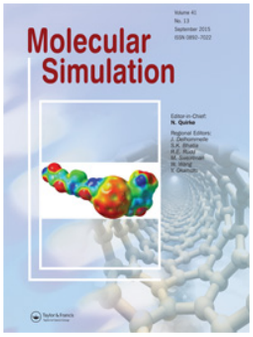

Molecular Simulation

ISSN: 0892-7022 (Print) 1029-0435 (Online) Journal homepage: https://www.tandfonline.com/loi/gmos20

Quantifying uncertainty in molecular dynamics simulations of grain boundary migration

C.P. Race

To cite this article: C.P. Race (2015) Quantifying uncertainty in molecular dynamics simulations of grain boundary migration, Molecular Simulation, 41:13, 1069-1073, DOI: 10.1080/08927022.2014.935774

To link to this article: https://doi.org/10.1080/08927022.2014.935774

Molecular Simulation, 2015
Vol. 41, No. 13, 1069–1073, http://dx.doi.org/10.1080/08927022.2014.935774

Taylor & Francis
Taylor & Francis Group

Quantifying uncertainty in molecular dynamics simulations of grain boundary migration

C.P. Race[1]*

Max-Planck Institute for Iron Research, Düsseldorf 40237, Germany

(*Received 31 March 2014; final version received 9 June 2014*)

Molecular dynamics simulations of simple bicrystal systems have been much used as a tool to explore how the migration of grain boundaries varies with their structure and with experimental conditions. In order to permit the exploration of a large parameter space, many studies are forced to rely on a small number of simulations (often a single simulation) for each configuration. The motion of a grain boundary is inherently statistical and any variability in the measured grain boundary velocity should be taken into account in subsequent analysis of trends in grain boundary mobility. Here we present the results of large numbers of simulations of equivalent boundaries, which show that this variability can be large, particularly when small systems are simulated. We show how a bootstrap resampling approach can be used to characterise the statistical uncertainty in boundary velocity using the information present in a single simulation. We show that the approach is robust across a variety of system sizes, temperatures and driving force strengths and types, and provides a good order-of-magnitude measure of the population standard deviation across multiple equivalent simulations.

Keywords: molecular dynamics; grain boundary mobility; bootstrap resampling

1. Introduction

The movement of grain boundaries has a major impact on the evolution of the mechanical properties of polycrystalline materials during processing and in application. Classical molecular dynamics (MD) simulations have been much used to investigate how the mobility of grain boundaries varies with their geometry, with the applied driving force for motion and with temperature.[1–13] Classical MD offers great flexibility in the conditions of the computer experiment, but it is computationally costly. Previous explorations of the variation of mobility across configuration space have thus often been restricted to a single simulation for each combination of geometry, temperature and driving force and to small systems of several 10s of 1000s of atoms. Because the migration of a grain boundary is a thermally activated process, the evolution of the boundary's position with time contains many fluctuations. Figure 1 shows some typical trajectories. At low temperatures the boundary remains essentially stationary for large periods of time before moving a fixed distance relatively quickly. Superimposed on this trajectory are statistical fluctuations around the positions of stable equilibrium.

If we are to be able to draw firm quantitative conclusions from results of MD simulations of grain boundary migration, we need a robust method to estimate the uncertainty in the grain boundary velocity. In Section 2 of this paper we present the results of large numbers of equivalent simulations, each typical of those used to explore grain boundary mobility across configuration space. These show that the scatter in average velocity can be very large. We consider how the result of a single simulation might be used to estimate this scatter. We demonstrate the inadequacy of the standard linear regression fitting statistics for this purpose and adopt instead an intuitively appealing *bootstrap resampling* approach.[14,15] We show that this approach consistently gives a good order-of-magnitude estimate of the scatter across a range of simulation conditions.

Bootstrap resampling is a well-established technique and is often employed in situations where the inherent randomness of some process must be considered, but where the data do not show the behaviour assumed in the definition of classical statistics. Bootstrap analysis is extensively used to analyse large volumes of data in the fields of molecular biology and bioinformatics [16] and in financial and economic analysis.[17] It has also been applied in seismology, for example to determine seismic velocities,[18] and in materials science (e.g. to study the strength of steel [19]).

2. Bootstrap resampling of grain boundary trajectories

We have carried out a large number of MD simulations, using the Lammps code,[20] of the migration of grain boundaries for a selection of geometries, driving forces and system sizes. The simulations that we have carried out are detailed in Table 1. Grain boundary motion was driven either using the artificial driving force of Janssens et al. [6] or an elastic driving force of the form used by Zhang et al. [5]. The interatomic forces are modelled using an

*Email: christopher.race@manchester.ac.uk

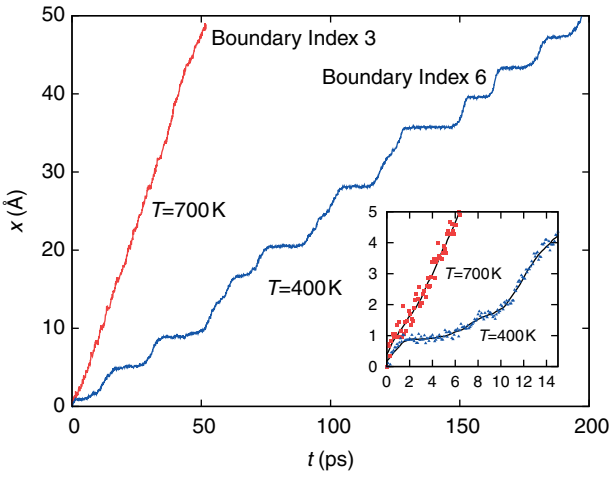

Figure 1. (Colour online) Sample data for the mean position of a [1 1 1] Σ7 symmetric tilt grain boundary as a function of time at two different temperatures. Thermal noise in the grain boundary position is evident in the inset.

embedded atom model potential fitted to properties of aluminium [21] (or nickel [22] in the case of the elastic driving force). Table 1 also shows the spread in the mean velocity across equivalent simulations for each simulation type. In all cases this spread is significant (up to 26% of the mean velocity), particularly in the case of smaller systems and at lower temperatures. This scatter in measured velocities must be taken into account when we calculate quantities of interest such as grain boundary mobilities and migration energy barriers. If it is not, then the strength even of qualitative conclusions about the behaviour of those quantities may be in doubt.

Due to computational restrictions, it will often be necessary to determine an estimate in the error in the velocity from a single trajectory of the grain boundary position, $x(t)$. Where previous studies include such error estimates, it is often unclear how they have been determined. Frequently the estimated errors appear to be too small when compared with sample trajectory data provided. It is possible that these quoted errors are the parameter error statistics from a least squares linear fitting (linear regression) of the sampled trajectory data, $x(t_0), x(t_1), \ldots, x(t_N)$. However, these statistics are highly dependent on the sampling frequency and should not be used. Figure 2 demonstrates this with an example trajectory. The grain boundary migration is made up of 13 discrete jumps during the course of the simulation, separated by waiting periods of varying duration. By measuring these waiting periods and calculating their mean and standard deviation, we obtain our best estimate of the velocity and the corresponding standard deviation: 0.26 ± 0.03 Å ps^{-1}. However, our data might sample this series of 13 events at a high frequency and the parameter error statistics from a linear regression would take no account of the high degree of correlation in the data. Figure 2 shows how the error estimates vary with sampling frequency. Even at a relatively low sampling rate, the estimated percentage error in the mean velocity is much less than the 11% spread suggested by the distribution in the waiting time.

The character of a grain boundary trajectory can vary considerably from boundary to boundary and depending on the conditions of the simulation. Sometimes discrete migration events can be resolved, sometimes more continuous motion is observed. Thermal noise will also be present to some degree. Given the complexity of the process under study, involving many atoms, we would like a method of estimating errors which does not rely on a particular statistical model of the grain boundary migration. Rather we will trade off statistical rigour for broad and simple applicability. The so-called *bootstrap resampling* methods [14] offer such a compromise and have the additional advantage that they are intuitively understandable, computationally inexpensive and trivial to apply.

Figure 3 illustrates the principle, which is to directly exploit the statistical information inherent in a single trajectory. We take our original sample of grain boundary positions $\{x(t_k)\}_{k=1}^{N}$ and randomly sample $n = N/l$ integers, $\{j_1, \ldots, j_n\}$, uniformly and with replacement

Table 1. Details of the number of simulations of different types in this study, including statistics for the mean and standard deviation of the average boundary velocities measured and calculated by bootstrap resampling (shown in parentheses).

Index	Boundary and force	Number of atoms	Dimensions (Å)	Temperature (K)	Number of simulations	Velocity (Å ps^{-1})	SD/velocity
1	A	10,752	121 × 13 × 14	700	80	0.84 (0.83)	0.13 (0.097)
2	A	21,504	121 × 26 × 14	700	40	0.88 (0.87)	0.093 (0.069)
3	A	43,008	121 × 26 × 28	700	40	0.95 (0.94)	0.064 (0.046)
4	A	86,016	121 × 52 × 28	700	40	0.95 (0.95)	0.061 (0.034)
5	A	258,048	121 × 79 × 56	700	40	0.95 (0.95)	0.036 (0.022)
6	A	43,008	121 × 26 × 28	400	40	0.21 (0.21)	0.26 (0.15)
7	B	14,400	108 × 54 × 29	1400	10	0.059 (0.059)	0.13 (0.13)

Note: Boundary A is a [1 1 1] Σ7 symmetric tilt simulated using an aluminium potential [21] and an artificial driving force [6]. Boundary B is a [1 0 0] Σ5 asymmetric tilt simulated using a nickel potential [22] and an elastic driving force [5].

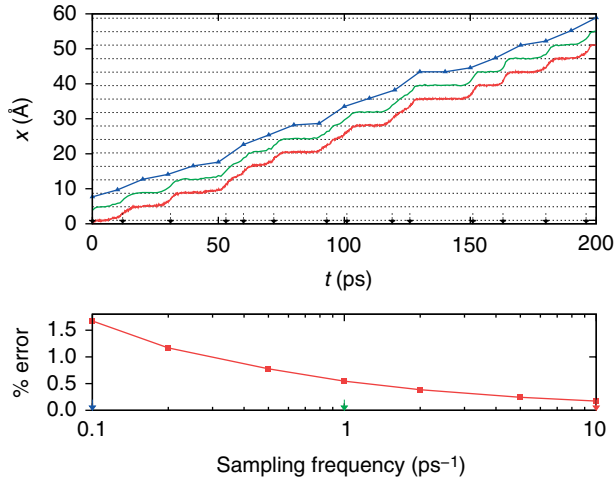

Figure 2. (Colour online) Top panel: An example grain boundary trajectory (for a boundary of Index 6 in Table 1) sampled at several different frequencies: bottom (red) 10 ps^{-1}; middle (green) 1 ps^{-1}; top (blue) 0.1 ps^{-1}. The arrows mark the times of discrete migration events between the stable positions indicated by horizontal gridlines (the plots have been shifted vertically to make them clearer). Bottom panel: The parameter error estimate as a fraction of the average velocity, obtained by least-squares linear regression of data sampling the trajectory in the top panel at different frequencies (the coloured arrows indicate the corresponding sampled trajectories in top panel).

from the range $[0, N-1]$. We can then construct a new trajectory from portions of length l of the original trajectory $x(t_{j_i}) \rightarrow x(t_{j_i+l})$, where we assume that the original

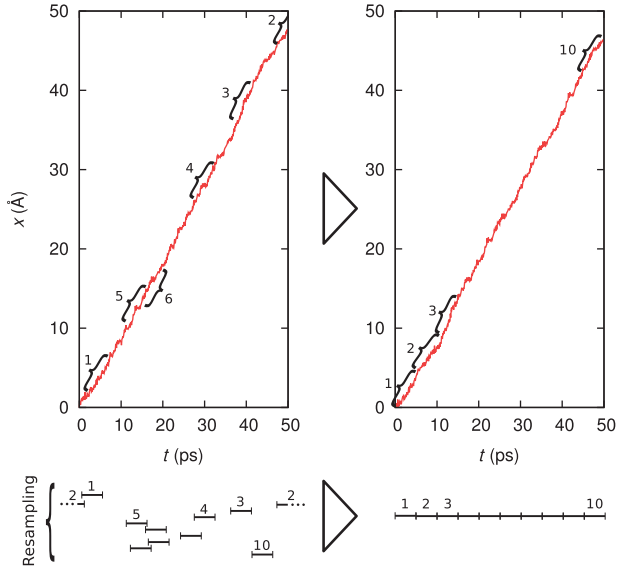

Figure 3. (Colour online) A schematic of the bootstrap resampling process for an example grain boundary trajectory. On the left is the original trajectory, below which are indicated the portions of the data randomly sampled with replacement and using periodic boundary conditions. On the right is a new trajectory constructed by joining these portions. A large set of such resampled trajectories can be constructed at very low computational cost.

trajectory wraps upon itself to avoid reduced sampling of the end portions. Figure 3 illustrates an example of this resampling and reconstruction process. It gives a trajectory with an average velocity,

$$v = \frac{1}{t_N - t_0} \sum_{i=1}^{n} (x(t_{j_i+l}) - x(t_{j_i})). \tag{1}$$

We can repeat this resampling process many times (10s of 1000s of samples are eminently achievable) to produce a large sample of velocities, which we assume to be representative of the sample we would obtain from running a large number of equivalent simulations. We can then calculate simple statistics, such as the standard deviation, from this sample.

The bootstrap resampling method is intuitive, but it does contain a free parameter: the resampling block length, l (or equivalently, the number of portions we cut from the original trajectory, $n = N/l$). Figure 4 shows that the estimate of the spread in velocities varies strongly with l in an undesirable way. In fact, much of this variation is due to thermal fluctuation of the grain boundary around its equilibrium position. We can model this noise as a binomial distribution of equally probable forward and backward deviations of amplitude a from the equilibrium position. For an original trajectory of N samples at a frequency f, use of a block length of l sampling periods will contribute N/l samples from the noise distribution to

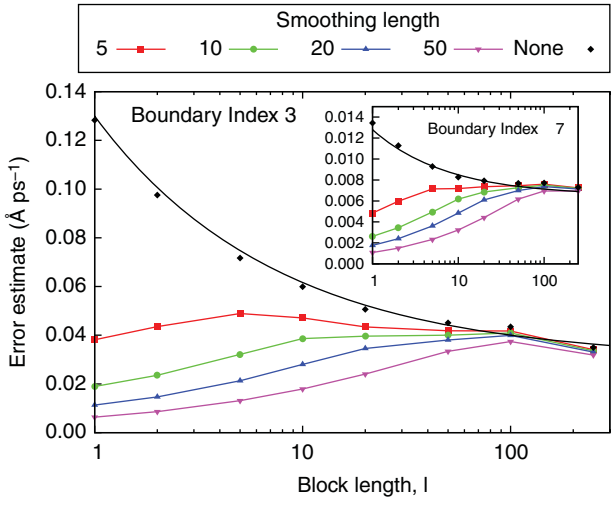

Figure 4. (Colour online) The effect of smoothing and sampling block length l on the estimated error in velocity from bootstrap resampling. The main plot shows results for a trajectory of a boundary with Index 3, the inset one with Index 7. Data for the original trajectory are shown by (black) diamonds. Most of the variation with l is due to thermal noise. A simple model for this noise contribution (discussed in the main text) is shown by the black curve. Other symbols (and lines to guide the eye) show the effect of smoothing the original data with a moving average over different numbers s of original sample points (as indicated in the key) before undertaking the resampling.

the variation in the velocity in the resampled set of trajectories. The magnitude of this contribution will be $0.5a\sqrt{N/l}$ over the duration of the trajectory N/f. Hence the thermal noise contributes $0.5af/\sqrt{Nl}$ to the variation in the velocity. This simple model of thermal noise fits the data well, as shown in Figure 4 for values $N = 600$, $f = 10$ ps^{-1} and $a = 0.5$ Å derived from a typical trajectory. A similarly good fit to the thermal noise in an alternative boundary (Index 7) is evident in the inset of Figure 4.

To remove the undesirable impact of this noise, we can smooth the original trajectory by calculating a moving average position over a window of width s. This introduces a second parameter into our bootstrapping method. Figure 4 shows how the estimated spread varies for this smoothed data and we see that the estimate becomes insensitive to the sampling block length, l, over a wide range. By selecting a smoothing length s of five thermal periods and a block length $l \geq s$, but less than a characteristic waiting time between migration events, we arrive at a well-defined method and robust algorithm. The same behaviour is evident for the alternative boundary (Index 7) in the inset of Figure 4.

Both the examples shown in Figure 4 are from relatively smooth trajectories like the 700 K (red) line shown in Figure 1. In these cases the smoothing removes the high-frequency variation in the position of the boundary, which in turn reduces the dependence of the bootstrap result on the block length. For trajectories like the 400 K (blue) line shown in Figure 1, which show 'stop–start' behaviour with discrete migration events, the smoothing has a less significant effect. In these cases, the block length, l, must be chosen so that it is not significantly smaller than a typical waiting period between migration events. If l is too small, then the resampled trajectory will still draw a large number of samples from the thermal noise (which provides the only source of variation in the boundary position between migration events) with a consequent increase in the bootstrap estimate of the uncertainty.

Figure 5 shows the result of applying the method to the simulations presented in Table 1. The bootstrap resampling consistently underestimates the true spread, as we would expect (this is a known property of such methods [14]), but it is order-of-magnitude correct (in contrast to the parameter estimates from linear regression). The performance of the method is stable over a range of temperatures, system sizes, driving forces and grain boundary geometries (over which there is an order of magnitude variation in the size of the true spread). The error bars in Figure 5 show the variability in the bootstrap estimate with the block length l (maximum and minimum values for $5 \leq l \leq 50$). This variability is significant, but no larger than the variation in the performance between simulation types. The inset of Figure 5 shows the trends in

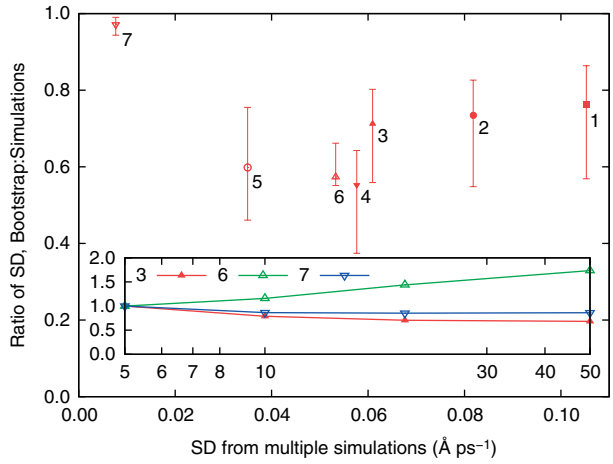

Figure 5. (Colour online) The ratio of the variation in the mean trajectory as estimated by bootstrap resampling to that calculated from multiple simulations under the conditions listed in Table 1. The error bars indicate the range of values obtained when the block length is varied between 5 and 50 periods of thermal noise, with the smoothing length fixed at five periods. The inset shows the bootstrap estimated error as a function of smoothing s and block length l when $l = s$ for simulations with Indices 3, 6 and 7. The behaviour is discussed in the main text.

the variability of the bootstrap result with block length. For smooth trajectories (at higher temperatures) the effect of thermal noise dominates and so resampling with fewer, larger blocks reduces the residual contribution of noise in the results. For stop–start trajectories the trend is reversed, because this behaviour is characteristic of lower temperatures in which thermal noise is insignificant.

We note that the versatility of the bootstrap approach means that it can also be used to estimate formal confidence intervals,[23] in this case for the grain boundary velocity. To do so we would simply need to determine the range of velocities in which some given percentage (our confidence level) of the resampled velocities was contained. However, we believe that this would perhaps imply a precision that is unjustified given the variation in the performance of the algorithm evident in Figure 5.

3. Conclusions

We have presented data from classical MD simulations of grain boundary migration, which show that there can be significant variation in the measured average velocity across a set of equivalent simulations. This spread is greater than is often indicated by the error bars in the published literature, perhaps because of the method used to calculate these errors.

We show how the use of a bootstrap resampling technique can give improved estimates of the scatter in velocity in a way that is robust to variation in the grain

boundary geometry, the driving force, the temperature and the system size. We have successfully applied the algorithm outlined here to a detailed analysis of the system-size- and driving-force-dependence of the migration kinetics of a [1 1 1] Σ7 symmetric tilt boundary. [24] The approach is intuitive, and cheap and easy to apply. It could equally well be used to process data from simulations of the thermally activated migration of other defects such as dislocations.

Acknowledgements

C.P.R. was funded by a Postdoctoral Fellowship of the Alexander von Humboldt Foundation.

Note

1. Current address: Dalton Nuclear Institute, University of Manchester, Oxford Road, Manchester M13 9PL, UK.

References

[1] Jhan R-J, Bristowe PD. A molecular dynamics study of grain boundary migration without the participation of secondary grain boundary dislocations. Scr Metall Mater. 1990;24(7):1313–1318.
[2] Deng C, Schuh C. Atomistic simulation of slow grain boundary motion. Phys Rev Lett. 2011;106(4):45503-1–45503-4.
[3] Schonfelder B, Gottstein G, Shvindlerman L. Atomistic simulations of grain boundary migration in copper. Metall Mater Trans A. 2006;37(15):1757–1771.
[4] Zhang H, Srolovitz DJ, Douglas JF, Warren JA. Characterization of atomic motion governing grain boundary migration. Phys Rev B. 2006;74(11):115404-1–115404-10.
[5] Zhang H, Mendelev MI, Srolovitz DJ. Computer simulation of the elastically driven migration of a flat grain boundary. Acta Mater. 2004;52(9):2569–2576.
[6] Janssens KGF, Olmsted D, Holm EA, Foiles SM, Plimpton SJ, Derlet PM. Computing the mobility of grain boundaries. Nat Mater. 2006;5(2):124–127.
[7] Zhang H, Upmanyu M, Srolovitz D. Curvature driven grain boundary migration in aluminum: molecular dynamics simulations. Acta Mater. 2005;53(1):79–86.
[8] Olmsted DL, Foiles SM, Holm EA. Grain boundary interface roughening transition and its effect on grain boundary mobility for non-faceting boundaries. Scr Mater. 2007;57(12):1161–1164.
[9] Trautt ZT, Upmanyu M, Karma A. Interface mobility from interface random walk. Science. 2006;314(5799):632–635.
[10] Zhang H, Mendelev MI, Srolovitz DJ. Mobility of Σ5 tilt grain boundaries: inclination dependence. Scr Mater. 2005;52 (12):1193–1198.
[11] Schönfelder B, Wolf D, Phillpot SR, Furtkamp M. Molecular-dynamics method for the simulation of grain-boundary migration. Interface Sci. 1997;5(4):245–262.
[12] Olmsted DL, Holm EA, Foiles SM. Survey of computed grain boundary properties in face-centered cubic metals–II: grain boundary mobility. Acta Mater. 2009;57(13):3704–3713.
[13] Zhou J, Mohles V. Towards realistic molecular dynamics simulations of grain boundary mobility. Acta Mater. 2011;59 (15):5997–6006.
[14] Efron B. Bootstrap methods: another look at the jackknife. Ann Stat. 1979;7(1):1–26.
[15] Efron B, Tibshirani R. Statistical data analysis in the computer age. Science. 1991;253(5018):390–395.
[16] Douady CJ, Delsuc F, Boucher Y, Doolittle WF, Douzery EJP. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol Biol Evol. 2003;20 (2):248–254.
[17] Cogneau P, Zakamouline V. Block bootstrap methods and the choice of stocks for the long run. Quant Finance. 2013;13 (9):1443–1457.
[18] Sacchi MD. A procedure for highresolution velocity analysis. Geophysics. 1998;63(5):1716–1725.
[19] Obraztsov SM, Konobeev YV, Birzhevoy GA, Rachkov VI. Bootstrap calculation of ultimate strength temperature maxima for neutron irradiated ferritic/martensitic steels. J Nucl Mater. 2006;359 (3):263–267.
[20] Plimpton S. Fast parallel algorithms for short-range molecular dynamics. J Comput Phys. 1995;117(1):1–19.
[21] Zope RR, Mishin Y. Interatomic potentials for atomistic simulations of the Ti–Al system. Phys Rev B. 2003;68(2):024102.
[22] Bonny G, Pasianot RC, Castin N, Malerba L. Ternary Fe–Cu–Ni many-body potential to model reactor pressure vessel steels: first validation by simulated thermal annealing. Philos Mag. 2009;89 (34–36):3531–3546.
[23] Henderson AR. The bootstrap: a technique for data-driven statistics. using computer-intensive analyses to explore experimental data. Clin Chim Acta. 2005;359(1–2):1–26.
[24] Race CP, von Pezold J, Neugebauer J. Role of the mesoscale in migration kinetics of flat grain boundaries. Phys Rev B. 2014;89 (21):214110-1–214110-11.