

Manuscript Number:

Title: Accurate Sampling with Noisy Forces from Approximate Computing

Article Type: Regular Article

Keywords: approximate computing; CP2K; fluctuation-dissipation theorem;
FPGA; i-PI; low precision arithmetic

Corresponding Author: Mr. Michael Lass,

Corresponding Author's Institution: Paderborn University

First Author: Varadarajan Rengaraj

Order of Authors: Varadarajan Rengaraj; Michael Lass; Christian Plessl;
Thomas D Kühne

Abstract: In scientific computing, the acceleration of atomistic computer simulations by means of custom hardware is finding ever growing application. A major limitation, however, is that the high efficiency in terms of performance and low power consumption entails the massive usage of low-precision computing units. Here, based on the approximate computing paradigm, we present an algorithmic method to rigorously compensate for numerical inaccuracies due to low-accuracy arithmetic operations, yet still obtaining exact expectation values using a properly modified Langevin-type equation.

Accurate Sampling with Noisy Forces from Approximate Computing

Varadarajan Rengaraj^{a,b}, Michael Lass^{b,c,*}, Christian Plessl^{b,c}, Thomas D.
Kühne^{a,c,d}

^a*Dynamics of Condensed Matter, Department of Chemistry,
Paderborn University, Warburger Str. 100, 33098 Paderborn, Germany*

^b*Department of Computer Science,
Paderborn University, Warburger Str. 100, 33098 Paderborn, Germany*

^c*Paderborn Center for Parallel Computing,
Paderborn University, Warburger Str. 100, 33098 Paderborn, Germany*

^d*Center for Sustainable Systems Design,
Paderborn University, Warburger Str. 100, 33098 Paderborn, Germany*

Abstract

In scientific computing, the acceleration of atomistic computer simulations by means of custom hardware is finding ever growing application. A major limitation, however, is that the high efficiency in terms of performance and low power consumption entails the massive usage of low-precision computing units. Here, based on the approximate computing paradigm, we present an algorithmic method to rigorously compensate for numerical inaccuracies due to low-accuracy arithmetic operations, yet still obtaining exact expectation values using a properly modified Langevin-type equation.

Keywords: approximate computing, CP2K, fluctuation-dissipation theorem, FPGA, i-PI, low precision arithmetic

1. Introduction

Molecular dynamics (MD) is a very powerful and widely used technique to study thermodynamic equilibrium properties, as well as the real-time dynamics of complex systems made up of interacting atoms [1]. This is done by numerically solving Newton's equations of motion in a time-discretized fashion via computing the nuclear forces of all atoms at every time step [2]. Computing these forces by analytically differentiating the interatomic potential with respect to the nuclear coordinates is computationally rather expensive, which is particularly true for electronic structure based *ab-initio* MD simulations [3–6].

*Corresponding author

Email addresses: rengaraj@campus.uni-paderborn.de (Varadarajan Rengaraj),
michael.lass@uni-paderborn.de (Michael Lass), christian.plessl@uni-paderborn.de
(Christian Plessl), tdkuehne@mail.upb.de (Thomas D. Kühne)

For a long time newly developed microchips became faster and more efficient over time due to new manufacturing processes and shrinking transistor sizes. However, this development slowly comes to an end as scaling down the structures of silicon based chips becomes more and more difficult. The focus therefore shifts towards making efficient use of the available technology. Hence, beside algorithmic developments [7–14], there have been numerous custom computing efforts in this area to increase the efficiency of MD simulations by means of hardware acceleration, which we take up in this work. Examples of the latter are MD implementations on graphics processing units (GPUs) [15–21], field-programmable gate arrays (FPGAs) [22, 23], and application-specific integrated circuits (ASICs) [24, 25]. While the use of GPUs for scientific applications is relatively widespread [26–28], the use of ASICs [29–32] and FPGAs is less common [33–38], but gained attention over the last years. In general, to maximize the computational power for a given silicon area, or equivalently minimize the power-consumption per arithmetic operation, more and more computing units are replaced with lower-precision units. This trend is mostly driven by market considerations of the gaming and artificial intelligence industries, which are the target customers of hardware accelerators and naturally do not absolutely rely on full computing accuracy.

In the approach presented in this paper, we make effective use of low-precision special-purpose hardware for general-purpose scientific computing by leveraging the approximate computing (AC) paradigm [39, 40]. The general research goal of AC is to devise and explore ingenious techniques to relax the exactness of a calculation to facilitate the design of more powerful and/or more efficient computer systems. However, in scientific computing, where the exactness of all computed results is of paramount importance, attenuating accuracy requirements is not an option. Yet, we will demonstrate that it is nevertheless possible to rigorously compensate for numerical inaccuracies due to low-accuracy arithmetic operations and still obtain exact expectation values as obtained by ensemble averages of a properly modified Langevin equation.

The remainder of the paper is organized as follows. In Section 2 we revisit the basic principles of AC before introducing our modified Langevin equation in Section 3. Thereafter, in Section 4, we describe the computational details of our computational experiments. Our results are presented and discussed in Section 5 before concluding the paper in Section 6.

2. Approximate Computing

A basic method of approximation and a key requirement for efficient use of processing hardware is the use of adequate data widths in computationally intensive kernels. While in many scientific applications the use of double-precision floating-point is most common, this precision is not always required. For example, iterative methods can exhibit resilience against low precision arithmetic as has been shown for the computation of inverse matrix roots [41] and for solving systems of linear equations [39, 42–44]. Mainly driven by the growing popularity of artificial neural networks [45], we can observe growing support of

Table 1: Bitwidth of common floating-point formats

Type	sign	exponent	mantissa
IEEE 754 Quadruple-precision	1	15	112
IEEE 754 Double-precision	1	11	52
IEEE 754 Single-precision	1	8	23
IEEE 754 Half-precision	1	5	10
Bfloat16 (truncated IEEE single-precision)	1	8	7

low-precision data types in hardware accelerators. In fact, recent GPUs targeting the data center have started supporting half-precision as well, nearly doubling the peak performance compared to single-precision and quadrupling it compared to double-precision arithmetics [46]. However, due to the low number of exponent bits, half-precision only provides a very limited dynamic range. In contrast, **bfloat16** provides the same dynamic range as single-precision, and just reduces precision. It is currently supported by Google’s Tensor Processing Units (TPU) [47] and support is announced for future Intel Xeon processors [48] and Intel AgileX FPGAs. A list of commonly used data types, together with the corresponding number of bits used to store the exponent and the mantissa, are shown in Table 1.

Yet, programmable hardware such as FPGAs, as a platform for custom-built accelerator designs [49–51], can make effective use of all of these, but also entirely custom number formats. Developers can specify the number of exponent and mantissa bits and trade off precision against the amount of memory blocks required to store values and the number of logic elements required to perform arithmetic operations on them.

In addition to floating-point formats, also fixed-point representations can be used. Here, all numbers are stored as integers of fixed size with a predefined scaling factor. Calculations are thereby performed using integer arithmetic. On CPUs and GPUs only certain models can perform integer operations with a peak performance similar to that of floating-point arithmetic, depending on the capabilities of the vector units / stream processors. Nevertheless, FPGAs typically can perform integer operations with performance similar to or even higher than that of floating-point. Due to the high flexibility of FPGAs with respect to different data formats and the possible use of entirely custom data types, we see them as the main target technology for our work. For this reason, we consider both floating-point and fixed-point arithmetic in the following.

3. Methodology

To demonstrate the concept of approximate computing, we introduce white noise to the interatomic forces that are computed while running the MD simulation. In this section, we describe in detail on how we introduce the noise to mimic in software the behaviour that would be observed when running the MD on the actual FPGA or GPU hardware with reduced numerical precision.

We classify the computational errors into two types: fixed-point errors, and floating-point errors. Assuming that \mathbf{F}_I are the exact and \mathbf{F}_I^N the noisy forces from a MD simulations with low precision on an FPGA for instance, fixed-point errors can be modelled by

$$\mathbf{F}_I^N = \begin{pmatrix} F_I^x \\ F_I^y \\ F_I^z \end{pmatrix} + \begin{pmatrix} c_1 \times 10^{-\beta} \\ c_2 \times 10^{-\beta} \\ c_3 \times 10^{-\beta} \end{pmatrix}, \quad (1)$$

whereas floating-point errors are described by

$$\mathbf{F}_I^N = \begin{pmatrix} F_I^x \times 10^{-\alpha_1} \\ F_I^y \times 10^{-\alpha_2} \\ F_I^z \times 10^{-\alpha_3} \end{pmatrix} + \begin{pmatrix} c_1 \times 10^{-(\alpha_1+\beta)} \\ c_2 \times 10^{-(\alpha_2+\beta)} \\ c_3 \times 10^{-(\alpha_3+\beta)} \end{pmatrix}. \quad (2)$$

Therein, c_1 , c_2 and c_3 are random values chosen in the range $[-0.5, 0.5]$, whereas F_I^x , F_I^y and F_I^z are the individual force components of \mathbf{F}_I , respectively. The floating-point scaling factor is denoted as α and the magnitude of the applied noise by β .

To rigorously correct the errors introduced by numerical noise we employ a modified Langevin equation. In particular, we model the force as obtained by a low-precision computation on a GPU or FPGA-based accelerator as

$$\mathbf{F}_I^N = \mathbf{F}_I + \boldsymbol{\Xi}_I^N, \quad (3)$$

where $\boldsymbol{\Xi}_I^N$ is an additive white noise for which

$$\langle \mathbf{F}_I(0) \boldsymbol{\Xi}_I^N(t) \rangle \cong 0 \quad (4)$$

holds. Throughout, $\langle \dots \rangle$ denotes Boltzmann-weighted ensemble averages as obtained by the partition function $Z = \text{Tr} \exp(-E/k_B T)$, where E is the potential energy, k_B the so-called Boltzmann constant, and T the temperature. Given that $\boldsymbol{\Xi}_I^N$ is unbiased, which in our case is true by its very definition, it is nevertheless possible to accurately sample the Boltzmann distribution by means of a Langevin-type equation [52–54], which in its general form reads as

$$M_I \ddot{\mathbf{R}}_I = \mathbf{F}_I + \boldsymbol{\Xi}_I^N - \gamma_N M_I \dot{\mathbf{R}}_I, \quad (5)$$

where $\dot{\mathbf{R}}_I$ are the nuclear coordinates (the dot denotes time derivative), M_I are the nuclear masses and γ_N is a damping coefficient, which is chosen to compensate for $\boldsymbol{\Xi}_I^N$. The latter, in order to guarantee an accurate canonical sampling, has to obey the fluctuation-dissipation theorem

$$\langle \boldsymbol{\Xi}_I^N(0) \boldsymbol{\Xi}_I^N(t) \rangle \cong 2\gamma_N M_I k_B T \delta(t). \quad (6)$$

Substituting Eq. 3 into Eq. 5 results in the desired modified Langevin equation

$$M_I \ddot{\mathbf{R}}_I = \mathbf{F}_I^N - \gamma_N M_I \dot{\mathbf{R}}_I, \quad (7)$$

Table 2: Values for γ_N^{fix} and γ_N^{float} as a function of β .

β	γ_N^{fix}	γ_N^{float}
0		0.00025
1	0.0004	0.000005
2	0.000009	0.000005
3	0.0000009	

which will be used throughout the remaining of this paper. This is to say that the noise, as originating from a low-precision computation, can be thought of as the additive white noise of a thitherto unknown damping coefficient γ_N , which satisfies the fluctuation-dissipation theorem of Eq. 6. The specific value of γ_N is determined in such a way so as to generate the correct average temperature, as measured by the equipartition theorem

$$\left\langle \frac{1}{2} M_I \dot{\mathbf{R}}_I \right\rangle = \frac{3}{2} k_B T. \quad (8)$$

4. Computational details

To demonstrate our approach we have implemented it in the CP2K suite of programs [55]. More precisely, we have conducted MD simulations of liquid Silicon (Si) at 3000 K using the environment-dependent interatomic potential of Bazant et al. [56, 57]. All simulations consisted of 1000 Si atoms in a 3D-periodic cubic box of length 27.155 Å. Using the algorithm of Ricci and Ciccotti [58], Eq. 5 was integrated with a discretized timestep of 1.0 fs with $\gamma_N = 0.001 \text{ fs}^{-1}$.

Whereas the latter settings were used to compute our reference data, in total six different cases of fixed-point and floating-point errors were investigated by varying the exponent β between 0 (huge noise) and 3 (tiny noise) that is, ranging from 1/1000 of the physical force up to the same magnitude as the force. As already alluded to above, the additive white noise is compensated by means of Eq. 7 by varying γ_N on-the-fly using a Berendsen-like algorithm until the equipartition theorem of Eq. 8 is satisfied [59–61]. Alternatively, γ_N can be computed by integrating the autocorrelation function of the additive white noise [62]. In Table 2 the resulting values of γ_N^{fix} for fixed-point and γ_N^{float} for floating-point errors are reported as a function of β .

5. Results and Discussion

As can be directly deduced from Table 2, the smaller values of γ_N for a given β immediately suggest the higher noise resilience when using floating-point as compared to fixed-point numbers. In Figs. 1 and 2, the Si-Si partial pair-correlation function $g(r)$, which describes how the particle-density varies as a function of distance from a reference particle (atoms, molecules, colloids, etc.), as computed using an optimal scheme for orthorombic systems [63], is shown

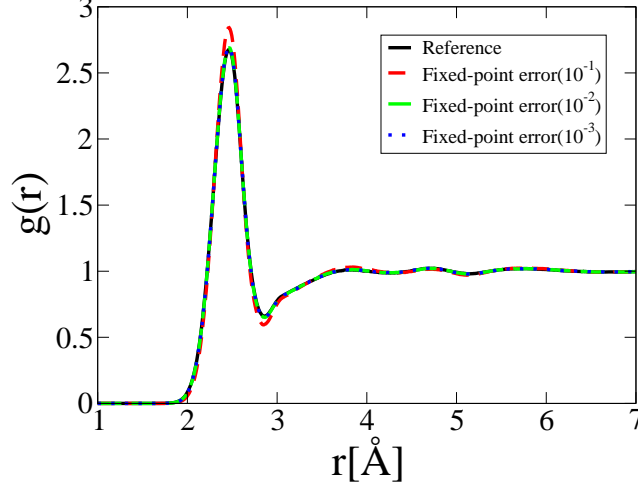


Figure 1: Partial pair correlation function for liquid Si at 3000 K with noisy forces introduced by fixed-point errors of magnitude 10^{-3} (blue), 10^{-2} (green) and 10^{-1} (red). For comparison, the results, as obtained by our reference calculations without noise, are shown in black.

for different values of β . As can be seen, for both fixed-point and floating-point errors, the agreement with our reference calculation is nearly perfect up to the highest noise we have investigated. As already anticipated earlier, the usage of floating-point errors is not only able to tolerate higher noise levels, but is also throughout more accurate.

To verify that the sampling is indeed canonical, in Fig. 3 the actual kinetic energy distribution as obtained by our simulations using noisy forces is depicted and compared to the analytic Maxwell distribution. It is evident that if sampled long enough, not only the mean value, but also the distribution tails are in excellent agreement with the exact Maxwellian kinetic energy distribution, which demonstrates that we are indeed performing a canonical sampling. To further assess the accuracy of the present method, we expand the autocorrelation of the noisy forces, i.e.

$$\langle \mathbf{F}_I^N(0) \mathbf{F}_I^N(t) \rangle \quad (9a)$$

$$= \langle (\mathbf{F}_I(0) + \mathbf{\Xi}_I^N(0)) (\mathbf{F}_I(t) + \mathbf{\Xi}_I^N(t)) \rangle \quad (9b)$$

$$= \langle \mathbf{F}_I(0) \mathbf{F}_I(t) \rangle + \langle \mathbf{F}_I(0) \mathbf{\Xi}_I^N(t) \rangle \quad (9c)$$

$$+ \langle \mathbf{F}_I(t) \mathbf{\Xi}_I^N(0) \rangle + \langle \mathbf{\Xi}_I^N(0) \mathbf{\Xi}_I^N(t) \rangle.$$

Since the cross correlation terms between the exact force and the additive white noise is vanishing due to Eq. 4, comparing the autocorrelation of the noisy forces

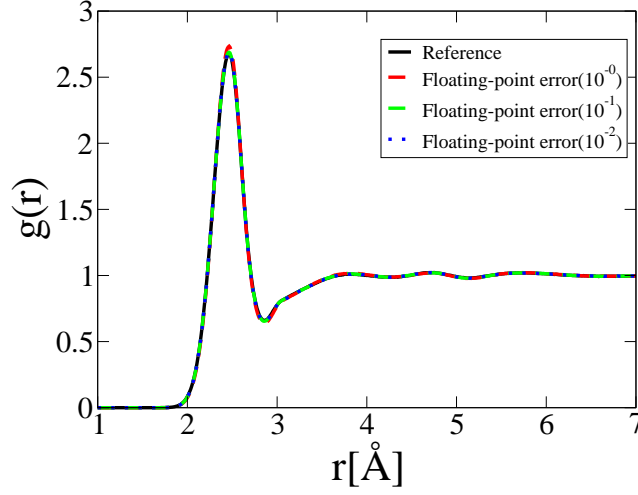


Figure 2: Partial pair correlation function for liquid Si at 3000 K with noisy forces introduced by floating-point errors of magnitude 10^{-2} (blue), 10^{-1} (green) and 10^{-0} (red). For comparison, the results, as obtained by our reference calculations without noise, are shown in black.

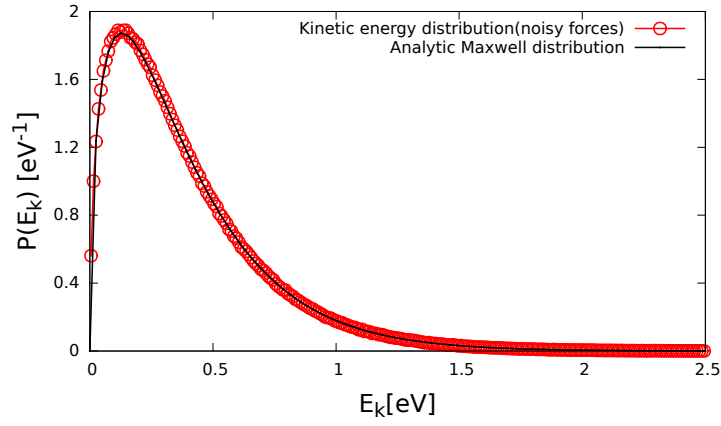


Figure 3: Kinetic energy distribution of liquid Si at 3000 K, as obtained by our simulations using noisy forces (circles). For comparison the analytic Maxwell distribution is also shown (line).

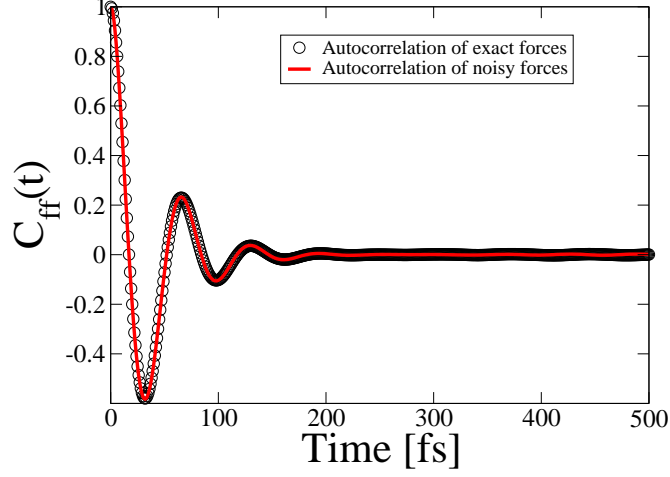


Figure 4: The Autocorrelation of the noisy forces $\langle \mathbf{F}_I^N(0) \mathbf{F}_I^N(t) \rangle$ (line), which are compared to the autocorrelation of the exact forces $\langle \mathbf{F}_I(0) \mathbf{F}_I(t) \rangle$ (circles).

$\langle \mathbf{F}_I^N(0) \mathbf{F}_I^N(t) \rangle$ with the autocorrelation of the exact forces $\langle \mathbf{F}_I(0) \mathbf{F}_I(t) \rangle$ permits to assess the localization of the last term of Eq. 9c. The fact that $\langle \mathbf{F}_I^N(0) \mathbf{F}_I^N(t) \rangle$ is essentially identical to $\langle \mathbf{F}_I(0) \mathbf{F}_I(t) \rangle$, as can be seen in Fig. 4, implies that $\langle \Xi_I^N(0) \Xi_I^N(t) \rangle$ is very close to a δ -function as required by the fluctuation-dissipation theorem in order to ensure an accurate canonical sampling of the Boltzmann distribution. In other words, from this it follows that our initial assumption underlying Eq. 7, to model the noise due to a low-precision calculation as an additive white noise channel, is justified.

6. Conclusion

We conclude by noting that the present method has been recently implemented in the universal force engine i-PI [64], which can be generally applied to all sorts of forces affected by stochastic noise such as those computed by GPUs or other hardware accelerators [15–21], and potentially even quantum computing devices [65–68]. The possibility to apply similar ideas to N-body simulations [69, 70] and to combine it with further algorithmic approximations [71] is to be underlined and will be presented elsewhere.

Acknowledgements

The authors would like to thank the Paderborn Center for Parallel Computing (PC2) for computing time on OCuLUS and FPGA-based supercomputer

NOCTUA. Funding from the Paderborn University’s research award for “Green IT” is kindly acknowledged. This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant Agreement No.: 716142).

References

- [1] B. J. Alder, T. E. Wainwright, Phase transition for a hard sphere system, *J. Chem. Phys.* 27 (5) (1957) 1208–1209. doi:10.1063/1.1743957.
- [2] A. Rahman, Correlations in the motion of atoms in liquid argon, *Phys. Rev.* 136 (2A) (1964) A405–A411. doi:10.1103/PhysRev.136.A405.
- [3] R. Car, M. Parrinello, Unified approach for molecular dynamics and density-functional theory, *Phys. Rev. Lett.* 55 (22) (1985) 2471–2474. doi:10.1103/PhysRevLett.55.2471.
- [4] T. D. Kühne, M. Krack, F. R. Mohamed, M. Parrinello, Efficient and accurate car-parrinello-like approach to born-oppenheimer molecular dynamics, *Phys. Rev. Lett.* 98 (6) (2007) 066401. doi:10.1103/PhysRevLett.98.066401.
- [5] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, J. D. Joannopoulos, Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients, *Rev. Mod. Phys.* 64 (4) (1992) 1045–1097. doi:10.1103/RevModPhys.64.1045.
- [6] T. D. Kühne, Second generation car-parrinello molecular dynamics, *WIREs Comput. Mol. Sci.* 4 (4) (2014) 391–406. doi:10.1002/wcms.1176.
- [7] M. E. Tuckerman, B. J. Berne, G. J. Martyna, Reversible multiple time scale molecular dynamics, *J. Chem. Phys.* 97 (3) (1992) 1990–2001. doi:10.1063/1.463137.
- [8] M. Snir, A note on n-body computations with cutoffs, *Theor. Comput. Syst.* 37 (2) (2004) 295–318. doi:10.1007/s00224-003-1071-0.
- [9] Y. Shan, J. L. Klepeis, M. P. Eastwood, R. O. Dror, D. E. Shaw, Gaussian split ewald: A fast ewald mesh method for molecular simulation, *J. Chem. Phys.* 122 (5) (2005) 054101. doi:10.1063/1.1839571.
- [10] D. E. Shaw, A fast, scalable method for the parallel evaluation of distance-limited pairwise particle interactions, *J. Comput. Chem.* 26 (13) (2005) 1318–1328. doi:10.1002/jcc.20267.
- [11] P. Gonnet, Pairwise verlet lists: Combining cell lists and verlet lists to improve memory locality and parallelism, *J. Comp. Chem.* 33 (1) (2012) 76–81. doi:10.1002/jcc.21945.
- [12] P. Gonnet, A quadratically convergent shake in $\mathcal{O}(n^2)$, *J. Comp. Phys.* 220 (2) (2007) 740–750. doi:10.1016/j.jcp.2006.05.032.

- [13] C. John, T. Spura, S. Habershon, T. D. Kühne, Quantum ring-polymer contraction method: Including nuclear quantum effects at no additional computational cost in comparison to ab initio molecular dynamics, *Phys. Rev. E* 93 (4) (2016) 043305. doi:10.1103/PhysRevE.93.043305.
- [14] T. D. Kühne, E. Prodan, Disordered crystals from first principles i: Quantifying the configuration space, *Annals of Physics* 391 (1) (2018) 120–149. doi:10.1016/j.aop.2018.01.016.
- [15] J. A. Anderson, C. D. Lorenz, A. Travesset, General purpose molecular dynamics simulations fully implemented on graphics processing units, *J. Comp. Phys.* 227 (10) (2008) 5342–5359. doi:10.1016/j.jcp.2008.01.047.
- [16] J. E. Stone, D. J. Hardy, I. S. Ufimtsev, K. Schulten, Gpu-accelerated molecular modeling coming of age, *J. Mol. Graphics Modell.* 29 (2) (2010) 116–125. doi:10.1016/j.jmgm.2010.06.010.
- [17] P. Eastman, V. S. Pande, Openmm: A hardware-independent framework for molecular simulations, *Comput. Sci. Eng.* 12 (2010) 34–39. doi:10.1109/MCSE.2010.27.
- [18] P. H. Colberg, F. Höfling, Highly accelerated simulations of glassy dynamics using gpus: Caveats on limited floating-point precision, *Comp. Phys. Commun.* 182 (5) (2011) 1120–1129. doi:10.1016/j.cpc.2011.01.009.
- [19] W. M. Brown, A. Kohlmeyer, S. J. Plimpton, A. N. Tharrington, Implementing molecular dynamics on hybrid high performance computers – particle–particle–particle–mesh, *Comp. Phys. Commun.* 183 (3) (2012) 449–459. doi:10.1016/j.cpc.2011.10.012.
- [20] S. Le Grand, A. W. Götz, R. C. Walker, Spfp: Speed without compromise—a mixed precision model for gpu accelerated molecular dynamics simulations, *Comp. Phys. Commun.* 184 (2) (2013) 374–380. doi:10.1016/j.cpc.2012.09.022.
- [21] A. J. Abraham, T. Murtola, R. Schulz, S. Pall, J. C. Smith, B. Hess, E. Lindahl, Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* 1–2 (2015) 19–25. doi:10.1016/j.softx.2015.06.001.
- [22] M. C. Herbordt, Y. Gu, T. VanCourt, J. Model, B. Sukhwani, M. Chiu, Computing models for fpga-based accelerators, *Comput. Sci. Eng.* 10 (6) (2008) 35–45. doi:10.1109/MCSE.2008.143.
- [23] M. C. Herbordt, Y. Gu, T. VanCourt, J. Model, B. Sukhwani, M. Chiu, Explicit design of fpga-based coprocessors for short-range force computations in molecular dynamics simulations, *Parallel Computing* 34 (4–5) (2008) 261–277. doi:10.1016/j.parco.2008.01.007.

- [24] D. E. Shaw, M. M. Deneroff, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson, K. J. Bowers, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, D. J. Ierardi, I. Kolossváry, J. L. Klepeis, T. Layman, C. McLeavey, M. A. Moraes, R. Mueller, E. C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, S. C. Wang, Anton, a special-purpose machine for molecular dynamics simulation, in: Proceedings of the 34th Annual International Symposium on Computer Architecture, ISCA '07, ACM, New York, NY, USA, 2007, pp. 1–12. doi:10.1145/1250662.1250664.
- [25] D. E. Shaw, J. P. Grossman, J. A. Bank, B. Batson, J. A. Butts, J. C. Chao, M. M. Deneroff, R. O. Dror, A. Even, C. H. Fenton, A. Forte, J. Gagliardo, G. Gill, B. Greskamp, C. R. Ho, D. J. Ierardi, L. Iserovich, J. S. Kuskin, R. H. Larson, T. Layman, L.-S. Lee, A. K. Lerer, C. Li, D. Killebrew, K. M. Mackenzie, S. Y.-H. Mok, M. A. Moraes, R. Mueller, L. J. Nociolo, J. L. Peticolas, T. Quan, D. Ramot, J. K. Salmon, D. P. Scarpazza, U. Ben Schafer, N. Siddique, C. W. Snyder, J. Spengler, P. T. P. Tang, M. Theobald, H. Toma, B. Towles, B. Vitale, S. C. Wang, C. Young, Anton 2: Raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC '14, IEEE Press, Piscataway, NJ, USA, 2014, pp. 41–53. doi:10.1109/SC.2014.9.
- [26] J. D. Owens, M. Houston, D. Luebke, S. Green, J. E. Stone, J. C. Phillips, Gpu computing, Proceedings of the IEEE 96 (5) (2008) 879–899. doi:10.1109/JPROC.2008.917757.
- [27] T. Preis, P. Virnau, W. Paul, J. J. Schneider, Gpu accelerated monte carlo simulation of the 2d and 3d ising model, J. Comp. Phys. 228 (12) (2009) 4468–4477. doi:10.1016/j.jcp.2009.03.018.
- [28] M. Weigel, Performance potential for simulating spin models on gpu, J. Comp. Phys. 231 (8) (2012) 3064–3082. doi:10.1016/j.jcp.2011.12.008.
- [29] F. R. Brown, N. H. Christ, Parallel supercomputers for lattice gauge theory, Science 239 (4846) (1988) 1393–1400. doi:10.1126/science.239.4846.1393.
- [30] P. A. Boyle, D. Chen, N. H. Christ, M. A. Clark, S. D. Cohen, C. Cristian, Z. Dong, A. Gara, B. Joo, C. Jung, C. Kim, L. A. Levkova, X. Liao, G. Liu, R. D. Mawhinney, S. Ohta, K. Petrov, T. Wettig, A. Yamaguchi, Overview of the qcdsp and qcdoc computers, IBM Journal of Research and Development 49 (2.3) (2005) 351–365. doi:10.1147/rd.492.0351.
- [31] P. Hut, J. Makino, Astrophysics on the grape family of special-purpose computers, Science 283 (5401) (1999) 501–505. doi:10.1126/science.283.5401.501.

- [32] T. Fukushige, P. Hut, J. Makino, High-performance special-purpose computers in science, *Comput. Sci. Eng.* 1 (2) (1999) 12–13. doi:10.1109/5992.753041.
- [33] F. Belletti, M. Cotallo, A. Cruz, L. A. Fernandez, A. Gordillo-Guerrero, M. Guidetti, A. Maiorano, F. Mantovani, E. Marinari, V. Martin-Mayor, A. Munoz-Sudupe, D. Navarro, G. Parisi, S. Perez-Gaviro, M. Rossi, J. J. Ruiz-Lorenzo, S. F. Schifano, D. Sciretti, A. Tarancon, R. Tripiccion, Janus: An fpga-based system for high-performance scientific computing, *Comput. Sci. Eng.* 11 (1) (2009) 48. doi:10.1109/MCSE.2009.11.
- [34] M. Baity-Jesi, R. A. R.A. Banos, A. Cruz, L. A. Fernandez, J. M. Gil-Narvion, A. Gordillo-Guerrero, D. Iniguez, A. Maiorano, F. Mantovani, E. Marinari, V. Martin-Mayor, J. Monforte-Garcia, A. Munoz-Sudupe, D. Navarro, G. Parisi, S. Perez-Gaviro, M. Pivanti, F. Ricci-Tersenghi, J. J. Ruiz-Lorenzo, S. F. Schifano, B. Seoane, A. Tarancon, R. Tripiccion, D. Yllanes, Janus ii: A new generation application-driven computer for spin-system simulations, *Comp. Phys. Commun.* 185 (2) (2014) 550–559. doi:10.1016/j.cpc.2013.10.019.
- [35] B. Meyer, J. Schumacher, C. Plessl, J. Forstner, Convey vector personalities - fpga acceleration with an openmp-like programming effort?, in: *Proc. Int. Conf. on Field Programmable Logic and Applications (FPL)*, FPL '12, IEEE, 2012, pp. 189–196. doi:10.1109/FPL.2012.6339259.
- [36] H. Giefers, C. Plessl, J. Förstner, Accelerating finite difference time domain simulations with reconfigurable dataflow computers, *SIGARCH Comput. Archit. News* 41 (5) (2014) 65–70. doi:10.1145/2641361.2641372.
- [37] T. Kenter, J. Förstner, C. Plessl, Flexible fpga design for fdtd using opencl, in: *Proc. Int. Conf. on Field Programmable Logic and Applications*, FPL '17, IEEE, 2017. doi:10.23919/FPL.2017.8056844.
- [38] T. Kenter, G. Mahale, S. Alhaddad, Y. Grynko, C. Schmitt, A. Afzal, F. Hannig, J. Förstner, C. Plessl, Opencl-based fpga design to accelerate the nodal discontinuous galerkin method for unstructured meshes, in: *Pro. Int. Symp. on Field-Programmable Custom Computing Machines*, Vol. 1 of FCCM '18, 2018, pp. 189–196. doi:10.1109/FCCM.2018.00037.
- [39] P. Klavík, A. C. I. Malossi, C. Bekas, A. Curioni, Changing Computing Paradigms Towards Power Efficiency, *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences* 372 (2018) (2014).
- [40] C. Plessl, M. Platzner, P. J. Schreier, Approximate computing, *Informatik Spektrum* 38 (5) (2015) 396–399. doi:10.1007/s00287-015-0911-z.
- [41] M. Lass, T. D. Kühne, C. Plessl, Using approximate computing for the calculation of inverse matrix p-th roots, *IEEE Embedded Systems Letters* 10 (2) (2018) 33–36.

- [42] C. M. Angerer, R. Polig, D. Zegarac, H. Giefers, C. Hagleitner, C. Bekas, A. Curioni, A fast, hybrid, power-efficient high-precision solver for large linear systems based on low-precision hardware 12 (2016) 72–82. doi:10.1016/j.suscom.2015.10.001.
- [43] A. Haidar, P. Wu, S. Tomov, J. Dongarra, Investigating half precision arithmetic to accelerate dense linear system solvers, in: 8th Workshop on Latest Advances in Scalable Algorithms for Large-Scale Systems, ScalA17, ACM, 2017. doi:10.1145/3148226.3148237.
- [44] A. Haidar, S. Tomov, J. Dongarra, N. J. Higham, Harnessing gpu tensor cores for fast fp16 arithmetic to speed up mixed-precision iterative refinement solvers, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis, SC '18, IEEE, 2018. doi:10.1109/SC.2018.00050.
- [45] S. Gupta, A. Agrawal, K. Gopalakrishnan, P. Narayanan, Deep learning with limited numerical precision, in: Proceedings of the 32nd International Conference on International Conference on Machine Learning, ICML'15, ACM, 2015, pp. 1737–1746.
- [46] NVIDIA Corporation, Tesla P100 data sheet (Oct. 2016).
- [47] The Next Platform, Tearing apart Google's TPU 3.0 AI coprocessor, Online Article (May 2018).
URL <https://www.nextplatform.com/2018/05/10/tearing-apart-googles-tpu-3-0-ai-coprocessor/>
- [48] Top 500, Intel lays out new roadmap for AI portfolio, Online Article (May 2018).
URL <https://www.top500.org/news/intel-lays-out-new-roadmap-for-ai-portfolio/>
- [49] R. Strzodka, D. Goddeke, Pipelined mixed precision algorithms on fpgas for fast and accurate pde solvers from low precision components, in: 14th Annual IEEE Symposium on Field-Programmable Custom Computing Machines, IEEE, 2006. doi:10.1109/FCCM.2006.57.
- [50] T. Kenter, G. Vaz, C. Plessl, Partitioning and vectorizing binary applications, in: Proc. Int. Conf. on Reconfigurable Computing: Architectures, Tools and Applications (ARC), Vol. 8405 of Lecture Notes in Computer Science, Springer, 2014, pp. 144–155. doi:10.1007/978-3-319-05960-0_13.
- [51] T. Kenter, H. Schmitz, C. Plessl, Pragma based parallelization - trading hardware efficiency for ease of use?, in: 2012 International Conference on Reconfigurable Computing and FPGAs, IEEE, 2012. doi:10.1109/ReConFig.2012.6416773.

- [52] F. R. Krajewski, M. Parrinello, Linear scaling electronic structure calculations and accurate statistical mechanics sampling with noisy forces, *Phys. Rev. B* 73 (2006) 041105. doi:10.1103/PhysRevB.73.041105.
- [53] D. Richters, T. D. Kühne, Self-consistent field theory based molecular dynamics with linear system-size scaling, *J. Chem. Phys.* 140 (13) (2014) 134109. doi:10.1063/1.4869865.
- [54] K. Karhan, R. Z. Khaliullin, T. D. Kühne, On the role of interfacial hydrogen bonds in “on-water” catalysis, *J. Chem. Phys.* 141 (22) (2014) 22D528. doi:10.1063/1.4902537.
- [55] J. Hutter, M. Iannuzzi, F. Schiffmann, J. VandeVondele, Cp2k: atomistic simulations of condensed matter systems, *WIREs Comput. Mol. Sci.* 4 (2014) 15–25. doi:10.1002/wcms.1159.
- [56] M. Z. Bazant, E. Kaxiras, Modeling of covalent bonding in solids by inversion of cohesive energy curves, *Phys. Rev. Lett.* 77 (1996) 4370–4373. doi:10.1103/PhysRevLett.77.4370.
- [57] M. Z. Bazant, E. Kaxiras, J. F. Justo, Environment-dependent interatomic potential for bulk silicon, *Phys. Rev. B* 56 (1997) 8542–8552. doi:10.1103/PhysRevB.56.8542.
- [58] A. Ricci, G. Ciccotti, Algorithms for brownian dynamics, *Mol. Phys.* 101 (2003) 1927–1931. doi:10.1080/0026897031000108113.
- [59] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.* 81 (8) (1984) 3684–3690. doi:10.1063/1.448118.
- [60] T. D. Kühne, M. Krack, M. Parrinello, Static and dynamical properties of liquid water from first principles by a novel car-parrinello-like approach, *J. Chem. Theory and Comput.* 5 (2) (2009) 235–241. doi:10.1021/ct800417q.
- [61] R. Z. Khaliullin, T. D. Kühne, Microscopic properties of liquid water from combined ab initio molecular dynamics and energy decomposition studies, *Phys. Chem. Chem. Phys.* 15 (38) (2013) 15746–15766. doi:10.1039/C3CP51039E.
- [62] H. Scheiber, Y. Shi, R. Z. Khaliullin, Communication: Compact orbitals enable low-cost linear-scaling ab initio molecular dynamics for weakly-interacting systems, *J. Chem. Phys.* 148 (23) (2018) 231103. doi:10.1063/1.5029939.
- [63] K. A. F. Röhrig, T. D. Kühne, Optimal calculation of the pair correlation function for an orthorhombic system, *Phys. Rev. E* 87 (2013) 045301. doi:10.1103/PhysRevE.87.045301.

- [64] V. Kapil, M. Rossi, O. Marsalek, R. Petraglia, Y. Litman, T. Spura, B. Cheng, A. Cuzzocrea, R. H. Meißner, D. M. Wilkins, B. A. Helfrecht, P. Juda, S. P. Bienvenue, W. Fang, J. Kessler, I. Poltavsky, S. Vandenbrande, J. Wieme, C. Corminboeuf, T. D. Kühne, D. E. Manolopoulos, T. E. Markland, J. O. Richardson, A. Tkatchenko, G. A. Tribello, V. V. Speybroeck, M. Ceriotti, i-PI 2.0: A universal force engine for advanced molecular simulations, *Comp. Phys. Commun.* 236 (2019) 214–223. doi:10.1016/j.cpc.2018.09.020.
- [65] A. M. Steane, Efficient fault-tolerant quantum computing, *Nature* 399 (1999) 124–126. doi:10.1038/20127.
- [66] E. Knill, Quantum computing with realistically noisy devices, *Nature* 434 (2005) 39–44. doi:10.1038/nature03350.
- [67] J. Benhelm, G. Kirchmair, C. F. Roos, R. Blatt, Towards fault-tolerant quantum computing with trapped ions, *Nature Physics* 4 (2008) 463–466. doi:10.1038/nphys961.
- [68] J. M. Chow, J. M. Gambetta, E. Magesan, D. W. Abraham, A. W. Cross, B. R. Johnson, N. A. Masluk, C. A. Ryan, J. A. Smolin, S. J. Srinivasan, M. Steffen, Implementing a strand of a scalable fault-tolerant quantum computing fabric, *Nature Commun.* 5 (2014) 4015. doi:10.1038/ncomms5015.
- [69] G. Efstathiou, M. Davis, C. S. Frenk, S. D. M. White, Numerical techniques for large cosmological N-body simulations, *The Astrophysical Journal* 57 (1985) 241–260. doi:10.1086/191003.
- [70] L. Hernquist, P. Hut, J. Makino, Discreteness Noise versus Force Errors in N-Body Simulations, *The Astrophysical Journal* 402 (1993) L85. doi:10.1086/186706.
- [71] M. Lass, S. Mohr, H. Wiebeler, T. D. Kühne, C. Plessl, A massively parallel algorithm for the approximate calculation of inverse p-th roots of large sparse matrices, in: *Proc. Platform for Advanced Scientific Computing Conference, PASC '18*, ACM, New York, NY, USA, 2018, pp. 7:1–7:11. doi:10.1145/3218176.3218231.