Reply to reviewer #2

A The authors (A) honestly thank reviewer #2 (R) for his/her critical reading, his/her extensive comments, and the time and effort spent with reviewing the manuscript.

R The authors present Monte Carlo simulations of a lattice gas using a dynamic model that breaks detailed balance. They determine a quantity called asymmetry parameter, which measures the breaking of detailed balance, and show that it is nonzero. They also present results for an off-lattice gas model which seems to behave in a similar manner. These findings are related to recent NMR experiments.

   I strongly disagree with the main statements in the paper. In my opinion the work suffers from serious conceptual deficiencies regarding both the design of the model and the interpretation of the results, which is why I would absolutely not recommend it for publication in a regular journal. However, I understand that in this journal, the referee reports will be published alongside with the paper. Therefore, a publication might be acceptable as long as some additional technical issues have been fixed.

   Let me start with the technical issues first

R The central quantity, the asymmetry parameter, is never properly defined. The only definition is found in Equation (5) which refers to the special case of a three-site exchange. The authors must add an equation defining the quantity which is actually measured in the simulations and shown in Figures 4 and 6.

A The quantity reported in Figs. 4 and 6 is the very quantity defined in Eqn. (5). It follows from the master Eqn. (1) for the special case of three sites at mass equilibrium, when the time derivative on the left-hand side of (1) vanishes. The general case for n-site exchange as described in the van Kampen text can be worked out accordingly by introducing general indices $(k_{ij}M_j - k_{ji}M_i)$. $k_{ij}M_j$ quantifies the number of particles moving from site j to site i per time. This number corresponds to the peak integral in a T_2-T_2 NMR exchange map after correction for the impact of NMR relaxation. Thus, it is a quantity that can be determined experimentally with the help of forward modelling and inverse Laplace transformation (see Eqn. 7). Note, that T_2 is a relaxation time for nuclear magnetization which describes the decay of the impulse response measured in NMR experiments. The denominator of Eqn. (5) is the sum over all particle moves, so that Eqn. (5) specifies the imbalance of the fractions of particles that move from pool j to pool i versus those that move from pool i to pool j. Most readers of this journal are familiar with the T_2-T_2 experiment or can refresh their memory with the references given in the manuscript. In practice, a two-dimensional exchange map measured with NMR contains peaks on the diagonal and cross-peaks right and left of the diagonal. It has been a question of concern in parts of the NMR community, if such an exchange map must always be symmetric with respect to its diagonal. If detailed balance is obeyed, then the corresponding cross-peak integrals right and left of the diagonal are the same and the map is symmetric. We investigate three-site exchange as the simplest case where T_2-T_2 exchange maps could formally

become asymmetric. In this case the master equation (1) at mass balance demands that the cross-peak integrals in the three-site T_2-T_2 NMR exchange map on each side of the diagonal always obey Eqn. (4). This was critically observed in each simulation by counting particle jumps to and from three different pools. The total number of observations in the simulations was chosen so that the agreement between the three symmetry parameters following from Eqn. (4) was usually at least within two decimals.

R   Likewise, a so-called "active site" seems to be an important ingredient either of the dynamical model or in the analysis (this does not become clear), but it is never defined. It has "different relaxation properties" but relaxation properties have not been introduced in the definition of the model before. As an "explanation", the caption of Figure 3 offers the following cryptic sentence: "If a particle cell contacts two different relaxation sites, the higher number overrides the lower number when identifying its relaxation environment." What does this mean in practice? Does the presence of an active site change the dynamics or is it just important for the analysis? And how exactly is this implemented?

A   The active site is a chemical term used in the context of heterogeneous catalysis, where diffusion of molecules in and out of nanopores is essential, and somewhere in the nanopore there is a chemical site that activates the catalysis of a chemical reaction. Often the active site is highly paramagnetic, i. e. molecules in contact with the site have a short NMR relaxation time T_2, which differs from the NMR relaxation time T_2 of molecules in contact with other parts of the pore wall (intermediate T_2) and molecules in the bulk fluid phase (long T_2). This is a simplified explanation, and the NMR experts know, that the pore size and the exchange kinetics play an important role here. The presence of the active site does not impact the dynamics, but it gives a handle on identifying three distinct pools of particles diffusing inside the pore. In our simulations the membership of a diffusing particle belonging to one of three pools is identified by their relaxation time. If the particle is without contact to other particles or in contact with only particles in the bulk, then it belongs to pool 1. If it is in contact with a particle from the bulk and with the wall, it belongs to pool 2. If it is in contact with either the wall (pool 2) or the bulk (pool 1) and the active site, then it is counted to belong to pool 3. This is what the statement in the caption to Fig. 3 is meant to express, and this is how the dynamics were implemented in the program.

R   Apart from the active site element, I think I roughly understand the dynamical model of the lattice simulations, but the off-lattice simulations (Section 2.2) are not well explained at all. Simulations of hard particle models would typically be done using event-driven algorithms, where the system is propagated from one elastic collision to the next. Apparently, this was not done here, instead fixed time steps were used, which reduces the accuracy of the simulations. How exactly were the collisions implemented? For example, did the authors accurately account for the impact parameter of each collision when calculating the new momenta of the participating particles, or did they pick them at random? What was the length of

A   The length of the time steps in the algorithm is in arbitrary units. The simulation was set up such that an initial speed of 0.035 corresponds to moving 0.035 arbitrary length units in 1 arbitrary time unit. This can be converted to SI units, but it is unnecessary as the focus is on the number of times the particles change relaxation sites. At each time step, the distance between each possible pair of particles was considered. If the center of each particle was within one diameter of another, the particles are considered to have collided. Immediately after a collision the projection of the velocity vector along the collision axis is reversed prior to propagating to the next step. At these low occupancy numbers, in the very rare occasion that more than two particles simultaneously collide, the projection of each particle's velocity vector on the collision axis is reversed prior to propagating to the next step. The impact of the observation interval on the asymmetry parameter has been reported in Fig. S6.

R   Error bars are missing throughout. They must be added in the graphs, also the numbers in the text should be given with errors, especially those for (nonzero) asymmetry parameters.

A   Thank you for making us aware of the missing error discussion. We apologize for not having noticed it disappear in the process of revising our draft manuscript multiple times before submission. In the submitted manuscript the simulation error has only indirectly been addressed in line 421, where the mirror symmetry of the observed asymmetry parameter with the position of the active site in the pore wall is pointed out. Also following the comments of reviewer #1, we will add a statement, that the total number of observations in the simulations was chosen so that the agreement between the three signed symmetry parameters stated in Eqn. (4) was usually met within at least two decimals. Thus, error bars in the graphs would be too small to be visible.

R   Given the complexity of the model, the code should not just be "available upon request", it should be published together with the manuscript. This holds especially for the off-lattice code.

These issues must be fixed before the paper can be published.

A   We will include both codes in the supplement.

Next I will summarize the conceptual deficiencies in the presentation of the paper.

Monte Carlo model :

   Description of the model :

R   *Helmholtz free energy*: On page 7, it is claimed that "the particle motion is governed by the Helmholtz free energy A". However, the Helmholtz free energy is a global thermodynamic quantity and does not govern local microscopic dynamics. Probably, the authors to refer to some kind of effective coarse-grained potential here.

   Dynamics and Boltzmann distribution: Same page, the authors state "The probability of a particle moving from one cell to another is given by the Boltzmann distribution $p = \exp(-\Delta A/k_B T)$". This statement does not make sense, as already apparent from the fact that the "probability" p can be larger than one, $p > 1$ for $\Delta A < 1$. It is also not consistent with the subsequent description of the algorithm, where it becomes clear that the probability of moving to a certain site also depends on the number of equivalent accessible sites etc.

A   This point has been addressed in line 220: "If for one or more jumps $p \geq 1$, the destination cell of the jump is picked at random from this subset of all potential jumps."

R   **Design of the model** :

   Internal energy: The internal energy change after moving one particle is described as $\Delta U = \mathbf{F}\Delta \mathbf{R}$ (page 7), where $\mathbf{F}$ is the a force acting on a particle that is constructed from the occupancy of neighboring sites. First, there is an obvious sign error there, probably a typo, it should really read $\Delta U = -\mathbf{F}\Delta \mathbf{R}$: The energy decreases if the particle follows the force. For example, in a gravitational field, if you roll downhill, your potential energy decreases.

A   Thank you for pointing out the sign error in $\Delta U = -\mathbf{F}\,\Delta \mathbf{R}$. We will correct that. Instead of the symmetric expression $\Delta U = -(\mathbf{F}_{final}-\mathbf{F}_{initial})\Delta \mathbf{R}$ we have chosen the computationally faster asymmetric difference neglecting $\mathbf{F}_{final}$ obtaining $\Delta U = \mathbf{F}_{Initial}\Delta \mathbf{R}$.

R   Second, and more seriously, it is easy to see that this specific force field, as it is formulated on a lattice, is not conservative. For example, consider a system where one particle is fixed at the origin, and a second particle undergoes a cyclic motion from $(1,0) \rightarrow (2,0) \rightarrow (2,1) \rightarrow (1,1) \rightarrow (1,0)$. Then the total internal energy change after the cycle is not zero, even though the final and initial configuration are exactly the same. Therefore, this lattice force field cannot be derived from a potential.

A   This is an interesting point. So, on average, energy can be injected into the system at each jump, which may give rise to the observed imbalance (asymmetry) of the three-site exchange. On the other hand, we observed the asymmetry also without prioritizing jumps based on probabilities calculated from our crude approximations of internal energy and entropy in the vacancy-diffusion case. This is why we checked our calculations with a different scenario, i. e. for confined gas diffusion of spheres colliding elastically and starting with an arbitrary velocity distribution. This distribution thermalized rapidly and produced asymmetry parameters in accordance with vacancy diffusion (Fig. 6e). Therefore,

if energy is injected onto the system, then it must be the case for both scenarios, vacancy diffusion and gas diffusion, in which particles diffuse randomly without thermodynamic constraints. So far, we cannot see where in both algorithms that can happen.

R   *Entropy*: The probability of moving to a neighbor lattice site is associated with an entropy change, which is estimated by the sum of step lengths to unoccupied neighbor cells. This specific form of entropy is entirely heuristic and again, it cannot be derived from an effective entropy potential. One should also note that it is not necessary to include translational entropy in a proper Monte Carlo algorithm: The algorithm will automatically account for it.

A   Thank you for your clarifying remark. Yes, the definition of entropy is entirely heuristic and only has essential features of the statistical entropy. Please also see our response to first reviewer: Our use of these crude approximations for S and $\mathbf{F}$ turned out to be important for discovering the nonzero asymmetry parameter in the vacancy-diffusion simulations in the first place as it resulted in a large asymmetry parameter. The gas-diffusion simulations were conducted without consideration of any interaction energy other than elastic collisions between spheres. Following these simulations, we executed the vacancy diffusion simulations again but now with $\Delta U=0=\Delta S$ and at 100 times higher number of jumps. Only then we discovered that there is still nonzero asymmetry albeit about one to two orders of magnitude lower (Fig. 6e, black curve). Our crude approximations of ΔU and ΔS for the calculation of jump probabilities do not impact the conclusion of nonzero asymmetry parameter, but the use of jump probabilities is discussed in the manuscript, because it was decisive in the discovery process of the computer simulation experiments. We will state that in the revised manuscript.

R   *Jump probability* (page 8): From the previous two points, it is already clear that the quantity ΔA in the expression for p cannot be associated with a well-defined effective potential A. However, even if such a potential existed, the choice of jump probabilities seems rather arbitrary. For example, page 8 says "If $0 < p < 1$, the destination cell is chosen at random from all those with the same largest jump probability $p < 1$". This is not well motivated. Why not choose from all cells with weighted probabilities according to their jump probability?

A   If the jump probability turned out to be larger than one it was artificially set to 1. Also, on occasion it happened that the largest jump probabilities were smaller than 1 and that the largest two or more probabilities were equal. In those two cases the destination cell was picked at random from the respective subsets to proceed in an unbiased way.

R   The algorithm described here is not motivated by any microscopic considerations. With the same right, assuming that ΔA could really be derived from a global effective potential function A, one could also use a standard Metropolis algorithm, which would satisfy detailed balance by construction.

A   Choosing an algorithm that satisfies detailed balance by construction precludes testing detailed balance.

R   *Summary*: The presented Monte Carlo algorithm does not satisfy detailed balance for two reasons: First, even though the notation suggests otherwise, the underlying quantities p are not associated with a well defined effective energy function A. Second, the jump probabilities are chosen heuristically according to some random rules which are not well-motivated. It is not surprising that these rules do not satisfy detailed balance, because imposing detailed balance usually requires special efforts.

In fact, these rules would not even guarantee global balance if A were well-defined. On the other hand, they do define some kind of stochastic Markovian dynamics, and according to the central limit theorem of finite Markov systems, the probability distribution will converge against some stationary fixed point, which however differs from the Boltzmann distribution N exp($-\beta A$). Furthermore, this stationary state would include persistent currents by default, because, as explained above, special efforts must be taken to remove them in such a model.

A   We repeat, that our choice of p does not impact our main conclusion of nonzero asymmetry parameter, but it is mentioned in the manuscript, because it was important in the discovery process of the computer simulation experiments. We found nonzero asymmetry parameters also when choosing the destination cell for a jump from all vacant neighbor cells at random.

Interpretation of the results :

R   *Thermodynamic equilibrium*: The term "thermodynamic equilibrium", by definition, refers to a stationary state *without currents*. One of the central postulates of thermodynamics is that every physical closed dynamical system reaches thermodynamic equilibrium at some point. This is a postulate and might be debated. However, a system with persistent currents as described in the manuscript would not be considered to be at thermodynamic equilibrium.

A   Perhaps the wording of current is misleading. The result of the simulations is that there is a cyclic exchange between random particles from one of three pools and the next pool when this movement happens under the impact of topological constraints. Our simulations suggest that this step is not balanced in detail on the particle scale, i. e. on the scale a few particle diameters from the pore wall. This means, that depending on the sign of the asymmetry parameter, a particle may jump with slight preference directly from the wall to the bulk at certain positions near the wall as compared to jumping form the bulk to the wall. Instead, the particles arriving at these special positions on the wall prefer to first move along the wall before jumping back to the bulk. These positions of a large asymmetry parameter depend on the geometry of the confinement. This is an explanation at the molecular level. The notion of currents refers to a macroscopic picture. We will check our manuscript if we have used the right wording, made this point sufficiently clear, and correct where necessary.

R   *Detailed balance and nonequilibrium thermodynamics:* As correctly stated in the manuscript, the lack of currents is associated with microscopic detailed balance – or, putting it the other way round, breaking detailed balance normally generates

currents. However, this also implies that entropy is constantly being produced, and dissipated, see, e.g., References [1-3].

A   See previous response and thank you for the literature references. In view of this issue, we refer to Feynman's rachet in the manuscript at line 154.

R   Dynamical systems with broken detailed balance have been discussed in nonequilibrium thermodynamics for many decades. Physically, they are used to describe open dissipative systems, for example, living systems or active systems [1,2], which are stabilized via a steady input of energy. It is easily possible to design stochastic dynamical systems that break detailed balance, as has been done, e.g., in the present manuscript or in Refs. [3,4]. In Monte Carlo simulations, implementing such dynamics can have the advantage that a desired probability distribution function can be sampled much more efficiently [4].

A   Thank you for these further references. Indeed, there are other steady-state cases that can be mentioned. We refer to some in the introduction, lines 40-44.

R   *Detailed balance and Monte Carlo:* The Monte Carlo method has been introduced by Metropolis et al as a method to efficiently sample a desired target probability distribution. The necessary ingredient for this is to impose global balance. Detailed balance is not strictly necessary. With the exception of kinetic Monte Carlo (which has not been used here), Monte Carlo dynamics is typically not realistic.

Nevertheless, Monte Carlo is also used to study dynamical systems in a coarse-grained sense. However, it is important to note that in this type of model, you get out what you put in. If you implement Monte Carlo moves that break detailed balance, then clearly, you will find that detailed balance is broken in your system. Therefore, Monte Carlo simulations designed to model dynamics at thermal equilibrium must be set up such that the Monte Carlo moves satisfy detailed balance.

Is detailed balance always fulfilled? As stated above, the claim that closed physical dynamical systems reach thermodynamic equilibrium is a postulate. It lies at the heart of the second law of thermodynamics, but being a postulate, it could be violated in certain cases. In fact, it is violated, e.g., for integrable systems such as linear harmonic chains. It has not been proved rigorously except for a few special cases.

On the other hand, the opposite claim that detailed balance might be broken in realistic (closed) physical system fundamentally challenges the foundations of thermodynamics. Such a claim cannot be based on Monte Carlo simulations. This is because, as explained above, Monte Carlo dynamics are inherently artificial, and it is much easier to implement dynamical models that break detailed balance than to implement models that satisfy detailed balance. The claim would have to be based on experiments, or on molecular simulations of a truly microscopic model, e.g., classical Hamiltonian dynamics or Schr¨odinger dynamics. In fact, there have been several claims in the past, based on atomistic simulations, that the second law might be broken in nanoscale systems. For example, spontaneous unidirectional currents through pores or the like were observed in

simulations. In all of these cases, it eventually turned out that the claimed effects could be attributed to numerical artefacts of the simulations.

A   We understand that you are saying that Monte Carlo methods cannot be used for this type of study on a molecular scale. We concede, that although both algorithms produce the same results, there is no proof, that for some reason still unknown to us they both produce the same false result. This is why we are currently working on experiments in addition to the one mentioned in the introduction to corroborate or falsify the computer-simulation results, see lines 575-578 in the manuscript. In view of your comments, we will formulate our conclusions more as a question.

R   The central question is whether a system can thermalize, which is a valid question especially for nanoscale systems and subject of active research. Specifically, the gas diffusion case discussed in the manuscript is related to the question whether a classical ideal gas can thermalize. This is one of the few cases which has been studied very intensely and for which rigorous results exist (the H-theorem, see [5]). Ideal gases do thermalize! In the manuscript, nonideal gases with excluded volume interactions are considered, which might change the situation, but I would be very surprised if it did. This is one of the reasons why it is so important that the authors describe their simulations for the gas diffusion simulations in more detail. If they maintain the claim that detailed balance is broken in these (off lattice) systems, they should prove it much more carefully, e.g., by systematic variation of the time step, by studying the relaxation of several quantities as a function of simulation time, and by a solid assessment of error bars.

A   Thermalization of the gas diffusion has been tested starting from both a Maxwell distribution and from a Dirac distribution of all particles possessing the same initial speed. For both large and small pores, the velocity distributions approach Maxwell-Boltzmann which is apparent within the first 1000 steps. We will add a sentence to that end in the revised manuscript and show the approach of an initial Dirac distribution to the Maxwell-Boltzmann distribution of speeds for the gas-diffusion case at three time intervals, initial, intermediate, and final, in a revision of the supporting information.

References:

1   C. W. Lynn et al, PNAS 2021, 118, e2108998118.

2   F. S. Gnesotto et al, Rep. Prog. Phys. 2018, 81, 066601.

3   L. Crochik et al, Phys. Rev. E 2005, 72, 057103.

4   M. Michel et al, J. Chem. Phys. 2014, 140, 054116.

5   G. Truesdell, R.G. Muncaster, *Fundamentals of Maxwell's kinetic theory of a simple monatomic gas*, Chapter XI, in *Pure and Applied Mathematics*, Volume 83, pp. 145-172 (1980).