

Supplementary Information for *Size-based sorting of dynamic bacterial clusters*

Elham Akbari¹, Jason P. Beech¹, Johannes Kumra Ahnlide², Sebastian Wrighton², Pontus Nordenfelt², and Jonas O. Tegenfeldt^{*1}

¹Division of Solid State Physics, Lund University, Sweden

²Department of Clinical Sciences, Division of Infection Medicine, Faculty of Medicine, Lund University, Lund, Sweden

CONTENTS

1	COMSOL simulations	2
2	Device fabrication and characterization	2
3	Experimental setup	4
4	Microscopy, image analysis and data processing	5
A	Image segmentation pipeline	5
B	Conventions used for the plots	7
C	Definitions of purity and yield	8
5	Dynamics of clusters in bulk	8
6	Cluster deformation in the DLD array	11

1. COMSOL SIMULATIONS

A COMSOL simulation was conducted on the inlet and outlet channels to ensure balanced resistance, which guarantees laminar, stable flow within the DLD device. Fig. S1 shows the flow streamlines in the inlets and outlets, as well as in the observation area. The observation array was added to the design to allow imaging of sorted particle fractions prior to sample collection at outlets. The factor 10 drop in flow velocity in the observation area (see simulations in figure S2) facilitate imaging of flowing particles while relative lateral positions, generated in DLD array, remain unchanged (see flow lines in figure S2).

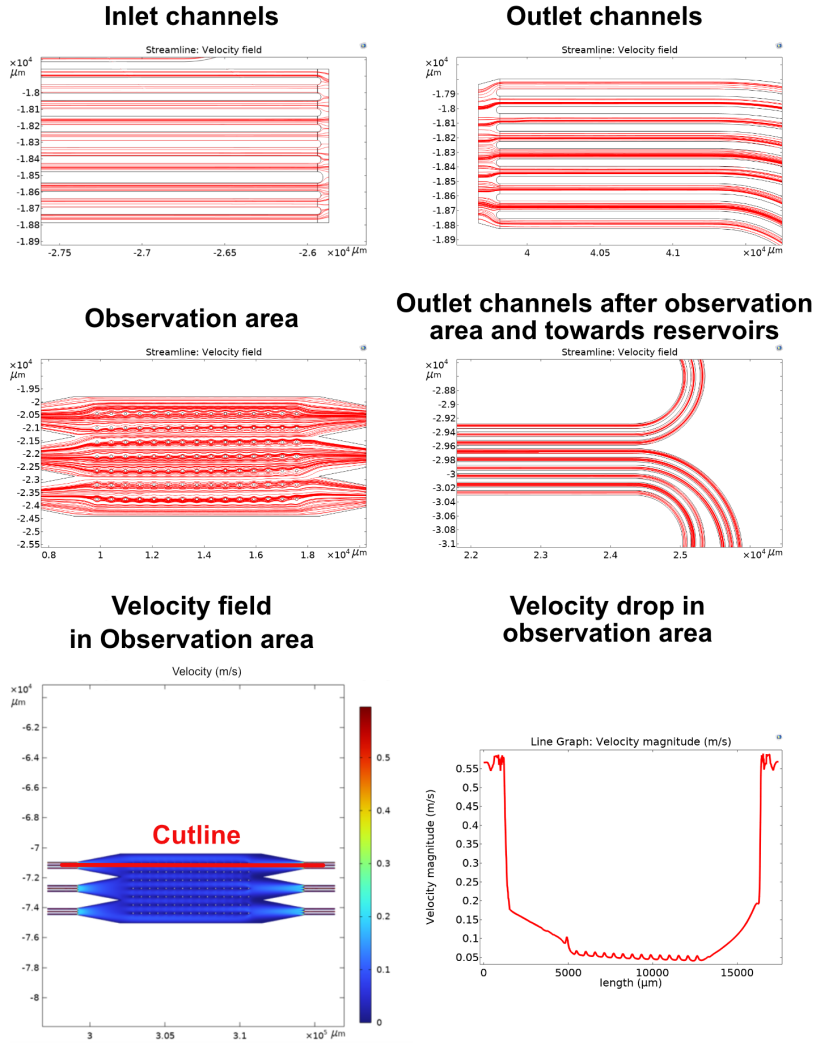


Fig. S1. Simulation of flow streamlines in inlet and outlet channels as well as inside the observation area. The streamlines are shown as red lines. The velocity drops within the observation also shown but relative lateral positions remain unchanged.

2. DEVICE FABRICATION AND CHARACTERIZATION

A master mould was fabricated in $100\mu\text{m}$ thick dry film resist (SUEx®, K100, DJ Microlaminates, Sudbury, Massachusetts, USA) laminated onto a 4" silicon wafer using a laminating machine (Catena 35, Acco UK Ltd, Buckinghamshire, UK) at 65°C and a pre-exposure bake of 5 minutes at 85°C on a hotplate (Model 1000-1 Precision Hot Plate, Electronic Micro Systems Ltd, West

Middlelands, UK) was performed to remove air bubbles and to help the film relax.

The channel design was transferred onto the mold through a mask (Delta Mask B.V., Enschede, The Netherlands) at 365 nm in a contact mask aligner (Karl Suss MJB4 soft UV, Munich, Germany) for 34 s at a lamp power of 30 mW/cm². Post exposure baking was done at 85°C for 5 minutes. Exposed SU8 was developed in mr-DEV 600 (Micro Resist Technology GmbH, Berlin, Germany) for 15 minutes plus 5 minutes in fresh developer followed by a rinse in flowing IPA (iso-propoanol alcohol) and drying with nitrogen to remove unpolymerized SU8. A final bake was done in a convection oven at 200°C for 2 hours. To prevent PDMS adhesion to the mold surface during casting a layer of aluminium oxide (1 nm) followed by a monolayer of Perfluorodecyltrichlorosilane (FDTs) was applied using an ALD system (Fiji – Plasma Enhanced ALD, Veeco, NY, USA).

Using this master, standard replica molding as developed by Xia et al[1] was used to cast PDMS devices (PDMS, Sylgard 184, Dow Corning, Midland, MI, USA). Cast PDMS pieces were removed from the master and access holes were punched for inlet and outlet connections. After activation by oxygen plasma (Zepto, Diener electronic GmbH & Co. KG, Ebhausen, Germany) (PDMS 15 s and glass slide 50 s) closed channels are created by bringing PDMS and glass into contact. Immediately after the bonding, the device was filled with poly(acrylamide)-g-(PMOXA, 1,6-hexanediamine, 3-aminopropyl dimethylsilanol) (abbreviated as PAcrAm-g-(PMOXA, NH₂, Si)) (SuSoS AG, Duebendorf, Switzerland). This molecule has charged amine groups that facilitate electrostatic absorption onto the surface, plus dimethylsilanols which covalently bind to hydroxylated surfaces. The resulting PAcrAm-g-(PMOXA, NH₂, Si) product is optimized for covalent binding onto glass (SiO₂) substrates but also adheres to other surfaces containing hydroxyl (-OH) groups, such as PDMS and other plasma-activated polymers[2]. The Poly(2-methyl-2-oxazoline) (PMOXA) group has molecular properties similar to poly(ethylene glycol) (PEG) and, like PEG, prevents surface protein fouling. Note that for clean room work, adherence to standard safety routines and precautions including wearing personal protective equipment is a requirement. This includes using protective gloves, goggles and face shields as appropriate. For volatile compounds, ventilated work space (fume hood or equivalent) must be utilized. Routines must be in place to ensure that users are properly trained accordingly.

The walls of the PDMS DLD device were imaged using optical microscopy. The images in Fig. S2 show the entire side wall of the device, the side wall of the DLD posts, and the x-y cross-sections of the posts. The schematic figure illustrates the directions of the x, y, and z axes and the dimensions of the DLD device that the images present.

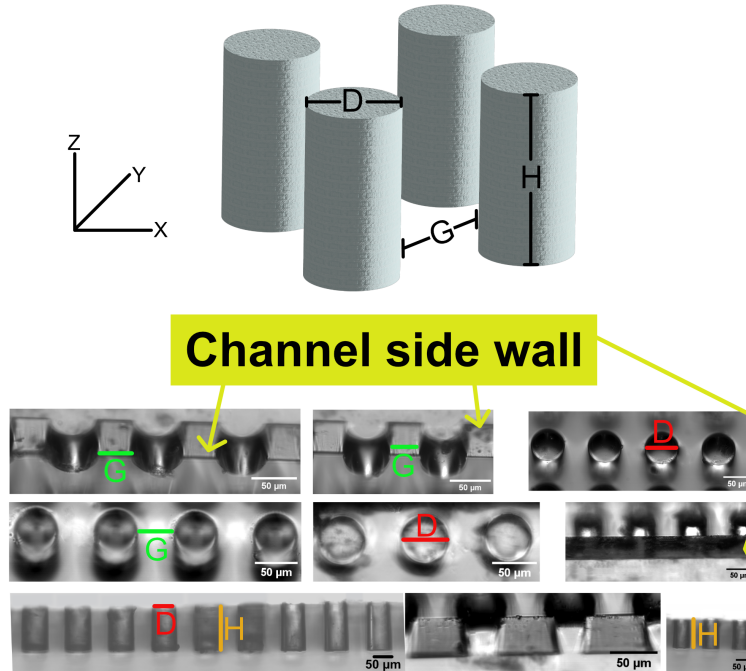


Fig. S2. Optical microscopic images of the DLD device walls. The images show the device side walls as well as posts' side walls.

We measured flow resistance in our DLD device using two methods. Firstly, using an inline flowmeter (Fluigent, Paris, France) volumetric flow rate was measured as a function of applied pressure in the range 0 - 50 mbar (the maximum value possible for this particular flowmeter) resistance was calculated as the inverse of the slope of a linear fit to the data, see Fig. S3). This yielded a value of 4.2 mbar·min/ μ l (after measuring and removing the resistances of the capillary tubes). Secondly, for higher pressures in the range 50 - 300 mbar, for given pressure and time, sample volumes were collected and measured (by weighing) and the resulting resistance calculated. Results (also with resistances of capillaries removed) are shown in Table S1. From this we conclude that resistance in our device does not change for pressures in the range 0 to 300 mbar indicating that volume changes due to deformations of the PDMS can be neglected.

Table S1. Lists of Input pressures applied to the DLD device, their corresponding volumetric flowrates, and calculated resistances, and the average of resistances calculated at different pressures.

Pressure (mbar)	Volume (μ l/min)	Resistance (mbar.min/ μ l)	Average device resistance
50 ± 1	10 ± 0.35	4.36 ± 0.48	4.41 ± 0.54
100 ± 1	23.1 ± 0.2	4.29 ± 0.21	
150 ± 1	34.8 ± 1.7	4.30 ± 0.03	
300 ± 1	69 ± 7.1	4.67 ± 0.15	

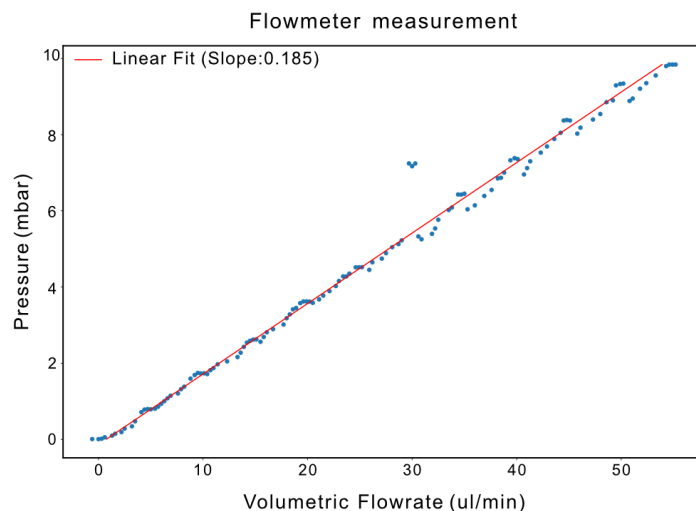


Fig. S3. In a plot of the applied pressure versus the resulting volumetric flow rate, we obtain the flow resistance from the slope of the curve. Data were collected for a setup with a capillary tube connected to the DLD device using Fluigent flow meter. Outliers were excluded for the linear fit.

3. EXPERIMENTAL SETUP

An image of the setup is shown in Fig. S4

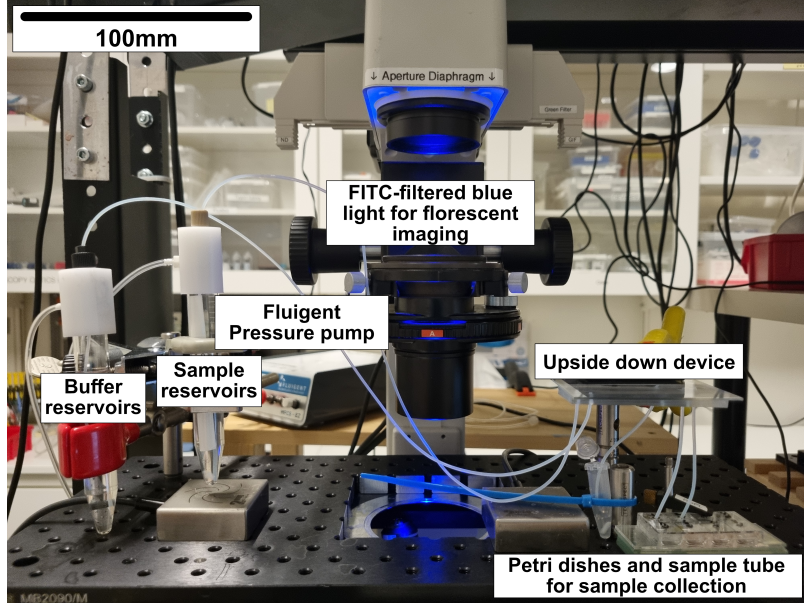


Fig. S4. Image of the experimental setup including reservoirs for sample and buffer, connecting capillary tubes, Fluigent pressure controller, fluorescence microscope, and tubes and petri dishes for sorted subpopulation collection.

4. MICROSCOPY, IMAGE ANALYSIS AND DATA PROCESSING

Objectives used in the study included 4X (Nikon Plan Apo λ , NA 0.2, Field of View (FoV) of 25mm), 10X (Nikon Plan Apo λ , NA 0.45, FoV of 819 μm), 20X (Nikon Plan Fluor, NA 0.45, FoV of 320 μm), and 50X (LU Plan, NA 0.55, FoV of 250 μm).

In the separation experiment, an automated image processing pipeline was used for the analysis of batches of microscopic images taken from the outlets and inlet reservoirs. An automated pipeline was also used for the analysis of images taken for the study of cluster dynamics. In addition to segmentation and object analysis, this pipeline includes division of the images by the fluorescent lamp intensity and subtraction of the background noise intensity, applying correction factors to remove the leakage of one color channel to another channel and detection of overlapping objects. Python Spyder (Anaconda 4) packages like numpy, matplotlib, skimage (more specifically skimage.measure including Regionprops), cv2, scipy, and pandas were used to develop the pipeline. Ilastik, which is a user-interface deep learning computer vision software, was used for the segmentation of images featuring low contrast between foreground and background or images with a high density of detectable objects[3].

Properties such as major axis length, minor axis length, area, position (Center X and Center Y), and circularity were extracted from the image analysis using Regionprops in Python cv2 skimage package. The major axis length is the length (in μm) of the major axis of the ellipse that has the same normalized second central moments as the region, returned as a scalar, while the minor axis length is the length (in μm) of the minor axis of the ellipse that has the same normalized second central moments as the region, returned as a scalar. The area is measured directly by Regionprops by counting the number of pixels within the segmented area of each object. Circularity is then calculated as the ratio of the area measured by Regionprops to the area of a perfect circle with a diameter equal to the major axis length measured by Regionprops.

The definitions of the extracted properties are shown in Fig. S5. Major and minor axis lengths are equal to the longest and shortest axis of an ellipse fitted to the segmented object. Area is the number of pixels within the segmented object multiplied by the area of a single pixel. Circularity is defined as the ratio between the segmented area to the area of a circle with a diameter equal to the major axis length.

A. Image segmentation pipeline

The separation outcome is evaluated by measuring the properties of the images of bacteria and clusters from the inlet and outlet samples. The images are fluorescent, and the analysis was

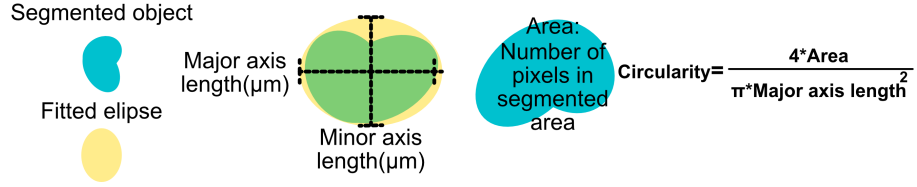


Fig. S5. Schematic illustration of the extracted properties from the segmented objects. Major axis length is the longest axis of the ellipse fitted to the segmented objects, minor axis length is the shortest axis of the ellipse. Area is equal to the number of pixels within the segmented area multiplied by the area of a single pixel. Circularity is defined as the ratio between the segmented area and the area of a circle with the diameter equal to the major axis length.

done using Python scripts and Ilastik software to segment the images. The analysis includes the following steps:

1. Import the image into a Python script and convert it to an 8-bit grayscale image.
2. Plot the intensity profile of the image, calculate the signal-to-noise ratio, and generate the histogram of the intensity profile for the input image.
3. Check if the intensity profile shows a distinct peak with a low level of background signal. This can be seen in Fig. S6-A-1. Fig. S6-A-2 shows peaks for background signals, indicating the existence of small objects that are not detectable in the image at a specific range of brightness and contrast levels.

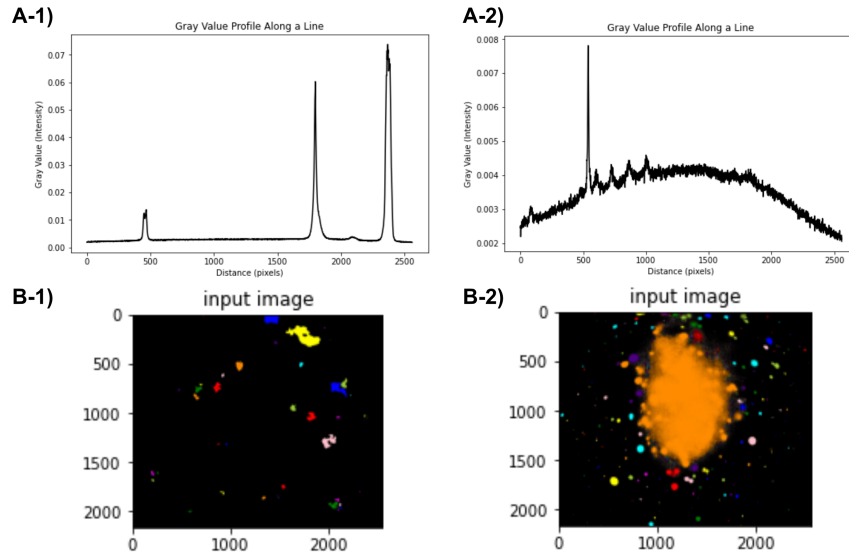


Fig. S6. A-1) Example of an intensity profile plot where the background does not show any distinct peaks, and the foreground generates sharp peaks. A-2) Example of an intensity profile plot where the background generates distinct signals. B-1) Example of successful segmentation where neither any large objects nor a lot of small objects (noise) were detected. B-2) Example of poor segmentation where a very large area was detected, indicating that the Otsu thresholding pipeline did not work well for this image.

To ensure that no small objects are missed in the segmentation process, the data is visualized using Fiji at different contrast setting . Fig. S7 shows two examples of images checked in Fiji at different contrast settings. It can be seen that for the example image from the displacement outlet, changing the contrast level does not reveal any small objects. On the other hand, the example

of the zigzag image shows that changing the brightness level leads to the appearance of some particles, indicating the presence of small objects and background noise that could be missed if the Otsu thresholding method is used.

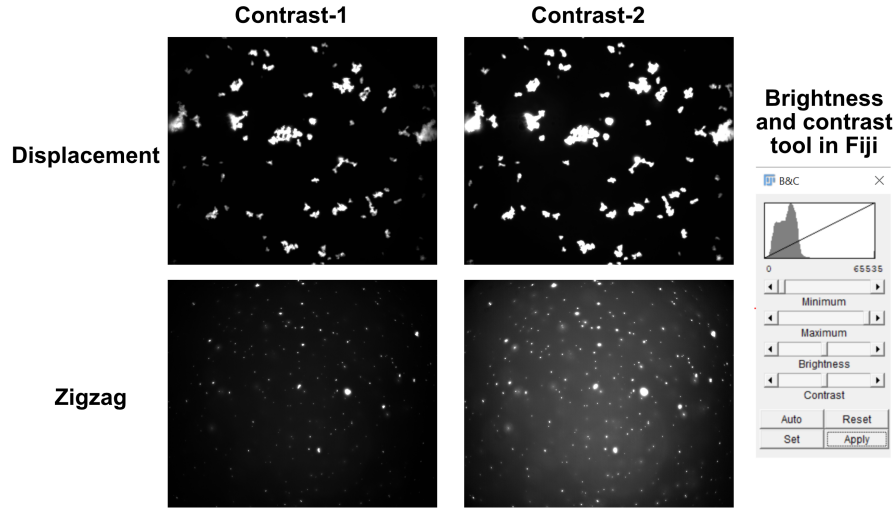


Fig. S7. Example of images from the displacement outlet and the zigzag outlet at different contrast level adjusted by brightness and contrast tool in Fiji.

4. (a) Segment the image using OTSU thresholding.
(b) Segment the images using Ilastik.
5. Analysing the segmented images by OTSU or Ilastik and extract the objects within the image using Regionprops.
6. (a) Check the range of the major axis length of the detected objects. If the major axis length is smaller than 1.8 or larger than 500, the image is not suitable for the pipeline, and manual segmentation should be applied again using Ilastik. The image will then be analyzed again from step 5. Fig. S6-B-1 shows an example of a segmented image with extracted object properties, indicating proper segmentation. Fig. S6-B-2 shows an example of a segmented image where the segmentation is not proper, and a very large object was detected, indicating that the image is not suitable for the segmentation pipeline.
(b) If the major axis length is in the range, the data can be saved as Pandas dataframe in .csv format.

Python script used for image analysis can be found in Github[4].

B. Conventions used for the plots

For our boxplots, we follow the convention that the bottom edge of the box presents 25th percentile (Q1) while the top edge presents 75th percentile (Q3). The height of the box is equal to Interquartile Range ($IQR = Q3 - Q1$) which shows the middle 50% of the data. The line inside the box shows the median and the whiskers extend to the smallest-largest observed values within $1.5 \times IQR$ of Q1 and Q3. The dots show the outlying values. This definition also applies to the following figures.

For the statistical analysis of the data, the histograms are normalized so that the sum of the bin heights is one corresponding to a probability mass function for a discrete variable where the sum of probabilities for all the values equals one.

Histograms in the plots are normalized so that they represent a probability density. The bin heights of the probability mass distribution is simply divided by the width of the bin. For example, for a histogram over the size distribution expressed in μm , the unit of the bin height is thus μm^{-1} .

In this way the shape of the histogram is not affected by changes in the bin width. This convention is used throughout the paper to present the results.

The meandering index is defined as the ratio of the net displacement to the total path length of the migration trajectory. Net displacement is the straight-line distance between the starting point and the endpoint of a cell's trajectory, while the total path length is the actual distance the cell travels along its path. The speed of a neutrophil cell is calculated as the total path length traveled by the cell per unit time.

C. Definitions of purity and yield

We define purity of a subpopulation in a reservoir as the number of particles in the desired size range divided the total number of particles in that reservoir.

The purity of the Small subpopulation is defined by the fraction of the particles in the Small outlet having a diameter less than the critical diameter.

$$purity(S) = P(d < D_C|S) \quad (S1)$$

The purity of the Large subpopulation is defined by the fraction of the particles in the Large outlet having a diameter greater than the critical diameter.

$$purity(L) = P(d > D_C|L) \quad (S2)$$

The equations for the purities, Eqs. S1-S2 correspond to the sums of the probability distributions in respective outlet for particles less than or greater than D_C . See Figure 3 in the main text.

The yield of a subpopulation in an outlet reservoir is defined as the fraction of particles in the inlet in the desired size range that are sent to the desired outlet.

The yield of the Small subpopulation is defined by the fraction of the particles in the inlet with a diameter less than the critical diameter that are sent to the Small outlet.

$$yield(S) = P(S|d < D_C) \quad (S3)$$

The yield of the Large subpopulation is defined by the fraction of the particles in the inlet with a diameter greater than the critical diameter that are sent to the Large outlet.

$$yield(L) = P(L|d > D_C) \quad (S4)$$

The equations for the yields, Eqs. S3-S4 correspond to the sums of the routing probabilities for particles less than or greater than D_C . See Figure 3 in the main text.

Due to the difficulties of estimating the fraction of sample that enters the device, we report the yield as a nominal yield assuming that the total from all outlets represent the total from the inlet. The reported yield therefore does not give information on any selective losses of the sample from the inlet reservoir to the outlet reservoir.

5. DYNAMICS OF CLUSTERS IN BULK

The aim of this experiment was to examine the effects of the buffer medium (the environment in which the bacteria reside) and time on the formation of clusters of group A streptococcus. This experiment took place in culture bottles and serves as the control for the GAS cluster dynamics study, the results of which are reported and discussed in the main text.

To achieve this, heat-killed bacterial samples usually used for the separation experiment were stained with two different fluorophores and were mixed and imaged over time. The number of clusters containing both colors was counted as an indication of bacterial adhesion and cluster formation. The results are presented by analyzing the ratio of clusters containing both colors to the total number of clusters.

One batch of bacteria was stained with Oregon Green (OG), which has an excitation wavelength of 488 nm, while another batch was stained with Wheat Germ Agglutinin (WGA) conjugates, with an excitation wavelength of 640 nm. One batch was diluted in PBS, and another in 1% BSA.

The batches were diluted using the same dilution factor as in the main experiment (100 μ l of bacteria in 4 mL of buffer). The two batches (one for each buffer) were then mixed and divided into separate bottles, each maintained under specific conditions. All mixed batches were stored at room temperature under two conditions: one set was constantly shaken (CS) at a speed of 20 rpm, while the other remained stationary (S). The batches were imaged with both two excitation wavelength each corresponds to one dye at specific time intervals shown in the table S2.

The experimental parameters are summarized in Table S2.

Table S2. Parameters of the experiment

Buffer	PBS, BSA 1%
Fluorophore dye	Oregon green (OG), CF-WG 633 (CF)
Time intervals of imaging	0h, 4h, 8h, 12h, 24h, 32h
Samples keeping condition	Stationary (S), Constant shaking (CS)

The analysis of the images was performed in several steps. These steps included normalizing the images, subtracting background noise, splitting RGB images into separate grayscale images, extracting information about the clusters in the images, and identifying the overlap area between the two channels, which corresponds to the detected clusters containing both colors. These clusters with both signals are the indication of bacterial adhesion and cluster formation.

As mentioned earlier, two fluorophores and two filters were used to image the samples. The signals resulting from the leakage of one wavelength into the other filter were quantified and removed to prevent erroneous results. This was achieved by imaging each fluorophore individually with both wavelengths and measuring the signal intensities in both channels. The code of data analysis can be found in Github[5].

Fig. S8 shows the flowchart explaining the image analysis pipeline.

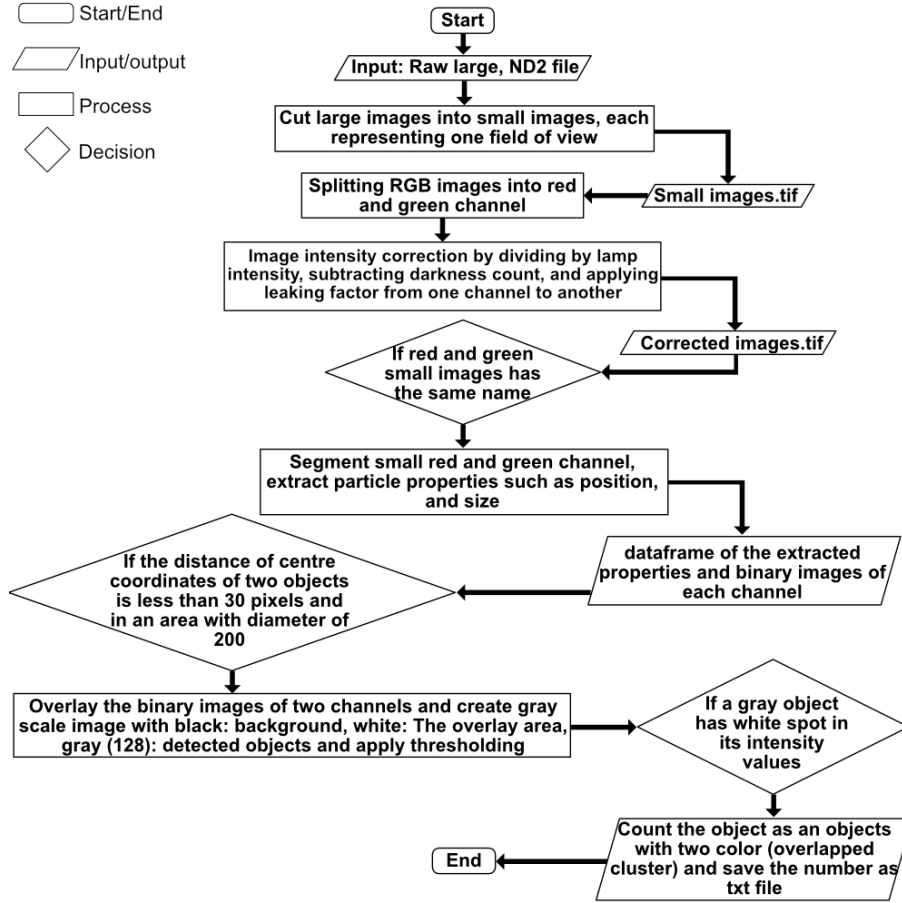


Fig. S8. Flowchart of the data analysis pipeline. The input data are RGB images with .ND2 format (Nikon NIS-Elements ND2 Image). And the output is the number of all detected objects and the number of detected objects showing signals in both excitation wavelengths (merged clusters)

The formation of clusters over time is investigated by calculating the ratio of clusters with mixed colors to the total number of clusters and displaying these ratios over time in bar charts for both buffer types and keeping conditions. Fig. S9 shows these bar charts, and as explained earlier, "S" is the abbreviation for "stationary," and "CS" stands for "constant shaking," both of which indicate the conditions under which the samples were kept throughout the experiment. The samples were kept at room temperature and each has the concentration of 100 μ L bacteria sample in 4mL of the buffer solution. The names of the buffers are also included in the title of each bar chart. For example, "Serum – CS" means the sample was in a 1% BSA buffer medium and kept under constant shaking conditions. Each sample was imaged three times and the error bars in each plot shows the variation between the extracted values from each image.

The overlapped clusters constitute at most 10% of the total detected objects across all samples tested over 32 hours. This suggests the stability of the sample in terms of size distribution over the time scale of this experiment. The same value ranges were also observed in the experiment where the mixed samples were separated by the device. The data of that experiment are shown and discussed in the main text.

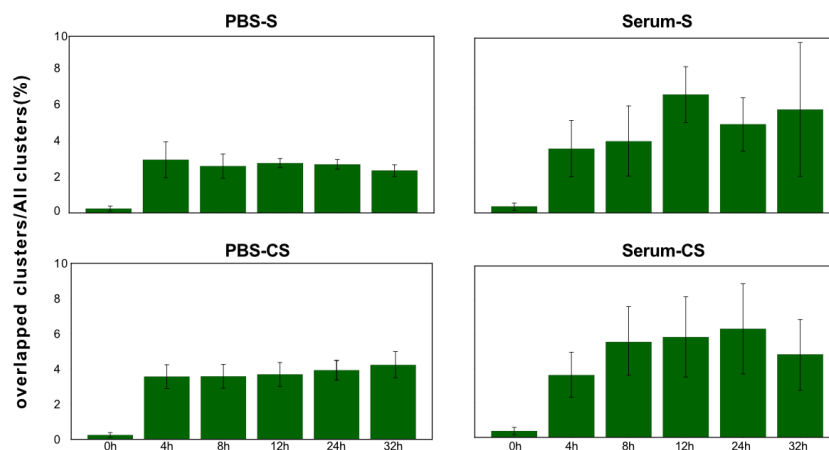


Fig. S9. Bar charts showing the mean ratio of merged clusters to total clusters and bacteria, with standard deviation.

6. CLUSTER DEFORMATION IN THE DLD ARRAY

Another dynamic change that the clusters underwent was deformation. Although the results of this observation are not within the scope of this paper, an example of deformation observed for bacterial clusters is shown in [S10](#). Figure [S10A](#) shows frames from fast-imaging videos, where the clusters are highlighted with red ellipses. Figure [S10B](#) shows the maximum projection of the segmented videos corresponding to the frames in Figure [S10A](#), where the bacterial clusters appear as white spots. The videos were recorded at 1000 frames/second, and the time interval between two frames, as well as between two consecutive steps in the binary images, is 2 ms.

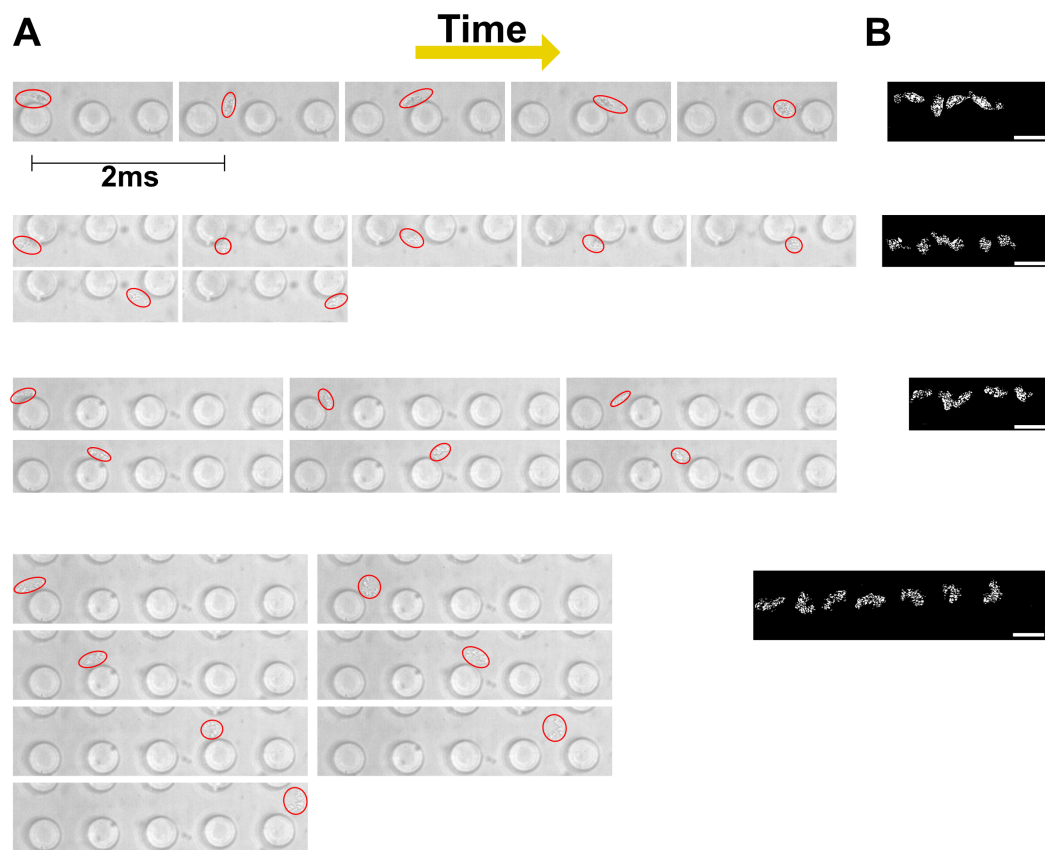


Fig. S10. Example image of bacterial clusters being deformed while sorted by the DLD posts. (A) Frames from high-speed videos (1,000 frames per second) tracking the movement of the clusters. The frame sequence progresses from left to right and top to bottom, with a 2 ms interval between consecutive frames. Bacteria in each frame are marked with red ellipses. (B) Binary images of the z-stacked maximum intensity projections of the videos corresponding to the frames shown in (A). The binary images highlight only the bacterial clusters and illustrate their shape changes during sorting. Scale bar= 50 μ m

REFERENCES

1. Y. Xia, J. J. McClelland, R. Gupta, *et al.*, "Replica molding using polymeric materials: A practical step toward nanomanufacturing," *Adv. Mater.* **9**, 147–149 (1997).
2. S. Weydert, S. Zurcher, S. Tanner, *et al.*, "Easy to apply polyoxazoline-based coating for precise and long-term control of neural patterns," *Langmuir* **33**, 8594–8605 (2017).
3. S. Berg, D. Kutra, T. Kroeger, *et al.*, "Ilastik: interactive machine learning for (bio) image analysis," *Nat. methods* **16**, 1226–1232 (2019).
4. E. Akbari, "Otsu thresholding code," <https://github.com/Elhamakbr/Elhamakbr/blob/8807b27951a318139d136eaca7d7408e706a8d58/Image%20analysis%20using%20OTSU%20thresholding>. Accessed: 2025-09-30.
5. E. Akbari, "GAS cluster dynamics in DLD device and batch," <https://github.com/Elhamakbr/Elhamakbr/blob/main/Image%20processing%3A%20Dynamics%20of%20clusters%20fromation>. Accessed: 2025-09-30.