Updated reply to the report of the First Referee

We thank the referee for reading our paper and spotting omissions and mistakes in our text. We hope that the following clarifications and corrections will enable full understanding of our approach. We address the comments paragraph by paragraph quoting the referee before our response.

> I have found a few gaps in the presentation which prevent a full grasp of their assumptions and numerical validation. Their arguments and derivations cannot be fully reproduced by qualified experts. For instance, the authors claim they sampled at finite temperature, but the value of the temperature is not provided anywhere in the text.

The text was intended as a demonstration of the proposed method not as a reproduction or prediction of a particular experimental result. This may perhaps explain unfortunate omission of sampled temperature in the text. While the temperature can in fact be extracted from the average energy in Fig. 3 and 4, it should also be stated in the text explicitly. The sampling temperature for all data in first version of the text is 300 K, chosen as standard ambient temperature. In the resubmitted version of the text we have corrected this omission. Furthermore, following suggestions of both referees, we have extended the presented data to higher temperatures - up to 2000K. The temperatures are indicated in the text.

> If the temperature is much smaller than the melting temperature (~3,000 K for SiC), the harmonic approximation is expected to be quite accurate and the the Gaussian "prior" sampling described in Sec. 3 should be almost exact. This seems to agree with the large acceptance ratio (up to 80%) observed by the authors.

The shape of the energy distribution is not determined by harmonicity of the potentials but by the Central Limit Theorem and the size of the system. The difference between prior distribution (which comes from our approximation of the *displacement* distribution) and the target distribution does not stem from the anharmonicity of the potential but mainly from the fact that the displacements of the atoms in the crystal are *not independent* (as stated in our text). Please note that we have no direct access to the potential energy of the system - we can only specify the geometry and calculate the resulting energy instead of directly generating the target energy distribution from Fig. 3. The high acceptance ratio comes from our selection of the displacement distribution and tuning algorithm described in the paper. The Metropolis-Hastings (M-H) algorithm can generate a target distribution from any prior which is non-zero over the domain. However, the acceptance ratio may be very low if you fail to use the prior which is a good approximation of target distribution. It is a well-known fact in the numerical statistics community that the good selection of the prior distribution is the key to an effective use of probability distribution sampling

algorithms. To further demonstrate the effectiveness of the algorithm for wide range of temperatures we have included results up to T=2000K. We have also extended explanation of the procedure used to generate Fig. 3 (which clearly has acceptance ratio below 80% claimed in the text for a typical values of delta). This figure was generated with artificially large value of delta (0.1 instead of 0.005-0.02) to make difference between prior and posterior distributions easily visible.

> A very high acceptance ratio is not necessarily an advantage of the approach as it may imply large correlation in the sample. The authors should discuss the position autocorrelation function obtained from their approach and compare against the one obtained using canonical MD sampling. The best acceptance rate is the one which minimises the autocorrelation time.

The issue of sample correlation would be indeed important if we used a 'random walk'-type algorithm for the prior generation (which is a popular variant of the M-H algorithm). Instead, we use *independent* samples and the only possible correlation between them arises from a very small change (less than 2%) in the variance of the position distribution. We discuss this issue in the paragraph starting in line 192 (212 in revised text). On the other hand, the MD derived data has obvious autocorrelations - note that we can derive phonon frequencies from the Fourier transform of the velocity autocorrelation function along the MD trajectory. Thus, it is necessary to separate sampling points on the trajectory by substantial intervals allowing for these correlations to die out. Nevertheless, all time steps between the sampling points still need to be calculated, which leads to large inefficiency of the MD as a configuration generator. To further clarify the issue we have expanded the explanation in the text in paragraph starting at line 212 (revised text).

> The agreement between the Monte Carlo and MD phonon dispersion shown in Fig. 5 is to be expected if the anharmonicity is negligible. The agreement on the phonon lifetimes is a tougher check, but it is hard to draw any quantitative conclusion from Fig. 6. The points are pretty scattered over a semi-log plot, which means that the error can be rather large. The authors should discuss a more quantitative estimator, e.g., the square root of the sum over the wave-vectors and bands of the square deviation of the Monte Carlo and MD phonon lifetimes.

Indeed, for the purely harmonic system the phonon frequency test is not useful since phonon frequencies are independent from the displacement size. However, if the system considered in the paper were close to harmonic, we would expect to obtain very long phonon lifetimes (since they are infinite in the harmonic system). The data in Fig. 6 demonstrates that many of the phonon modes exhibit lifetimes below 10ps - showing non-negligible anharmonicity in the model. This fact provides justification for the validity of the phonon frequency test. Furthermore, expanded temperature data of new Fig. 5 (up to 2000K), demonstrates

that anharmonicity induced by high temperatures has some small influence on the convergence of phonon data but not on the converged results (lower row of Fig. 5) which shows good agreement between MD and HECSS data in full range of temperatures. Additionally, we have included in the text RMS errors for frequencies obtained with both methods. The phonon lifetimes are very sensitive to the accuracy of the model. This is especially true in case of large values which indicate small deviations from harmonicity and usually carry large error bars. Unfortunately, the large range (close to four orders of magnitude) of the values of lifetimes makes the simple RMS measure of differences very misleading - since the differences at high end of the range will dominate the sum. However, we agree that the previous Figure 6 was indeed not very clear. Thus, we have replaced it with the separate plot for three temperatures (T=100, 300, and 600K) splitting the small-sample data set to a separate row. We think that the new Fig.6 clearly demonstrates good agreement between data obtained with MD and HECSS procedure over 4 orders of magnitude in phonon lifetime.

> Lines 34-35: The authors mention "running a 30000 steps MD". Why exactly this number of steps?

The number of steps (30 000) used in the introduction was a typical relaxation time of a long-run MD suggested by the often used "rule of thumb" in MD calculations (50 times period of typical vibrations in the system). For 3C-SiC: $f \approx 10\text{THz} = 10^{13}\text{Hz}$ \$ t=10^{-13}\$s = 100fs; 50 * 100fs = 5ps. With 1fs time step that equals 5000 steps *minimum* run where we can use at most half of it for actual data (you need to provide time to obtain thermal equilibrium). If we need approx. 30 data points (as required by anharmonic calculations, see Fig. 6) and they should be separated by at least 1ps interval (at least 10 typical vibrations) we get approximately 30 000 steps. The cited number itself has no 'magical' value and results from the setup of the calculations presented in the paper. To avoid impression that the number 30 000 has any special meaning, we have replaced the number by the phrase: "thousands of MD steps".

> Eq. (1): Not all the symbols have been introduced in the text.

Eq. (1): we have added to the revised text a sentence introducing the missing symbols: x_n - generalized coordinate or momentum, H - Hamiltonian, T - temperature, k_B - Boltzmann constant, δ_{mn} - Kronecker delta.

> Eq. (2): The authors implicitly assume a two-body force field. What would happen in the case of a many-body force field like the embedded atom model or Tersoff potentials?

Eq. (2): We make no two-body assumption, neither implied nor explicit. The formulation of equipartition theorem, Eq. (1), explicitly concerns *single* coordinates (the only non-zero term due to the Kronecker delta) and makes no assumption on the form of the Hamiltonian H. The Taylor expansion, Eq. (2), is not a complete expansion in all coordinates q (note the scalar q symbol). It is the Taylor expansion in a *single* coordinate with coefficients (C_n - proportional to *partial* derivatives of energy with respect to this coordinate) which are func-

tions of all the other coordinates in the system. What is more, the calculations presented in the paper use the mentioned Tersoff potential developed in refs 17, 18. We understand that due to the formulation of the surrounding text this may not be entirely clear and may confuse the reader. To avoid this, we have added a clarifying sentence below Eq. (2).

> Lines 80-81: The sentence "Only experience can tell us how good this approximation is and how wide its applicability range is" is not correct and underrate the role of a large body of numerical analysis.

The unfortunate sentence 80-81 brings nothing of importance to the text. Thus we have removed it in the resubmitted version and reformulated the surrounding paragraph.

> Lines 102-103: The sentence "The reasonable class is very broad here, certainly containing all physically interesting cases by virtue of requiring a finite variance and a well-defined mean" is not correct, unless phase transitions are excluded. The variance of several quantities diverges close to a phase transition.

Variance of several quantities is indeed divergent in some phase transitions. In cases where the transition involves divergent heat capacity this includes energy variance. Thus, our phrase :"...all physically interesting cases..." was indeed wrong. The sentence has been corrected and we clearly state that in cases where energy variance diverges, the procedure cannot be used. We thank the referee for spotting this important fact.

> Eq. (5): The equal sign is not correct as for N$\rightarrow\infty$the variance of the distribution of the right-hand side is zero.

The Eq. (5) was an attempt to formally write asymptotic relation of the Central Limit Theorem described in the paragraph 99-102. The CLT is indeed not a limit relation but asymptotic distribution convergence relation and the Eq. (5) should use appropriate notation for such relations as convergence in distribution:

$$\sqrt{3N}\left(\frac{1}{N}\sum_i E_i - \langle E\rangle\right) \xrightarrow{d} \mathcal{N}(0,\sigma).$$

The mistake in notation has no consequences for the arguments and conclusions presented in the text. The Eq. (5) has been corrected in the revised text.

> Line 125: A "single server" is not a well-defined object: how many CPU's were used?

"Single server" mentioned in line 125 was used as a rough indication of the computational effort involved in the described task. It is nothing out of ordinary: 2x4 cores CPU and 32GB RAM. This is actually a fairly under-powered and old machine, less powerful than some of newer generation laptops. The information has been added to the sentence in revised text.

> In comparing the computational efficiency of MD and Monte Carlo methods, one has to take into consideration the time spent generating random numbers, which may be more computationally demanding than propagating a trajectory, given the same number of force-field evaluations.

In some cases the random number generation may be indeed fairly expensive but in the case of typical systems of tens of atoms, the energy and forces evaluation is much more time-consuming. For instance, the random number generator we have used (from SciPy.stats library) takes 180μs to generate 3000 random numbers required to create one sample for the 5x5x5 supercell of 3C-SiC. The single evaluation of energy for the same cell (1000 atoms) takes 4ms (20 times longer) using ASAP3 with OpenKIM model from our calculations. A more sophisticated interaction model is bound to be even more time-consuming. Furthermore, molecular dynamics requires calculating multiple time steps per every generated sample. Considering this facts, we maintain that the proposed HECSS approach offers substantial advantage over MD as a source of configuration data.