grateful for a JSPS Postdoctoral Fellowship for Research Abroad. The authors thank Dr. Mark A. Watson and Mr. Sangwoo Shim for their useful comments.

#### References

- K. Hongo, R. Maezono, K. Miura, "Random Number Generators Tested on Quantum Mote Carlo Simulations," *J. Comput. Chem.*, **31**[11], 2186-2194 (2010).
- K. Miura, Full Polynomial Multiple Recursive Generator (MRG) Revisited; MCQMC 2006, Ulm, Germany (2006), [Unpublished].
- M. Matsumoto, T. Nishimura, "Mersenne Twister: A 623dimensionally equidistributed uniform pseudorandom number generator," ACM Trans. Model Comput. Simulat., 8[1], 3-30 (1998).
- M. Lüscher, "A Portable High-Quality Random Number Generators for Lattice Field-Theory Simulations," *Comput. Phys. Commun.*, **79**[1], 100-110 (1994).
- 5) J. E. Gentle, *Random Number Generation and Monte Carlo Methods*, 2nd Ed. Springer, New York (2003).
- D. P. Landau, K. Binder, A Guide to Monte Carlo Simulations in Statistical Physics, Cambridge University Press: Cambridge (2000).
- D. E. Knuth, *The Art of Computer Programming, Volume 2, Seminumerical Algorithms*, 3rd ed., Addison-Wesley Publishing Company, Reading, Massachusetts (1998).
- B. L. Hammond, W. A. Lester, Jr., P. J. Reynolds, *Monte Carlo Methods in Ab Initio Quantum Chemistry*, World Scientific, Singapore (1994).
- W. M. C. Foulkes, L. Mitas, R. J. Needs, G. Rajagopal, "Quantum Monte Carlo simulations of solids," *Rev. Mod. Phys.*, 73[1], 33-83 (2001).
- 10) G. Marsaglia, The Marsaglia Random Number CDROM, including the DIEHARD Battery of Tests of Randomness, Department of Statistics, Florida State University, Tallahassee, Florida (1995). Available at: http://stat.fsu.edu/~geo/diehard.html.

- NIST, A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications, NIST Special Publication 800-22, National Institute for Standards and Technology, Gaithersburg, Maryland (2000).
- 12) C. Kalle, S. Wansleben, "Problems with the random number generator RANF implemented on the CDC cyber 205," *Comput. Phys. Commun.*, **33**[4], 343-346 (1984).
- G. Parisi, F. Rapuano, "Effects of the random number generator on computer simulations," *Phys. Lett.*, B157[4], 301- 302 (1985).
- 14) A. Hoogland, A. Compagner, H. W. J. Blöte, "Smooth finitesize behaviour of the three-dimensional Ising model," *Physica*, A132[2-3], 593-596 (1985).
- 15) T. Filk, M. Marcu, K. Fredenhagen, "Long range correlations in random number generators and their influence on Monte Carlo simulations," *Phys. Lett.*, **B165**[1-3], 125-130 (1985).
- A. Milchev, K. Binder, D. W. Heermann, "Fluctuations and lack of self-averaging in the kinetics of domain growth," *Z. Phys.*, B63[4], 521-535 (1986).
- A. M. Ferrenberg, D. P. Landau, Y. J. Wong, "Monte Carlo simulations: Hidden errors from "good" random number generators," *Phys. Rev. Lett.*, 69[23], 3382-3384 (1992).
- P. Grassberger, "On Correlations in Good Random Number Generators," *Phys. Lett.*, A183[1], 43-46 (1993).
- P. D. Coddington, "Analysis of Random Number Generators Using Monte-Carlo Simulation," *Int. J. Mod. Phys.*, C5[3], 547-560 (1994).
- P. D. Coddington, "Tests of random number generators using Ising model simulations," *Int. J. Mod. Phys.*, C7[3], 295-303 (1996).
- 21) F. James, "RANLUX A FORTRAN Implementation of the High-Quality Pseudrandom Number Generator of Luscher," *Comput. Phys. Commun.*, **79**[1], 111-114 (1994).
- 22) R. J. Needs, M. D. Towler, N. D. Drummond, P. López Ríos, CASINO version 3.0 User Manual, University of Cambridge: Cambridge (2009).
- 23) N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, "Equation of State Calculations by Fast Computing Machines," *J. Chem. Phys.*, **21**[6], 1087-1092 (1953).
- 24) N. Nemec, "Diffusion Monte Carlo: Exponential Scaling of Computational Cost for Large Systems," *Phys. Rev.*, B81[3], 035119 (2010).
- 25) M. Mascagni, "Serial and Parallel Random Number Generation," in *Quantum Monte Carlo Methods in Physics and Chemistry*, M. P. Nightingale, C. J. Umrigar, (Eds), Kluwer Academic, Dordrecht, The Netherlands, 425-446 (1999).
- H. Flyvbjerg, H. G. Petersen, "Error-Estimates on Averages of Correlated Data," J. Chem. Phys., 91[1], 461-466 (1989).
- G. Marsaglia, A. Zaman, "A New Class of Random Number Generators," Ann. Appl. Prob., 1[3], 462-480 (1991).
- 28) S. Tezuka, P. L'Ecuyer, R. Couture, "On the Add-with-Carry and Subtract-with-Borrow Random Number Generators," ACM Trans. Model. Comput. Simulat., 3[4], 315-331 (1993).
- 29) I. Vattulainen, K. KanKaala, J. Saarinen, T. Ala-Nissila, "A Comparative Study of Some Pseudorandom Number Generators," *Comput. Phys. Commun.*, 86[3], 209-226 (1995).