Supplementary Information : Interplay of multiple clusters and initial interface positioning for forward flux sampling simulations of crystal nucleation

Katarina E. Blow[1], Gareth A. Tribello[2], Gabriele C. Sosso[3] and David Quigley[1]
[1)] *Department of Physics, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, United Kingdom*
[2)] *Centre for Quantum Materials and Technologies, School of Mathematics and Physics, Queen's University Belfast, Belfast, BT7 1NN, United Kingdom*
[3)] *Department of Chemistry, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, United Kingdom*

This supplementary information contains several graphs of relevance to this paper. Figure S1 shows the number of boundary crossings as a function of time. Figure S2 presents the effect of using all nuclei for the fluxes in the λ' test (compared to only the primary nuclei in Figure 4). Figure S3 shows retaining atomic velocities and allowing configurations not at $\lambda_0 + \frac{1}{2}$.Figure S4 gives the influence of λ_A placement on the λ' test, and Figure S5 shows the effect of λ_A placement on subsequent interfaces. Figure S6 shows the effects of sampling interval on initial fluxes and the λ' test. Figure S7 displays the weighted histogram distribution function of single solid particles. Finally, Figure S8 displays the radial distribution function of hard spheres.

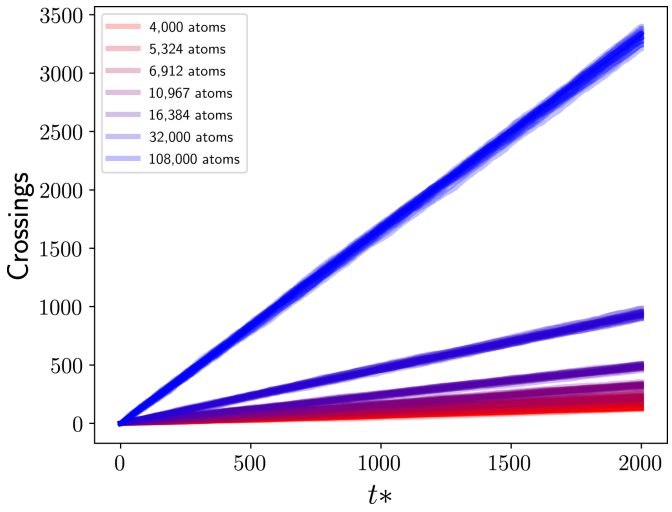

FIG. S1: Number of crossings of $\lambda_0 = 36.5$ as a function of simulation time for different system sizes. All clusters were considered. Counted crossings are the first positive crossing of λ_0 after a return to λ_P (i.e. the count corresponding to $\Phi_{0|\lambda_P}$). Each faint line corresponds to one independent simulation.

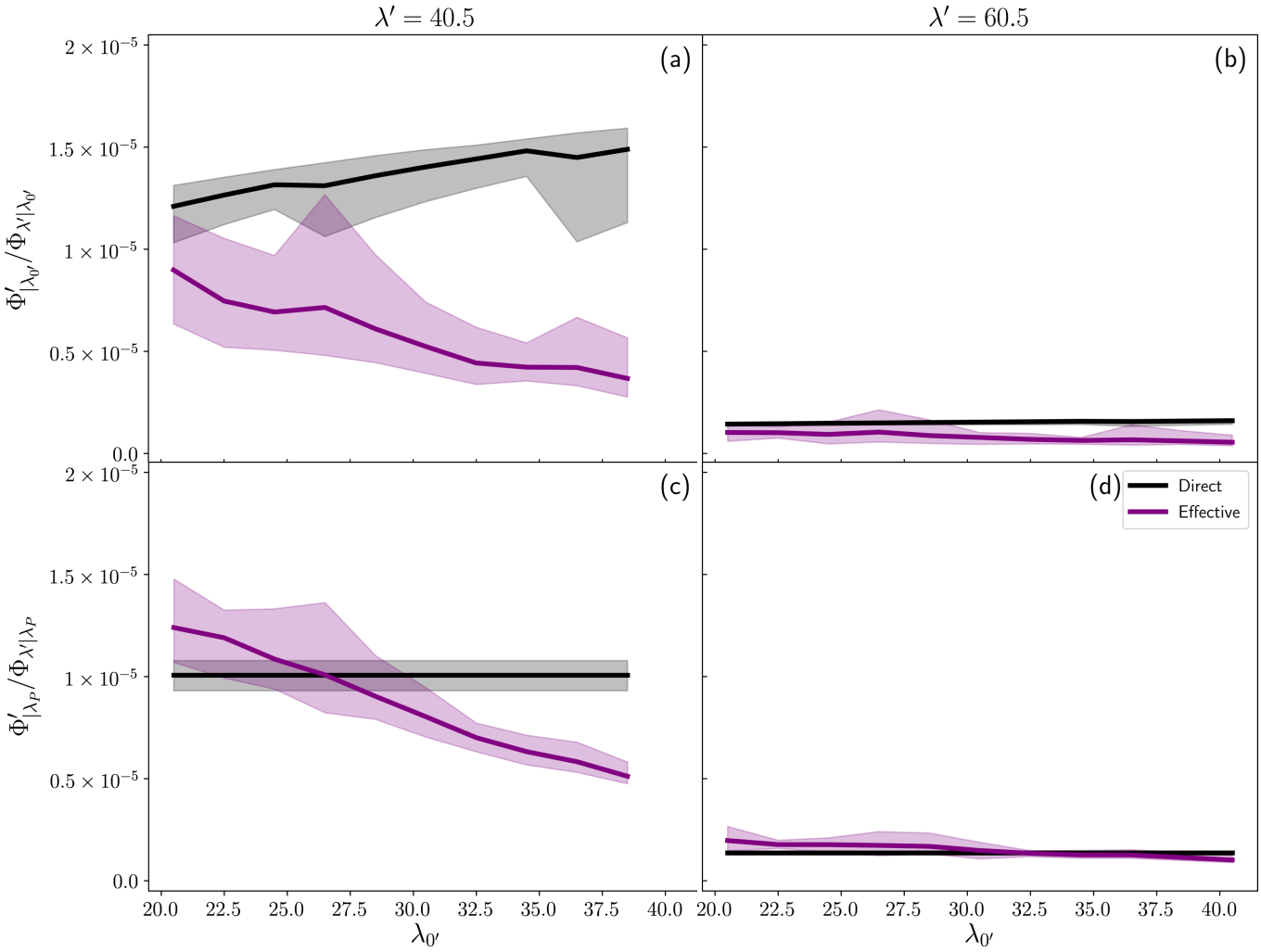

FIG. S2: Fluxes using different values of λ_A and λ'; both the direct flux, $\Phi'_{|\lambda_A}$ and the effective flux $\Phi_{\lambda'|\lambda_A}$ are present on the same scale. (a) $\lambda_A = \lambda_{0'}$, $\lambda' = 40.5$, (b) $\lambda_A = \lambda_{0'}$, $\lambda' = 60.5$, (c) $\lambda_A = \lambda_P$, $\lambda' = 40.5$, (d) $\lambda_A = \lambda_P$, $\lambda' = 60.5$. The black line represents the mean of the direct flux of all nuclei through λ' (requiring a return to λ_A between subsequent crossings). The purple line represents the mean of the effective flux of all nuclei through λ' ($P(\lambda'|\lambda_{0'}) \times \Phi_{0'|\lambda_A}$, where $\Phi_{0'|\lambda_A}$ includes contributions from all nuclei). The shaded region represents the uncertainty on the fluxes as a result of the different volumes of the simulations considered here. The 32,000 and 108,000 atom systems only contribute to relevant fluxes for $\lambda_{0'} > \lambda_P$, even for the panels where the edge of the liquid basin is defined as $\lambda_A = \lambda_{0'}$. For effective fluxes, this is a result of the procedure used to generate initial configurations. For direct fluxes with $\lambda_A = \lambda_{0'}$, this is for a more consistent comparison. Note that the direct flux for $\lambda_A = \lambda_P$ is the only flux independent of $\lambda_{0'}$ and as such the 32,000 and 108,000 atom systems contribute for the entire range. The legend applies to all panels. This is equivalent to Figure 4 with total nuclei crossings.

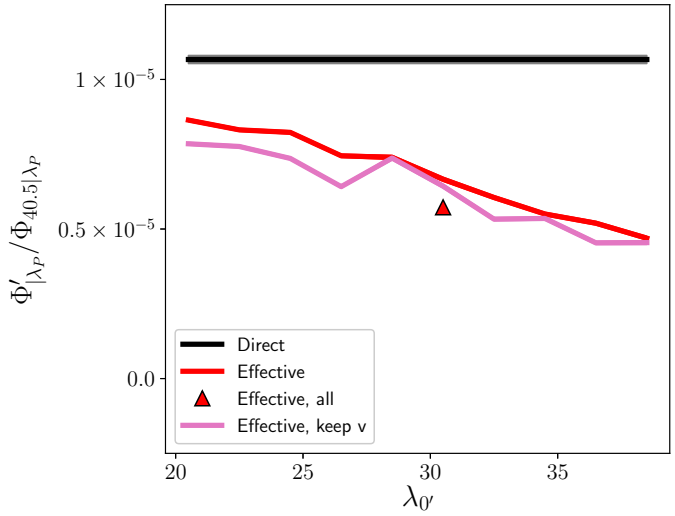

FIG. S3: Direct and effective fluxes of primary nuclei through $\lambda' = 40.5$, for a 4,000 atom system at $\lambda_A = \lambda_P$. The statistical uncertainty is represented by a shaded area, and where not visible is less than line width. "All" represents the effective flux of stored configurations which have crossed $\lambda_{0'}$ from below λ_P but which have a value between $\lambda_{0'} + \frac{1}{2}$ and λ' (i.e. which do not lie on the interface). "Keep v" represents retaining the initial velocities of configurations stored at $\lambda_{0'} + \frac{1}{2}$ which have just crossed λ_0 from below λ_P.

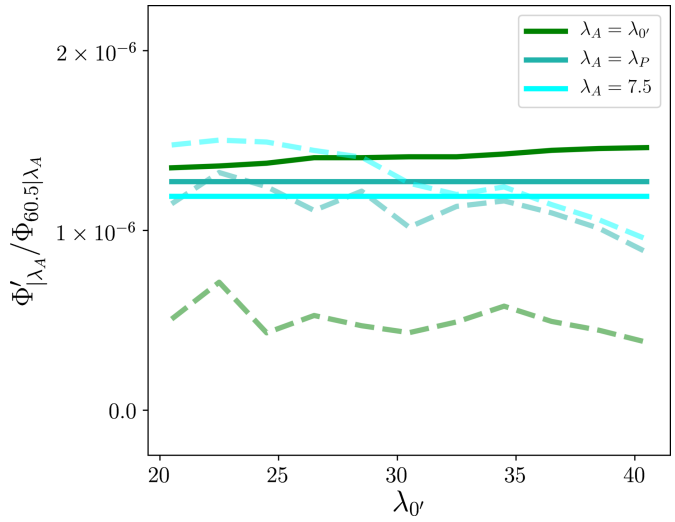

FIG. S4: Direct and effective flux through $\lambda' = 60.5$ for different values of λ_A for a 4,000 atom system. Darker solid lines represent directly observed flux, while the fainter dashed lines correspond to the effective flux. All fluxes consider only primary nuclei, and statistical uncertainties are not shown. Note that the direct flux for $\lambda_A = \lambda_{0'}$ is not constant with $\lambda_{0'}$.

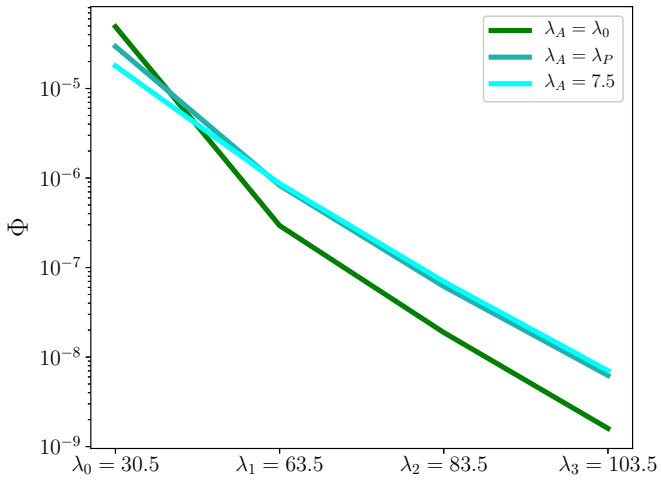

FIG. S5: omparison of fluxes (using primary clusters only) through several FFS interfaces for different values of λ_A for a 4,000 atom system at $T* = 0.86$ and $p* = 5.68$. All fluxes consider only primary nuclei, and statistical uncertainties are not shown. The choice of interfaces has not been optimised in any way, apart from ensuring that $\lambda_1 > 2\lambda_0$ to minimise the influence of merging, where the OP was computed every 100 MD timesteps (of $t* = 0.002$). 10,000 trials were initialized from each interface, and at each subsequent interface the first 500 configurations were stored (although there were in some cases fewer than 500 stored configurations). At each interface, the stored configurations were run for 100 MD timesteps before the OP was computed and it was determined if a crossing had occurred.

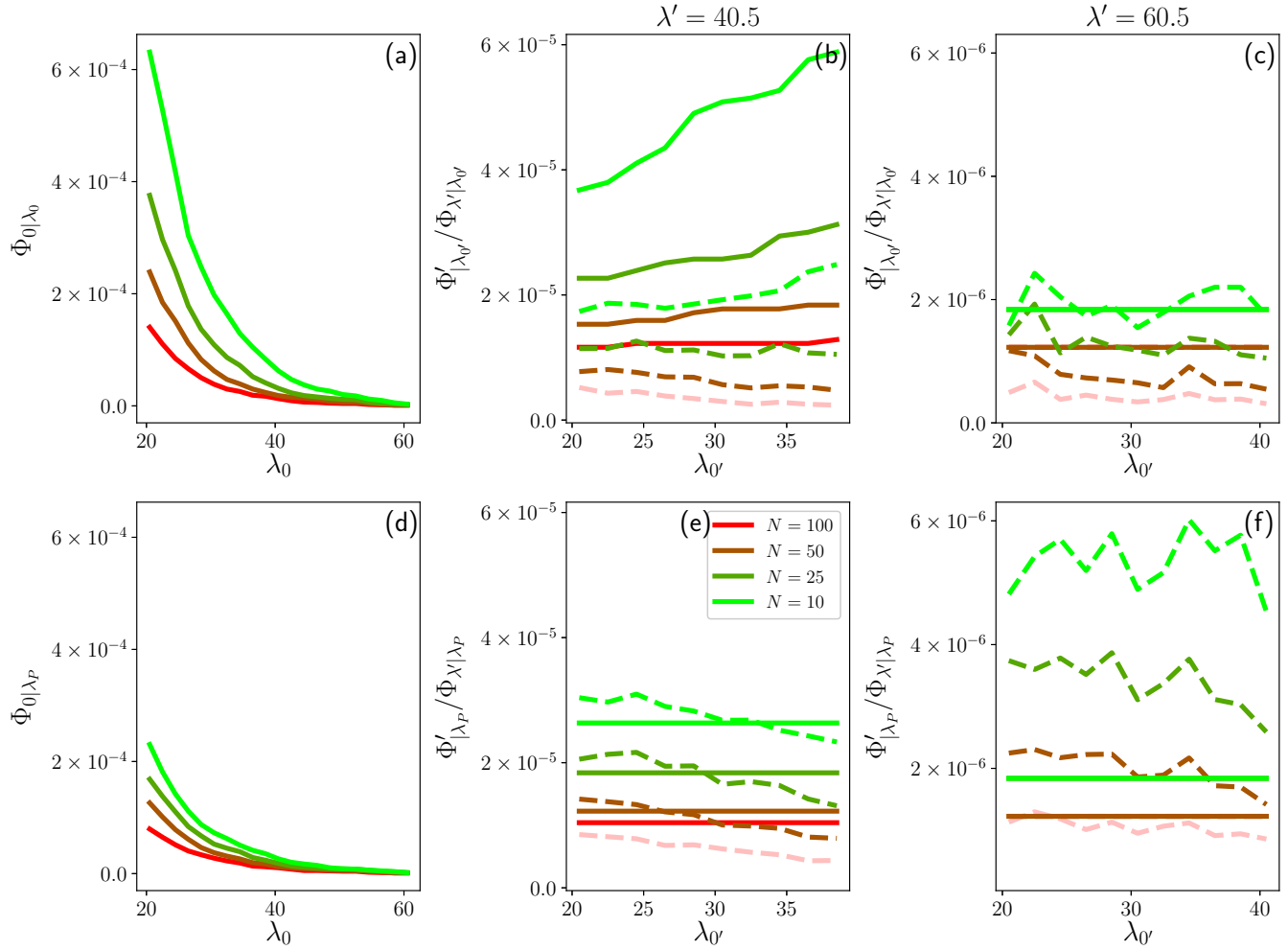

FIG. S6: Fluxes of primary nuclei, for a 4,000 atom system with a range of sampling intervals using different values of λ_A and λ_1. N represents the number of time steps between snapshots (OP calculations). (a) Primary flux, $\lambda_A = \lambda_0$ (b) direct and effective fluxes, $\lambda_A = \lambda_{0'}$, $\lambda' = 40.5$, (c) direct and effective fluxes $\lambda_A = \lambda_{0'}$, $\lambda' = 60.5$, (d) primary flux, $\lambda_A = \lambda_P$, (e) direct and effective fluxes $\lambda_A = \lambda_P$, $\lambda' = 40.5$, (f) direct and effective fluxes $\lambda_A = \lambda_P$, $\lambda' = 60.5$. Darker solid lines represent directly observed flux, while the fainter dashed lines correspond to the effective flux. Data was generated from a single 200,000 time step run so statistical uncertainties are not present. The legend applies to all panels. In panel (f) the $N = 25$ and $N = 10$ direct fluxes and $N = 50$ and $N = 100$ direct fluxes overlap at this scale, as do the $N = 50$ and $N = 100$ direct fluxes in panel (c). Note that in panel (f) the large discrepancy between direct and effective fluxes for $N = 10$.

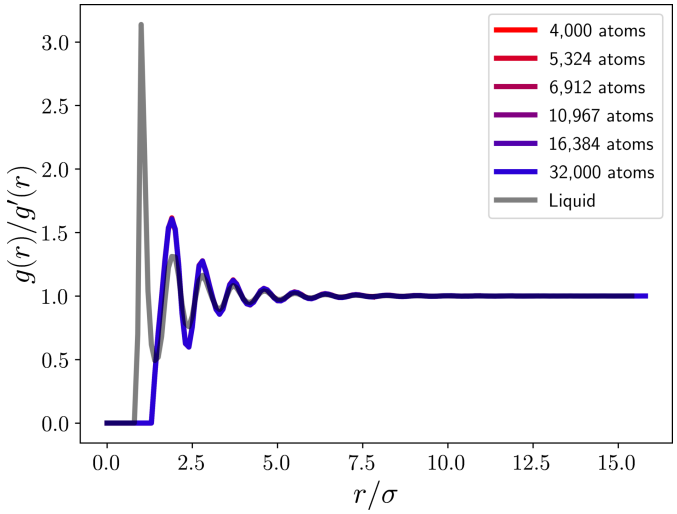

FIG. S7: $g'(r)$ of single solid particles, and $g(r)$ of the liquid (same scale) at $T* = 0.86$ and $p* = 5.68$. The 108,000 atom system has been omitted due to computational cost. The liquid $g(r)$ was calculated as the $g(r)$ of all particles for the first 50 frames of a 32,000 atom trajectory (as used for analysis of solid particles/clusters). The incidence of a peak in the liquid $g(r)$ before the solid $g'(r)$ is due to the fact that all solid particles within 1.432 σ are considered neighbors and therefore belong to the same cluster. Statistical uncertainties in the solid $g'(r)$ are present but less than the line width.

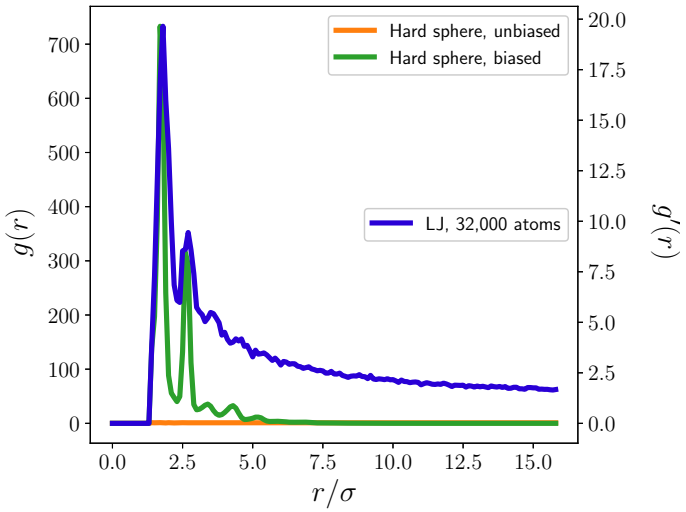

FIG. S8: $g'(r)$ of clusters of size 20 and above for the 32,000 atom LJ system at $T* = 0.86$ and $p* = 5.68$, and $g(r)$ of hard spheres of radius 0.716 in a box of side length 31.8. This size of the hard sphere radius was chosen to map the extent of the hard spheres to the maximum distance between connected atoms. For the LJ system, only the 32,000 atom system is shown for ease of comparison. The hard sphere data corresponds to the minimum distances between all images of 25,000 instances of multiple spheres, with the number of spheres selected according to a binomial distribution. The orange line corresponds to the pair-correlation function of all images of hard spheres placed randomly throughout the simulation volume. The green line corresponds to preferentially placing new hard spheres at two set distances (with applied Gaussian noise for peak smoothing) around existing spheres. These distances are 1.8 and 2.7 (1.5×1.8), although this choice is purely for mapping to the separation of the peaks in the LJ system. Statistical uncertainties in the solid $g'(r)$ are present but less than the line width.