Journal of Computational Science 71 (2023) 102060

Contents lists available at ScienceDirect

Journal of Computational Science

journal homepage: www.elsevier.com/locate/jocs

Establishing metrics to quantify spatial similarity in spherical and red blood cell distributions

Sayan Roychowdhury [a], Erik W. Draeger [b], Amanda Randles [a,*]

[a] Duke University, Durham, NC, USA
[b] Lawrence Livermore National Laboratory, Livermore, CA, USA

ARTICLE INFO

Keywords:
Red blood cells
Microvascular simulation
Cell packing

ABSTRACT

As computational power increases and systems with millions of red blood cells can be simulated, it is important to note that varying spatial distributions of cells may affect simulation outcomes. Since a single simulation may not represent the ensemble behavior, many different configurations may need to be sampled to adequately assess the entire collection of potential cell arrangements. In order to determine both the number of distributions needed and which ones to run, we must first establish methods to identify well-generated, randomly placed cell distributions and to quantify distinct cell configurations. We utilize metrics to assess (1) the presence of any underlying structure to the initial cell distribution and (2) similarity between cell configurations. We propose the use of the radial distribution function to identify long-range structure in a cell configuration and apply it to a randomly distributed and structured set of red blood cells. To quantify spatial similarity between two configurations, we make use of the Jaccard index, and characterize sets of red blood cell and sphere initializations. As an extension to our work submitted to the International Conference on Computational Science (Roychowdhury et al., 2022), we significantly increase our data set size from 72 to 1048 cells, include a similar set of studies using spheres, compare the effects of varying sphere size, and utilize the Jaccard index distribution to probe sets of extremely similar configurations. Our results show that the radial distribution function can be used as a metric to determine long-range structure in both distributions of spheres and RBCs. We determine that the ideal case of spheres within a cube versus bi-concave shaped cells within a cylinder affects the shape of the Jaccard index distributions, as well as the range of Jaccard values, showing that both the shape of particle and the domain may play a role. We also find that the distribution is able to capture very similar configurations through Jaccard index values greater than 95% when appending several nearly identical configurations into the data set.

1. Introduction

Computational blood flow models are a powerful tool for answering biomedical questions. For microvessel simulations, where individual cell diameters are on the same order of magnitude as vessel size, the presence of cells plays a significant role in the non-Newtonian behavior of blood. In this regime, velocity profile blunting has been observed due to the motion of cells towards the vessel centerline [1] and blood viscosity has been shown to be dependent on vessel diameter and hematocrit (volume percentage of cells in the blood) [2]. Additionally, cell-to-cell [3] and cell-to-vessel interactions [4] have been shown to affect the underlying blood flow profile. Therefore in small vessel simulations, blood must be modeled as a suspension of cells rather than a continuum fluid. Fluid–structure interaction (FSI) models, such as the immersed boundary method [5] or dissipative particle dynamics [6], which fully couple deformable particles with a background fluid, have been shown to accurately model cells in microfluidic [7] and microcirculatory systems [8]. Blood flow simulations using FSI models provide a wealth of information, as both microscopic and macroscopic quantities, such as individual cell position and deformation, and fluid pressure and velocity profiles, can be precisely tracked and studied over time [9,10]. More importantly, these models allow for the isolation and controlled variation of specific parameters such as cell size or stiffness, enabling researchers to probe the effects of individual parameters on the quantity of interest. Much of the *in silico* work in microvessels with cell FSI models has been focused on red blood cells (RBCs), including studies on the effects of cell deformability and shape [11–13], partitioning at junctions in the vasculature [14,15], aggregation mechanics [16], and development of a cell-depleted layer [17,18]. Simulation has also been used to study the motion of other particles

* Corresponding author.
E-mail address: amanda.randles@duke.edu (A. Randles).

https://doi.org/10.1016/j.jocs.2023.102060
Received 2 November 2022; Received in revised form 13 April 2023; Accepted 7 May 2023
Available online 16 May 2023

in the presence of RBCs such as platelets [19,20], leukocytes [21,22], and circulating tumor cells [10,23–26].

While FSI models of cells in complex geometries are not new, advances in computational efficiency and capability [27–30] have only recently made this approach practical for comprehensive studies of realistic systems. The inclusion of explicit cells in particular introduces several new obstacles. The main challenge is simply one of statistics: the motion of microparticles diffusing through a vessel is an inherently stochastic process, thus trajectories must be sampled a sufficient number of times to capture average behavior.

The same problem holds when abstracting from cells to granular materials, such as grains of sand or rice, or atomic particles. When interested in the microscopic behavior at the level of individual particles, a single simulation is unable to capture the range of possibilities. Using a cellular example, when tracking cancer cells *in silico*, the distance to a vessel wall directly influences the cell's likelihood of adhesion [31] and subsequent escape into nearby tissue. We previously demonstrated the effects of varying cell positions while studying combinations of hemodynamic parameters and the motion of a tumor cell [23]. Even when all bulk fluid parameters were held constant, the trajectory of the individual cell was found to vary significantly based on the relative configurations of neighboring cells.

In addition to increasing the overall computational cost, the need for a representative ensemble of starting configurations introduces new potential sources of error that must be managed. This challenge is particularly acute for systems with higher hematocrit values, where random coordinate generation must be done carefully to avoid artificial structure that would bias the observed dynamics. Similar to the well-known equilibration problem in molecular dynamics [32], flow simulations through tortuous vascular geometries have the added complication that one cannot easily gather equilibrated statistics simply by running a closed system longer in time. Instead, one must generate a number of distinct sets of equilibrated starting points to be run independently [33]. To this end, we propose a method to generate many cell configurations and the use of the radial distribution function (RDF) to characterize the structure in a particular configuration. The final challenge is to define quantitative metrics to rigorously compare individual cell configurations and to characterize the complete set as a whole. For this purpose, we propose the use of the Jaccard index (J) to quantify spatial similarity between individual configurations as an appropriate metric for describing and comparing sets of cell configurations.

A comprehensive method for sampling the entire space of potential configurations is to generate a large data set of starting configurations. While this method is useful, it is important to establish that the data set is truly representative of the entire space. For instance, it is possible to generate hundreds of configurations, each with numerous cells, but with only minor differences such as a single cell being slightly rotated. In this scenario, despite the high number of configurations, they would only be representative of a limited pool of possibilities. In order to ensure that the data set is representative of the entire space, it is crucial to have a systematic, quantitative approach for capturing the volumetric similarity between configurations.

This is where J comes into play as a useful tool for quantifying the volumetric similarity between configurations. By calculating the Jaccard index between each pair of configurations, it is possible to determine whether the data set is diverse and representative of the entire space, or if it is confined to a local set of possibilities. This approach can provide valuable insights into the breadth of the data set and can help to improve the variety in starting conditions.

In our previous work [34], we described several of these methods, including a procedure to quickly generate many configurations of particles, an application of the RDF on individual configurations to assess the long-range structure or randomness in particle placement, and a lattice-based numerical method to calculate volumetric spatial similarity between configurations using the Jaccard index between bi-concave shaped red blood cells. RBCs serve as a uniquely shaped material, but since they are by far the most common type of cell in the bloodstream, any type of microvascular simulation including cells will require the packing of RBCs.

In this work, we extend our study as follows. First, we now apply these metrics not just to RBCs, but also to ideal spheres. Spheres, which are often used as an ideal model for many different particles, are used as a baseline comparison to the irregularly-shaped RBCs. We compare the volumetric spatial similarity between sphere configurations again using the Jaccard index, but this time with an analytical method to calculate sphere overlap. Both sets of particles are used to study how the number of configurations affects the distribution of Jaccard index values and in order to compare the two, we maintain similar levels of volume fraction for both systems. Then, we significantly increase the size of the data set used for volumetric comparison, from 72 individual configurations to over 1000. Within this data set, we also break down the results by packing fraction in order to minimize the effects of hematocrit variation. Next, we study the effects of changing sphere size on the J index distributions. Finally, using the quantitative J value distribution, we show that we can describe entire systems of configurations to identify extremely similar cases by appending minutely perturbed configurations to the data set. Thus, we make the step towards choosing the most representative set of configurations that can describe the entire space of possibilities.

2. Methods and metrics

First, we will describe the approach we use to generate many cell configurations. Then we introduce the main methods employed in this research to (1) assess these configurations are random, (2) quantify spatial similarity, (3) characterize a set of configurations through the distribution of their similarities. These three techniques are summarized in Fig. 1.

2.1. Generating initial configurations of materials

Dense packing of non-overlapping shapes is a long-standing research problem of active interest [35–37]. Although a maximally dense packing of same-size spheres can reach volume fractions of 63%, typical values for RBCs are approximately 40%–45% in major vessels [38]. However at this scale, blood is typically treated as a whole continuous fluid [39]. In the microvessel regime, the volume fraction can range anywhere from 5 to 35% [40]. For this work, we chose a target volume fraction of 25% within an idealized 30 μm diameter microvessel. While the metrics we present should be generally-applicable and independent of density, this regime is sufficiently dense to ensure significant overlap in both correlated and uncorrelated distributions but dilute enough to simplify random configuration generation. For spheres at these packing fractions, a simple random distribution of cell centers along with a collision check can quickly hit these values. However due to the irregular, bi-concave shape of RBCs, packing many versions of configurations at the same hematocrit is not as simple as the spherical case, especially in microvessels. The unique RBC shape is not amenable to quick collision checks, and the tight cylindrical domain can significantly limit the hematocrit when randomly packing the vessel. A cursory packing is unable to hit 25% hematocrit in this vessel, and longer computational time would be required to iteratively fit more RBCs in. Multiplying this by the number configurations, the computational time spent purely in this pre-processing step quickly increases before any simulation work has even begun.

Here we describe a better procedure for generating and characterizing packed configurations of RBCs in arbitrary vessels. Rather than generating individual configurations on demand, we instead start with a large system of packed cells from which we can fill vessels of arbitrary size and shape. This technique has the advantage of letting us generate

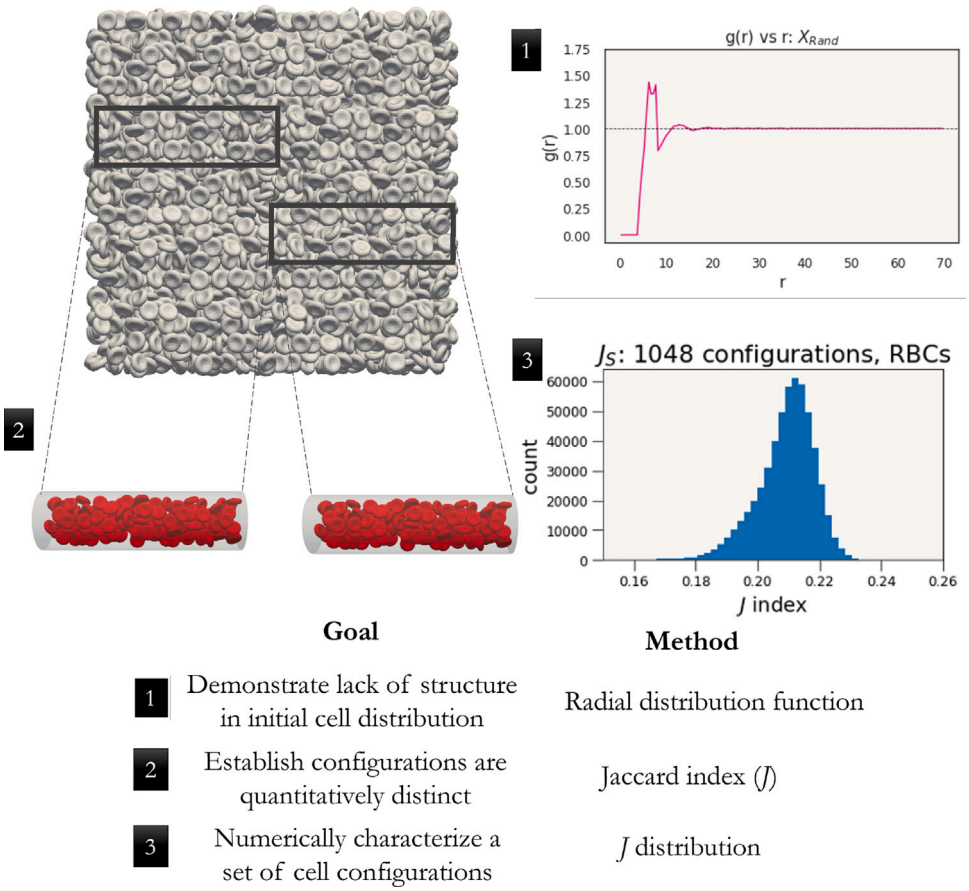

Fig. 1. Summary of three techniques used to determine the best set of cell configurations to describe the ensemble. (1) The radial distribution function $g(r)$ is used to assess randomness in a distribution of cells. (2) The Jaccard index J is used to quantify spatial similarity between two configurations. (3) A set of pairwise J values are used to numerically describe a large set of cell configurations and presented as a distribution.

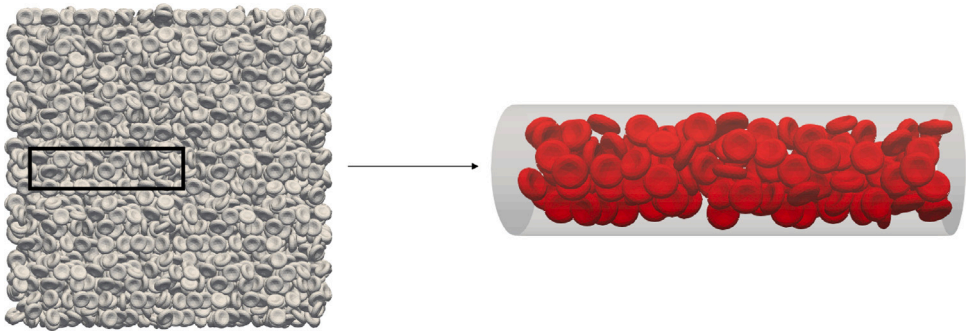

Fig. 2. An example of a cell initialization taken by submerging the vessel within a large, pre-generated packing domain. Only cells that fit completely inside the vessel are returned and used as the starting point for a simulation.

a packed domain in the simplest possible geometry prior to simulation while avoiding the code complexity an on-the-fly implementation would require. The source domain is created to be several times larger than the vessel of interest. The standalone implementation provided by Birgin et al. [41] is used to pack ellipsoids that tightly encompass the RBC's biconcave shape, returning a set of non-overlapping positions and orientations. Although the fully enclosed RBC represents approximately 70% of the encompassing ellipsoid volume, a distribution with a packing fraction of up to 60% is enough to reach the high end of microvascular hematocrit levels. An example of this packing and a corresponding cell initialization is shown in Fig. 2. Testing vessels ranging in diameter from 20 to 50 μm shows the ability to reach realistic hematocrits from 20% to 35% consistently.

This approach of separately generating a packed source domain has the advantage of easily allowing for rigorous *a priori* analysis before performing expensive microvascular simulations.

Once the bulk source geometry has been generated, individual RBC configurations can be created by submerging the target vessel in the source domain at different locations and selecting all cells contained within. This process remains the same for both simple and complex geometries, establishing a straightforward method for generating many different configurations prior to running HPC simulations.

2.2. Assessing randomness within an individual configuration

To avoid initializing FSI runs with non-physical starting configurations, the source bulk system must not have any long-range order

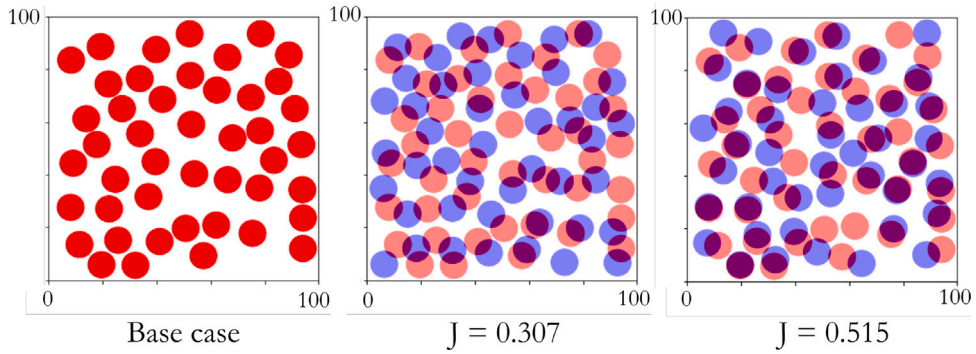

Fig. 3. A 2-D example of differing Jaccard indices compared to a base configuration (left) using circular particles. Two other distinct configurations are generated and overlaid on the base configuration (middle and right). The middle configuration is less similar to the base case than the right configuration and can be confirmed visually by identifying the overlapping violet regions and computed using J.

consistent with crystalline packing. The radial distribution function $g(r)$ is a well-established metric in the simulation of fluids [42] used for confirming liquid structure, defined as:

$$g(r) = \frac{dn_c}{4\pi r^2 dr\rho}$$

where dn_c returns the number of cells within a shell of thickness dr and ρ is the bulk density, and returning the probability of finding another particle at a distance r. Long-range structure is reflected in the form of multiple peaks well beyond the average particle spacing. Particles are assumed to start from a fully disordered liquid state, reflected by a $g(r)$ that quickly converges to a constant value of unity.

2.3. Quantifying spatial similarity between configurations

After generating a set of multiple material configurations, the next step is to verify that each of these packings are distinct by quantifying their spatial similarity to each other; specifically, the fraction of volume shared by two configurations. The Jaccard index, or intersection over union, is used to measure the similarity between two discrete sample sets, defined as:

$$J(C_i, C_j) = \frac{|C_i \cap C_j|}{|C_i \cup C_j|}$$

where C_i and C_j are independent samples of the same space. For a configuration of spheres, the shared volume in space by spheres S_1 and S_2 with radius r separated by distance d can be calculated by

$$V_{overlap}(S_1, S_2) = \frac{1}{12}\pi(4r + d)(2r - d)^2$$

However, due to the irregular biconcave disk shape of RBCs, a simple analytical algorithm for overlap check given cell positions and orientation angles does not exist. We propose the use of the Jaccard index to quantify the volume overlap between sets of RBCs by comparing the interior grid points. This is similar in approach to the algorithms used by the image segmentation community [43], such as the Dice similarity index. Therefore in this calculation, a numerical method is utilized, where each configuration of RBCs is mapped to a 3-D grid, and overlap is calculated by the number of grid points shared. A grid resolution is chosen based on those of typical lattice-based RBC simulations. If a lattice point is inside of a cell, it is marked as occupied, and two sets of occupied points pertaining to two particular configurations can be used to calculate J via

$$J(C_i, C_j) = \frac{l_{C_i,C_j}}{l_{C_i} + l_{C_i} - l_{C_i,C_j}}$$

where l_{C_i} and l_{C_j} refer to the number of grid points occupied by cells in configuration i and j respectively while l_{C_i,C_j} is the number of lattice points occupied in both configurations.

While there are many potential metrics for comparing structural similarity such as the Sorensen-Dice similarity score, overlap coefficient, and Salton's cosine index, we choose to use the Jaccard index. The Sorensen-Dice coefficient and Jaccard index are monotonically related, but in order to treat the similarity score between two configurations as "overlapping volume in space divided by total potential volume occupied by cells", we choose to use J. We expect that using the Dice score would generate a similar distribution which can be recovered from the Jaccard index using $S = 2J/(1+J)$ and provide the same level of information. The overlap coefficient only represents the percentage of overlap related to the smaller set, which does not allow fair comparison as the set sizes may change. Finally, Salton's cosine index is used to measure similarity between a sequence of numbers, but in our case, where the data is simply a set of binary values (fluid vs cell-occupied), Jaccard index is easier to implement.

2.4. Using the Jaccard index to cut correlated configurations and examples

Since $J(C_i, C_j)$ represents the percentage of overlapping cell volume between C_i and C_j, $J(C_i, C_j) = 100\%$ if two arrangements are identical and zero if there is no shared cell volume in space.

A threshold value ϕ is chosen to label whether or not two configurations are correlated; if $J(C_i, C_j) > \phi$, the pair is marked similar. For example, two test configurations that contained the same group of cells shifted a micron led to J over 90%, and would be marked as a similar pair. Because the likelihood of two configurations of cells both occupying a certain space increases with volume fraction occupied by the material, we expect that ϕ is not a static value.

A two-dimensional visual example is shown in Fig. 3, displaying distinct configurations of circles with significantly different J values. This example uses spheres with a radius of 4 units in a 100 by 100 domain. Compared to the base configuration, there is a clear difference in overlap, which can be identified visually and captured quantitatively through an analytical computation of J.

Another two-dimensional example for calculating J using RBCs is provided in Fig. 4. Each initialization contains a single cell marked red for configuration 1 and blue for configuration 2. The corresponding lattice points are marked with the color of containing cell. Once these two lattices are overlaid, the shared points are marked in yellow. We note that this is a purely hypothetical example where the lattice spacing is enlarged for visual clarity.

Although these examples are visualized in two dimensions for clarity, the sphere and RBC configurations are performed in three-dimensional space. The Jaccard index provides a numerical method to identify spatial similarity rather than a qualitative comparison.

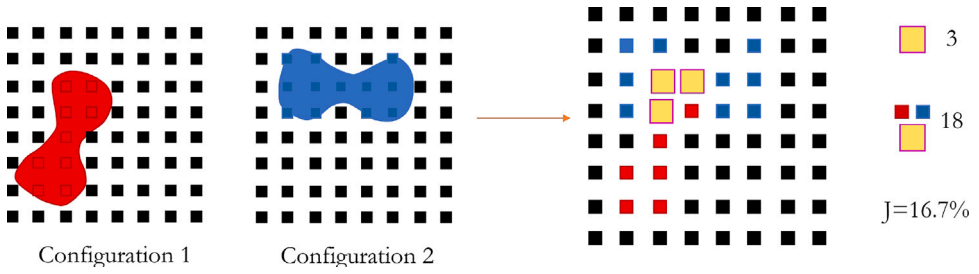

Fig. 4. A 2-D example for the calculation of Jaccard index J between two configurations of RBCs. After the RBCs are mapped to their corresponding lattices, there are 18 total points which contain a cell, of which 3 are shared in both configurations which are marked by the larger grid points. $J = 3/18 = 16.7\%$ in this example.

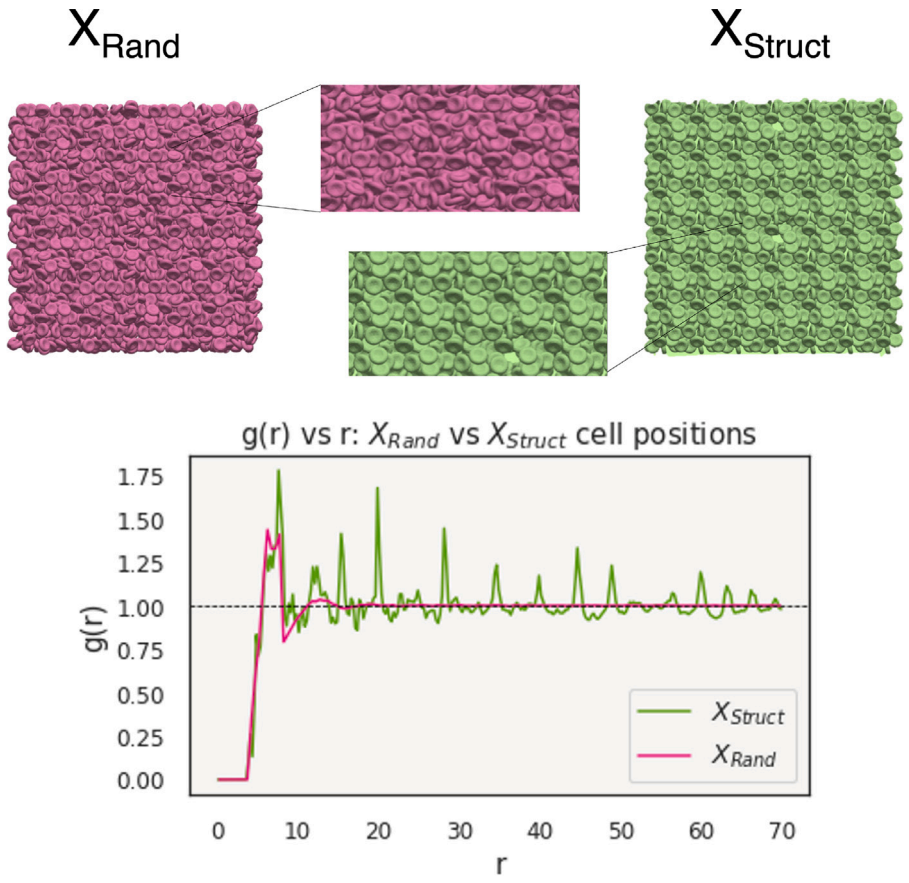

Fig. 5. A random distribution X_{rand} (top left) and ordered distribution X_{struct} (top right) of cells packed within a cube of side length 200 μm. The corresponding radial distribution functions are shown below the two. X_{rand}'s $g(r)$ shows a single peak and trails off to 1 quickly, analogous to a random liquid-like state, while X_{struct}'s $g(r)$ displays several peaks, indicating that the distribution contains a repetitive structure.

2.5. Using Jaccard indices to describe an entire set of configurations

Since the Jaccard index is applied between two particular configurations, J needs to be calculated on a pairwise basis before it can be used to quantify the entire distribution of configurations. For a set of configurations $S = \{C_1, \ldots, C_n\}$, we define J_S, the set of Jaccard similarity scores, as:

$$J_S = \{J(C_i, C_j) \big| i, j = 1 \ldots n, i \neq j\}.$$

To quantify the similarity of a particular configuration C_i with respect to all the others, the mean Jaccard index $\bar{J}(C_i)$ is calculated as:

$$\bar{J}(C_i) = \frac{1}{n-1} \sum_{j=1}^{n} J(C_i, C_j), j \neq i$$

for a set of n configurations. Given two similar configurations C and C' such that $J(C, C') > \phi$, and mean Jaccard indices such that $\bar{J}(C) < \bar{J}(C')$, configuration C' would be considered first for removal from the set.

2.6. Computing details

Simulations are set up on the Summit supercomputer using HARVEY, a massively parallel computational fluid dynamics solver [44,45]. After cells are created in the beginning of the simulation, lattice points are iteratively marked as either occupied by fluid or cell. These resulting positions and occupancies are saved into output files for each initialization. A parallel python script is used to perform the calculation of Jaccard index between every pair of samples.

2.7. Procedure

In this section, we present the methodology used to set up the study. We first describe with the procedure to efficiently obtain numerous RBC

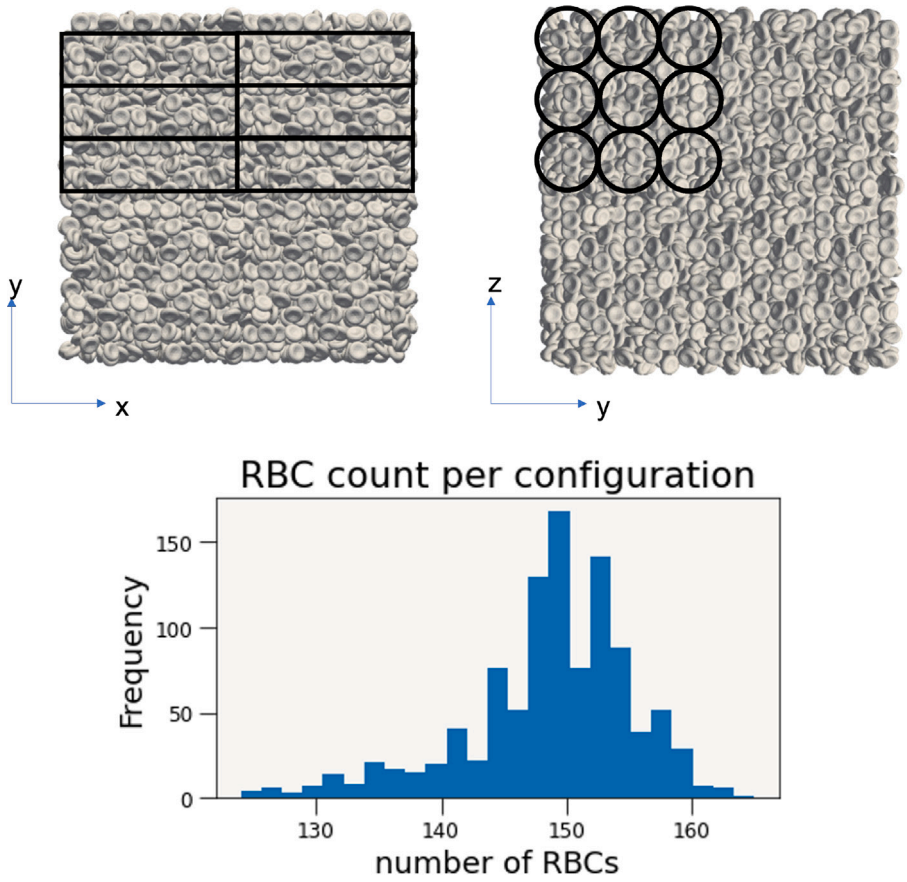

Fig. 6. A numerical method to quickly generate many packings of RBCs is utilized in this work. Up to 200 different initializations for a cylinder of diameter 30 μm and length 100 μm pulled from a packing domain of 200^3 μm^3. Cells within this region are used to generate RBC configurations of initial positions and orientations at a target hematocrit of 25%. A subset of the cutouts are shown in the x-y and y-z planes for this set of initializations. A histogram of the number of RBCs in each configuration is provided, showing that there is variation in the cell count within each cutout.

configurations, establish randomness, and create a comparison with spheres. Finally, we describe the method for using J to identify distinct configurations.

2.7.1. Applying the RDF to quantify structure in the source domains

For RBCs, configurations are sourced from multiple large, dense packings. Three source arrangements are generated in a cubic domain of side length 200, 300, and 400 μm with over 3.5×10^4, 1×10^5, and 2.5×10^5 cells, respectively. These random distributions are created by packing the domain and then applying an external force to perturb the initial arrangement of cells from Birgin et al. [41]. For comparison to the randomly packed 200 μm side length cube's cells, labeled X_{rand}, a structured set of cells X_{struct} is produced by tessellating a small set of RBCs across the 200 μm side length cubic domain. Fig. 5 displays the two examples of these cubic source domains. The radial distribution function is then applied to each set of cell centers splitting dr into 0.5 μm buckets, and the corresponding $g(r)$ functions are shown in the plot in Fig. 5.

2.7.2. Utilizing jaccard index to quantify discrete distributions of RBC and sphere configurations

Non-overlapping cutouts representing an ideal microvessel geometry with diameter 30 μm and length 100 μm are created from the three source packing domains by aligning the cylinder lengths along the x-y plane. Fig. 6 shows an example of cutouts from the 200 μm side length cubic domain X_{Rand}; there are 72 different configurations generated from this particular domain: 6 from the y- and z-planes, and 2 in the x-plane. In total, 1048 configurations are created with a hematocrit spanning approximately 15%–25%, ranging from 125 to 160 RBCs, shown in the histogram in Fig. 6, and are mapped on to a lattice with grid spacing 0.25 μm. The most common cell count is around 150 RBCs. The 1048 are sourced from all 3 packing domains while the 72 are a subset from the 200 μm side length cube from the original work [34].

2.7.3. Assessing the effect of more samples and different sphere radius on Jaccard index distributions

For a baseline comparison, multiple configurations of 160 spheres are numerically generated at a 25% packing density in the same vessel domain. To observe the evolution of the distribution's shape with respect to set size, we sample 72, 200, and 1000 configurations of 160 spheres, and an analytical calculation is used to compute J for each pair of configurations.

2.7.4. Using Jaccard index to identify and remove duplicate configurations

The power of predetermining an entire J distribution lies in its ability to detect extremely similar configurations. When preparing a representative ensemble of simulations, it is important to start with a distinct set of configurations for two reasons. First, including identical groups in the data set may bias the outcome if there is an uneven sample weighting, leading to a skewed set of results. Second, running a group of duplicate simulations without randomness built into the system would simply return the same result, effectively wasting compute resources. By using this method of calculating the shared volume in space to identify similar configurations, we can mitigate these effects.

To show that applying the J score can identify these repeated configurations, we then append several nearly identical configurations to the existing data set of randomly distributed spheres. Since none of these J indices for the original 1000 sphere configurations fall near

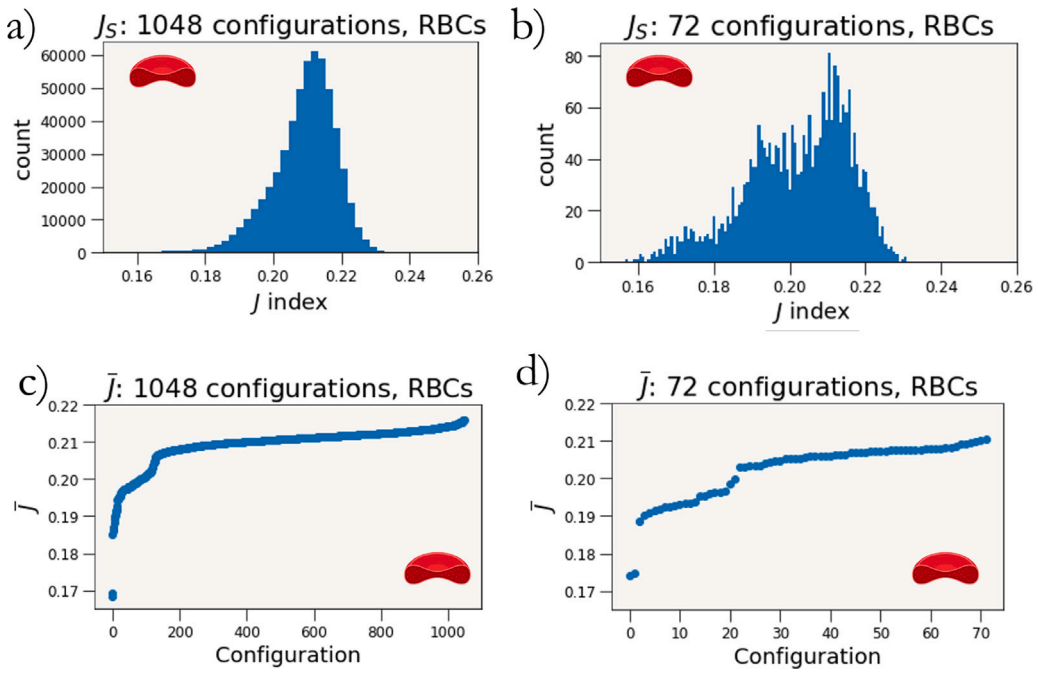

Fig. 7. The distribution J_S of pairwise J values for (a) 1048 and (b) 72 RBC configurations in the 30 μm diameter and 100 μm length cylinder. The respective average overlap of a particular configuration to all of the others are shown in figures (c) and (d).

100%, none are extremely similar to one another. To test the method's ability to parse out these types of configurations, we choose one of the already created configurations and slightly perturb the original positions to generate 100 new configurations in addition to the previous 1000. After these are created, the pairwise J indices are calculated again.

3. Results and discussion

3.1. Applying the RDF to quantify structure in the source domains

Within the three random packings, $g(r)$ contains a single peak near the lengthwise diameter of the RBC that quickly trails off to unity, indicative of a liquid-like, random distribution of particles, as seen in the plot in Fig. 6. For clarity, only X_{struct} for the 200 μm side length domain is presented, but all three random domains display the same behavior. In the case of X_{struct}, multiple discrete peaks are visible, signifying the presence of long-range structure in the distribution of cells. The structure found in the large packing domain remains in the individual configurations generated from them. A qualitative comparison between the two source domains can be performed visually, but the use of the radial distribution functions provide a quantitative confirmation for the presence of ordered structure. Sampling cells from the random distribution provides a better initial set of the positions and orientations of red blood cells for running HPC simulations.

In order to compare a potential correlation in the orientations, We perform a numerical comparison between the angles of every RBC in the each of the source domains. This angle is found by taking the dot product of the normals and computing the angle between RBCs. We find that there is a balanced distribution of angles in between each set of RBCs.

Since the procedure to generate many cell configurations in a microvessel utilizes a subset of the cells in the large domain, it is important to confirm the randomness of the initial cell arrangements. The packing found in X_{struct} is non-physiological, and would generate many structured cell initializations as inputs to HPC simulations. Moving forward, we sample configurations from the three random distributions for RBCs after confirming the lack of long-range structure in its distribution of cells.

3.2. Utilizing jaccard index to quantify discrete distributions of RBC and sphere configurations

The entire set of pairwise Jaccard similarity index values are presented in Fig. 7a, while a subset corresponding to the original 72 configurations' pairwise J scores are shown in Fig. 7b [34]. There is a range of J values from 15 to 24% for all of the pairwise combinations. This shows that there is no significant overlap in any two configurations, likely because completely non-overlapping subdomains are chosen. There is also a skew in the entire distribution towards higher J values, with the most common J index lying around 21%. The 72-configuration distribution also contains a skew in the distribution with peaks between 21 and 22%. We expect the shape of this distribution to stray from normal due to the variety in the number of cells, especially due to the cases where a J index is calculated between two data points with significantly different hematocrits. A skewed shape emerges in both cases which is likely due to the variation in volume fraction resulting from the cylindrical cutout approach, although the bi-concave shape of RBCs may also have an effect on the distribution.

Fig. 7c and d show $\bar{J}$, the average pair overlap of a configuration with all other configurations. This provides a method to compare individual configurations' spatial coverage against the full set. We expect that both J_S and $\bar{J}$ distributions will change based on vessel geometry and hematocrit. However, this study establishes that a pairwise Jaccard index distribution can be used as a quantitative metric to describe a set of cell configurations. We posit that selecting configurations with low $\bar{J}$ could be used to sample the configurational phase space more efficiently.

To mitigate the effects of the variation in hematocrit or volume fraction, the configurations are segmented into specific cell counts ±1, ranging from 135 to 160 cells, corresponding to hematocrits of approximately 18% to 25% respectively. The resulting pairwise J scores for each bucket are shown in Fig. 8. Additionally, the exact same number of configurations for each bucket are generated with spheres occupying

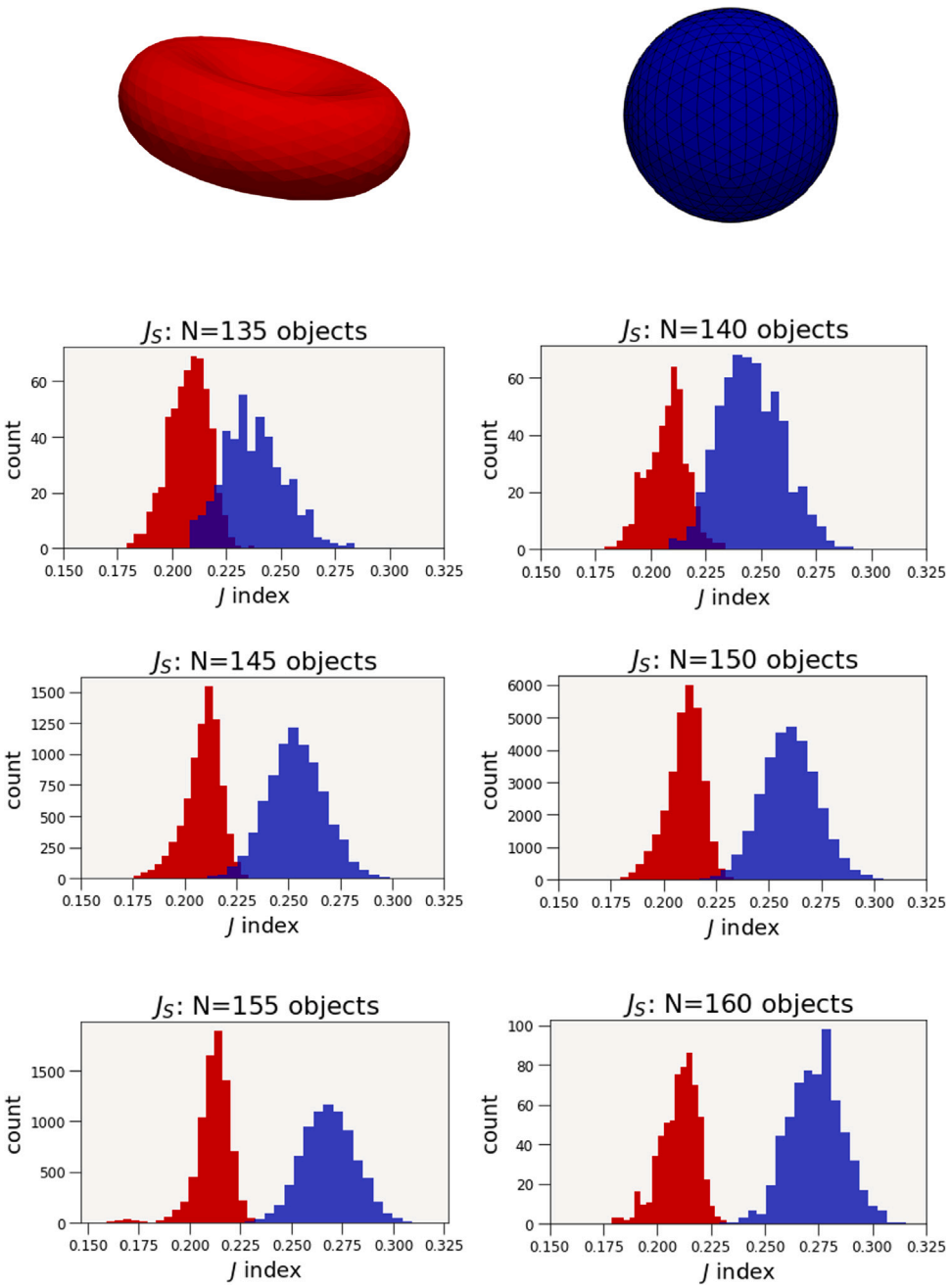

Fig. 8. The distribution J_S of pairwise J values filtered by the number of objects; 135 through 160 RBCs are plotted in red. The same domain is filled with sphere configurations at an equivalent volume fraction and the corresponding J_S distributions are plotted along the same axes in blue.

the same volume fraction and the corresponding J data is plotted alongside the RBC data. This allows for a close comparison of the effect of the particle shapes on J distribution as the hematocrit changes.

The overlap index of configurations of randomly-placed spheres generally follows a normal distribution. The distribution of RBCs, on the other hand, is skewed away from normal in each case. This may be an artifact of the packing algorithm used to populate the source distribution or may be a fundamental difference in how biconcave disks pack into a confined geometry; more work will be needed to elucidate the underlying cause. The magnitude of the average overlap differs significantly between the shapes as well despite all systems having the same volume packing fraction. Additionally, we notice that while the average J index remains roughly the same for the different RBC counts, as the number of spheres increases, the average J shifts higher. It should be noted that the RBC geometry likely has a systematic underestimation of the overlap due to discretization error, though this is not expected to be large. The spherical overlap was computed analytically as a function of distance between sphere centers.

3.3. Assessing the effect of more samples and different sphere radius on jaccard index distributions

Results for the sphere configurations are presented in Fig. 9a–c in the left hand column. To study the effect of different sphere radii at the same volume fraction, we then pack the ideal microvessel with spheres half the diameter of previous. The resulting J_S distributions at the same number of configurations are presented in the right hand column of Fig. 9.

We observe that as more samples of sphere configurations are introduced, the J_S converges towards a normal distribution, spanning

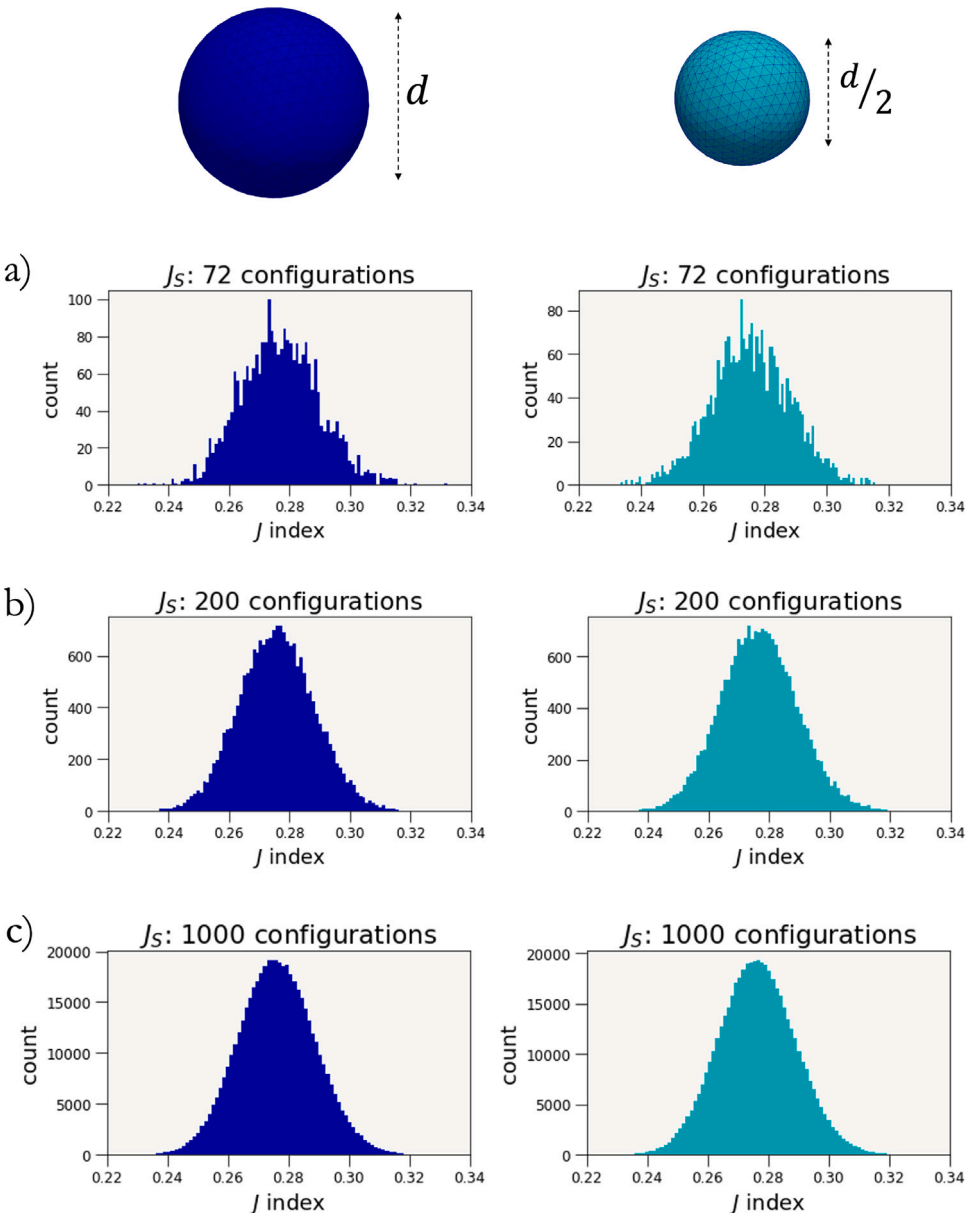

Fig. 9. The distribution J_S of pairwise J values for (a) 72 (b) 200 and (c) 1000 sphere configurations at a 25% volume fraction. The same number of configurations are created with a set of spheres half the diameter of the original and the corresponding J_S distribution is plotted. Both sphere sizes tend to the same normal distribution. As more configurations are added into each data set, from 72 up to 1000, the shape tends to a clear normal distribution.

a range from 22 to 34%. For both sets of sphere diameters, the distributions look extremely similar across all configuration counts. We expect that as the sphere size is increased to larger diameters such that only a handful of objects can be placed within the domain, there will only be a finite set of possible configurations to place the spheres. There will be sectioned off volumes of the domain that will always contain a sphere, no matter the configuration, which could lead to changes in the J distribution. For relevant simulations however, the domain will typically be several times the size of the cell, as is the case with the microvessel.

3.4. Using jaccard index to identify and remove duplicate configurations

Results for the J_S and $\bar{J}$ distributions for the updated set are shown in Fig. 10. In the J_S plot, in addition to the original normal distribution around $J = 30\%$, there is now a clear emergence of another set of peaks to the far right of the plot. Since the new configurations are nearly identical to their original counterparts, they cause the appearance of several pairwise J scores greater than 98%. In the $\bar{J}$ plot, we find that the average J per configuration has increased, but significantly more in a subset of the 1100. One can iteratively remove configurations starting with the highest $\bar{J}$ and see the effects on the J_S distribution. Once there are no more J pair scores that reside in the region close to 100%, the data set can be considered distinct. Applying this method to the 1100 sphere configuration set, the 100 additional configurations are tagged for removal one at a time. Once all of the J pairs above 98% are removed, the J_S distribution returns to one from the original 1000 configurations.

Typical qualitative comparison between configurations, especially on larger scales is often impossible, thus making the use of the quantitative spatial similarity a strong metric for identifying which configurations are extremely similar and may be dropped from the test set. With no prior knowledge of the similarity of this 1100 configurations test case, the application of J indices leading to the plots of J_S and $\bar{J}$ can pinpoint similar pairs. Applying the J scores can help pre-process the data set by identifying these identical configurations and removing

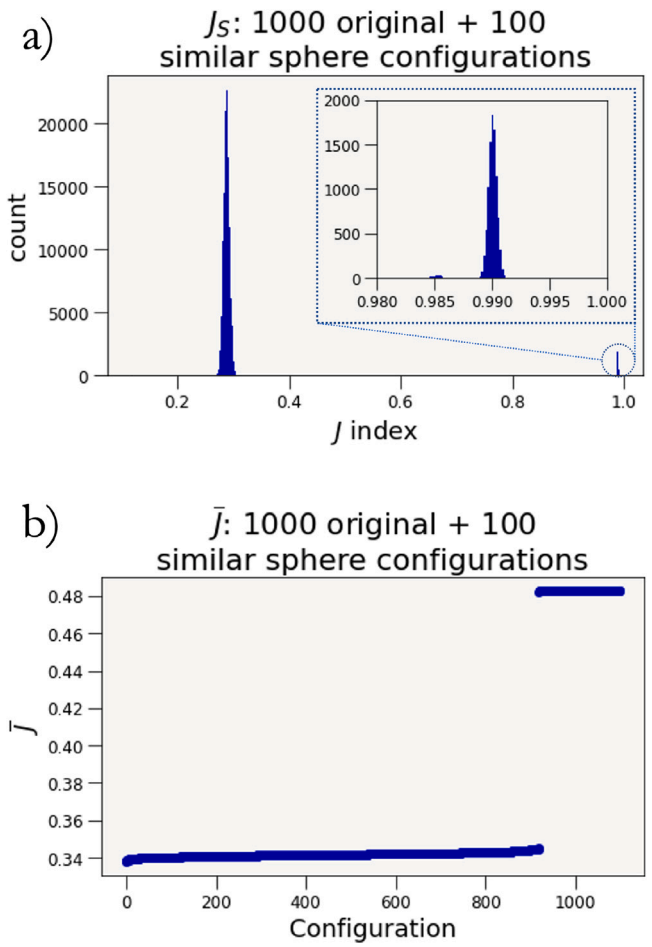

Fig. 10. Distribution of (a) J_S and (b) $\bar{J}$ on a per configuration basis for sphere configurations augmented with a set of 100 configurations with slightly perturbed sphere positions. The J_S distribution now displays a group of J values over 98%, corresponding to the extremely similar configurations. This is also reflected in the values for $\bar{J}$, as the average J significantly increases for a subset of configurations.

them from the initial set of configurations in order to minimize bias and better sample the entire space.

4. Conclusion

In this study, we apply several metrics to spheres and red blood cells within an idealized microvessel. The radial distribution function is used to qualify structured arrangements of particles, while the Jaccard index is utilized to capture shared cell volume between two configurations. We create over 1000 RBC configurations and study pairwise J values, comparing them to randomly generated groups of spheres. We also vary the sphere diameter and add slightly perturbed configurations to the sphere data set. In all of the sphere and RBC studies, we perform pairwise J calculations and plot the distribution of Jaccard index values, showing that this method can be used to define the space of particle configurations.

This study sets the groundwork for identifying the optimal set of initial particle arrangements for a specific group of simulation parameters for microvascular simulations. Without applying this type of quantitative metric, it is nearly impossible to visually identify every single configuration which is extremely similar, especially when the data sets grow to massive sizes. This type of metric can be applied to large sets of initial configurations, effectively pre-processing the data set in order to minimize the number of similar conditions being run. One of the limitations of this work is the dependence of J_S and $\bar{J}$ on several factors, such as packing fraction, vessel geometry, and discretization of the geometry. However, the focus of this work is to introduce these methods as a quantitative metric, and further studies can be performed to determine the effects of varying distributions. Next steps for this work include performing simulation studies with sets of spatially uncorrelated particles configurations to determine how these affect certain outputs, such as motion of individual cells. Future work will also study the effect of different vessel sizes, shapes, and hematocrit on distributions of the Jaccard similarity index.

CRediT authorship contribution statement

Sayan Roychowdhury: Conceptualization, Methodology, Writing – review & editing, Software, Formal analysis, Investigation, Validation, Data curation, Visualization, Writing – original draft. **Erik W. Draeger:** Conceptualization, Methodology, Writing – review & editing, Software, Formal analysis, Investigation, Validation. **Amanda Randles:** Conceptualization, Methodology, Writing – review & editing, Resources, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank Daniel Puleri, Samreen Mahmud, and Cyrus Tanade for their feedback and discussion. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Computing support for this work came from the Lawrence Livermore National Laboratory (LLNL) Institutional Computing Grand Challenge program. The work of Sayan Roychowdhury and Amanda Randles were supported by the National Science Foundation under award number 1943036. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF.

References

[1] G.J. Tangelder, D.W. Slaaf, A. Muijtjens, T. Arts, M. Oude Egbrink, R.S. Reneman, Velocity profiles of blood platelets and red blood cells flowing in arterioles of the rabbit mesentery, Circ. Res. 59 (5) (1986) 505–514.
[2] A.R. Pries, D. Neuhaus, P. Gaehtgens, Blood viscosity in tube flow: dependence on diameter and hematocrit, Am. J. Physiol.-Heart Circ. Physiol. 263 (6) (1992) H1770–H1778.
[3] J.J. Bishop, P.R. Nance, A.S. Popel, M. Intaglietta, P.C. Johnson, Effect of erythrocyte aggregation on velocity profiles in venules, Am. J. Physiol.-Heart Circ. Physiol. 280 (1) (2001) H222–H236.
[4] J. Panés, M. Perry, D.N. Granger, Leukocyte-endothelial cell adhesion: avenues for therapeutic intervention, Br. J. Pharmacol. 126 (3) (1999) 537.
[5] C.S. Peskin, The immersed boundary method, Acta Numer. 11 (2002) 479–517.
[6] I.V. Pivkin, G.E. Karniadakis, Accurate coarse-grained modeling of red blood cells, Phys. Rev. Lett. 101 (11) (2008) 118105.
[7] T. Krüger, D. Holmes, P.V. Coveney, Deformability-based red blood cell separation in deterministic lateral displacement devices—A simulation study, Biomicrofluidics 8 (5) (2014) 054114.
[8] D.A. Fedosov, H. Noguchi, G. Gompper, Multiscale modeling of blood flow: from single cells to blood rheology, Biomech. Model. Mechanobiol. 13 (2) (2014) 239–258.
[9] P. Bagchi, Mesoscale simulation of blood flow in small vessels, Biophys. J. 92 (6) (2007) 1858–1877.
[10] M. Pepona, P. Balogh, D.F. Puleri, W.F. Hynes, C. Robertson, K. Dubbin, J. Alvarado, M.L. Moya, A. Randles, Investigating the interaction between circulating tumor cells and local hydrodynamics via experiment and simulations, Cell. Mol. Bioeng. 13 (5) (2020) 527–540.

[11] H. Lei, G.E. Karniadakis, Quantifying the rheological and hemodynamic characteristics of sickle cell anemia, Biophys. J. 102 (2) (2012) 185–194.
[12] B. Czaja, M. Gutierrez, G. Závodszky, D. de Kanter, A. Hoekstra, O. Eniola-Adefeso, The influence of red blood cell deformability on hematocrit profiles and platelet margination, PLoS Comput. Biol. 16 (3) (2020) e1007716.
[13] T. Krüger, D. Holmes, P.V. Coveney, Deformability-based red blood cell separation in deterministic lateral displacement devices—A simulation study, Biomicrofluidics 8 (5) (2014) 054114.
[14] P. Balogh, P. Bagchi, Analysis of red blood cell partitioning at bifurcations in simulated microvascular networks, Phys. Fluids 30 (5) (2018) 051902.
[15] J. Yang, S.S. Yoo, T.-R. Lee, Effect of fractional blood flow on plasma skimming in the microvasculature, Phys. Rev. E 95 (4) (2017) 040401.
[16] J. Zhang, P.C. Johnson, A.S. Popel, Red blood cell aggregation and dissociation in shear flows simulated by lattice Boltzmann method, J. Biomech. 41 (1) (2008) 47–55.
[17] D.A. Fedosov, B. Caswell, A.S. Popel, G.E. Karniadakis, Blood flow and cell-free layer in microvessels, Microcirculation 17 (8) (2010) 615–628.
[18] D. Katanov, G. Gompper, D.A. Fedosov, Microvascular blood flow resistance: role of red blood cell migration and dispersion, Microvasc. Res. 99 (2015) 57–66.
[19] K. Vahidkhah, S.L. Diamond, P. Bagchi, Platelet dynamics in three-dimensional simulation of whole blood, Biophys. J. 106 (11) (2014) 2529–2540.
[20] S. Fitzgibbon, A.P. Spann, Q.M. Qi, E.S. Shaqfeh, In vitro measurement of particle margination in the microchannel flow: effect of varying hematocrit, Biophys. J. 108 (10) (2015) 2601–2608.
[21] A. Jain, L.L. Munn, Determinants of leukocyte margination in rectangular microchannels, PLoS One 4 (9) (2009) e7104.
[22] J.B. Freund, Leukocyte margination in a model microvessel, Phys. Fluids 19 (2) (2007) 023301.
[23] S. Roychowdhury, J. Gounley, A. Randles, Evaluating the influence of hemorheological parameters on circulating tumor cell trajectory and simulation time, in: Proceedings of the Platform for Advanced Scientific Computing Conference, 2020, pp. 1–10.
[24] L. Xiao, C. Lin, S. Chen, Y. Liu, B. Fu, W. Yan, Effects of red blood cell aggregation on the blood flow in a symmetrical stenosed microvessel, Biomech. Model. Mechanobiol. 19 (1) (2020) 159–171.
[25] J. Gounley, E.W. Draeger, A. Randles, Numerical simulation of a compound capsule in a constricted microchannel, Procedia Comput. Sci. 108 (2017) 175–184.
[26] P. Balogh, J. Gounley, S. Roychowdhury, A. Randles, A data-driven approach to modeling cancer cell mechanics during microcirculatory transport, Sci. Rep. 11 (1) (2021) 1–18.
[27] L. Grinberg, J.A. Insley, V. Morozov, M.E. Papka, G.E. Karniadakis, D. Fedosov, K. Kumaran, A new computational paradigm in multiscale simulations: Application to brain blood flow, in: Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis, 2011, pp. 1–5.
[28] J. Ames, D.F. Puleri, P. Balogh, J. Gounley, E.W. Draeger, A. Randles, Multi-GPU immersed boundary method hemodynamics simulations, J. Comput. Sci. 44 (2020) 101–153.
[29] L. Lu, M.J. Morse, A. Rahimian, G. Stadler, D. Zorin, Scalable simulation of realistic volume fraction red blood cell flows through vascular networks, in: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2019, pp. 1–30.
[30] J. Gounley, E.W. Draeger, A. Randles, Immersed boundary method halo exchange in a hemodynamics application, in: International Conference on Computational Science, Springer, 2019, pp. 441–455.
[31] L. Xiao, Y. Liu, S. Chen, B. Fu, Effects of flowing RBCs on adhesion of a circulating tumor cell in microvessels, Biomech. Model. Mechanobiol. 16 (2) (2017) 597–610.
[32] L. Stella, S. Melchionna, Equilibration and sampling in molecular dynamics simulations of biomolecules, J. Chem. Phys. 109 (23) (1998) 10115–10117.
[33] K. Gordiz, D.J. Singh, A. Henry, Ensemble averaging vs. time averaging in molecular dynamics simulations of thermal conductivity, J. Appl. Phys. 117 (4) (2015) 045104.
[34] S. Roychowdhury, E.W. Draeger, A. Randles, Establishing metrics to quantify underlying structure in vascular red blood cell distributions, in: International Conference on Computational Science, Springer, 2022, pp. 89–102.
[35] A. Donev, I. Cisse, D. Sachs, E.A. Variano, F.H. Stillinger, R. Connelly, S. Torquato, P.M. Chaikin, Improving the density of jammed disordered packings using ellipsoids, Science 303 (5660) (2004) 990–993.
[36] H. Malmir, M. Sahimi, M. Tabar, Microstructural characterization of random packings of cubic particles, Sci. Rep. 6 (1) (2016) 1–9.
[37] K. Miśkiewicz, R. Banasiak, M. Niedostatkiewicz, K. Grudzień, L. Babout, An algorithm to generate high dense packing of particles with various shapes, in: MATEC Web of Conferences, Vol. 219, EDP Sciences, 2018, p. 05004.
[38] S. P.R., Red cell indices, in: H.K. Walker, W.D. Hall, J.W. Hurst (Eds.), Clinical Methods: The History, Physical, and Laboratory Examinations, third ed., Butterworths, 1990, Ch. 152.
[39] S.N. Doost, D. Ghista, B. Su, L. Zhong, Y.S. Morsi, Heart blood flow simulation: a perspective review, Biomed. Eng. Online 15 (1) (2016) 1–28.
[40] C. Desjardins, B.R. Duling, Microvessel hematocrit: measurement and implications for capillary oxygen transport, Am. J. Physiol.-Heart Circ. Physiol. 252 (3) (1987) H494–H503.
[41] E.G. Birgin, R.D. Lobato, A matheuristic approach with nonlinear subproblems for large-scale packing of ellipsoids, European J. Oper. Res. 272 (2) (2019) 447–464.
[42] M.P. Allen, D.J. Tildesley, Radial distribution function in a planar interface, in: Computer Simulation of Liquids, Oxford University Press, 2017, pp. 453–470, Ch. 14.
[43] V. Yeghiazaryan, I.D. Voiculescu, Family of boundary overlap metrics for the evaluation of medical image segmentation, J. Med. Imaging 5 (1) (2018) 015006.
[44] A. Randles, L.V. Kale, J. Hammond, W.D. Gropp, E. Kaxiras, Performance analysis of the lattice Boltzmann model beyond Navier-Stokes, 2013, pp. 1063–1074.
[45] A. Randles, E.W. Draeger, P.E. Bailey, Massively Parallel Simulations of Hemodynamics in the Primary Large Arteries of the Human Vasculature, J. Comput. Sci. 9 (2015) 70–75.

Sayan Roychowdhury is a Biomedical Engineering Ph.D. candidate at Duke University in the Randles Lab. His work leverages the power of supercomputers to perform microscale bloow flow and microfluidic device simulations, with a focus on fluid–structure interaction models to capture cellular dynamics. He received his undergraduate degree in Electrical Engineering and Computer Science from the University of Illinois at UrbanaChampaign. Previously, he worked as a graduate research assistant at Lawrence Berkeley National Laboratory in the computational materials group, developing predictive models for inorganic solid crystal materials.

Erik W. Draeger is the Director of the High Performance Computing Innovation Center and RADIUSS project at Lawrence Livermore National Laboratory as well as the Scientific Computing group leader at the Center for Applied Scientific Computing. He is also the Deputy Director of Application Development for the Exascale Computing Project, jointly overseeing a portfolio of 22 Office of Science applications, 4 NNSA applications, and 7 co-design projects. Erik earned a Bachelor's degree in Physics from the University of California, Berkeley in 1995 and received a Ph.D. in theoretical condensed matter physics from the University of Illinois, UrbanaChampaign in 2001. He has over a decade of experience developing scientific applications to achieve maximum scalability and time to solution on next-generation architectures. He has been a finalist for the Gordon Bell Prize six times since 2005 and won the prize in 2006.

Amanda Randles is the Alfred Winborne Mordecai and Victoria Stover Mordecai Assistant Professor of Biomedical Sciences and Biomedical Engineering at Duke University. She has courtesy appointments in the departments of Mechanical Engineering and Material Science, Computer Science and Mathematics, and is a member of the Duke Cancer Institute. Focusing on the intersection of high performance computing, machine learning, and personalized modeling, her group is developing new methods to aid in the diagnosis and treatment of a diseases ranges from cardiovascular disease to cancer. Amongst other recognitions, she has received the NIH Pioneer Award, the NSF CAREER Award, and the ACM Grace Hopper Award. She was named to the World Economic Forum Young Scientist List and the MIT Technology Review World's Top 35 Innovators under the Age of 35 list and is a Fellow of the National Academy of Inventors. Amanda received her Ph.D. in Applied Physics from Harvard University as a DOE Computational Graduate Fellow and NSF Fellow. Before that, she received her Master's degree in Computer Science from Harvard University and her Bachelor's degree in Computer Science and Physics from Duke University. Prior to graduate school, she worked as a software engineer at IBM on the Blue Gene supercomputing team. She has contributed to over 80 peer-reviewed papers, over 100 granted US patents, and had over 100 pending patent applications.