Supporting Information

for "Fast, Label-Free Tracking of Single Viruses and Weakly Scattering Nanoparticles in a Nano-Fluidic Optical Fiber

Sanli Faez, Yoav Lahini, Stefan Weidlich, Rees F. Garmann, Katrin Wondraczek, Matthias Zeisberger, Markus A. Schmidt, Michel Orrit, Vinothan N. Manoharan

Additional details on methods

A. *Calibration of background scattering*

The background scattering from the empty core can be used as a reference for the intensity of guided light inside the fiber. It also indicates the scattering losses of the fiber. We first measure the instrument response function by measuring the background scattering from a piece of commercial single mode fiber (HB 450) with known attenuation (primarily due to Rayleigh scattering) of 10 dB/km at 633 nm in the same configuration used for our nano-fluidic fibers. This attenuation corresponds to a loss factor of 2.3×10^{-9} μm^{-1}. We consider the attenuation value for 670 nm to be the same. At this wavelength, we recorded 1550 photo-electrons per millisecond per micrometer for each milliwatt of power coupled into this fiber on the camera. For this measurement, the light intensity that is coupled into the single mode of the fiber was roughly 1 mW and was accurately measured at the exit of the fiber. After determining the instrument response function, we repeat the same measurement with our opto-fluidic fibers in exactly the same configuration. The images for the two fibers are depicted in Fig. S1a,b, and the sum of all pixels perpendicular to the fiber axis is plotted as a function of axial position along the fiber in Fig. S1c. From these measurements, the loss factor for our fiber is shown to be 50 dB/km, which is roughly five times the loss factor in the commercial single mode fibers. This extra loss is likely due to nanometric ripples on the inner glass interface, which is a common outcome of the fiber drawing process.
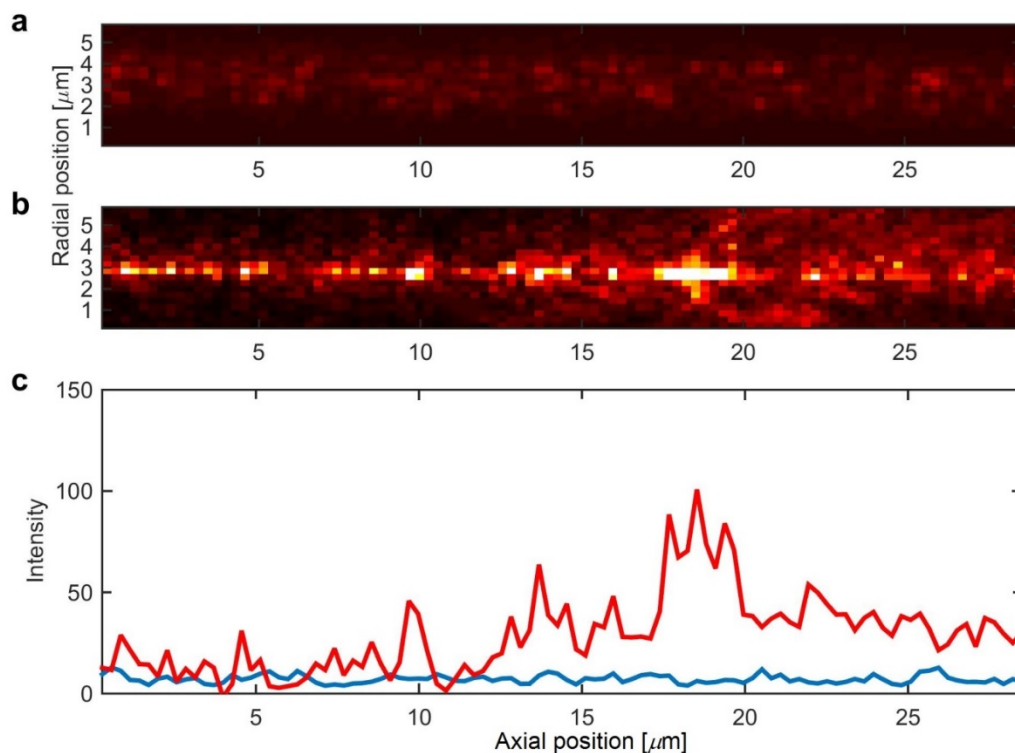
Figure S1: (a) Background scattering from the impurities in the core of a commercial single mode fiber. (b) As in (a), but for an empty nano-channel fiber. (c) Plot of the total scattered intensity in (a) and (b), showing that the nano-channel fiber has an approximately 5-fold higher background than the step-index single-mode fiber. The higher background of the nano-channel fiber is likely caused by scattering from the inner interface between glass and air.

B. *Particles and their characterization*

The experiments on nanoparticles used commercial, sulfate-functionalized polystyrene particles (Interfacial Dynamics Corp.). The suspensions are first diluted in deionized water to a volume fraction between 2×10^{-4} and 5×10^{-4}. The diluted suspensions are mixed at a 1:1 ratio with 100 mM borate buffer solution of pH 8.3 to reduce the screening length; without the buffer some of the particles do not enter the capillary, presumably because electrostatic repulsion prevents them from doing so. The hydrodynamic diameter of the particles is characterized by dynamic light scattering using a Zetasizer nano S (Malvern Instruments) and the electrophoretic mobility using a Zetasizer nano Z (Malvern Instruments). The characterization results are summarized in Table S1.

Table S1, Particle characterization

Particles	Reported Diameter[1] [nm]	Reported CV[1] [%]	Measured Diameter[2] [nm]	Measured CV [%]	Zeta potential[3] [mV]	Mobility[3] [10^{-9} m^2/VS]	Conductivity[3] [mS/mm]
51 nm Latex[4]	51	5	66	22	-40±17	-31±14	7.6
35 nm Latex[4]	35	4.2	35	34	-59±19	-46±15	7.5
19 nm Latex[4]	19	15	15	33	-75±5	-59±4	13.4
CCMV[5]	n.a.	n.a.	26±5	<5	-9.4±16.5	-7±13	0.253

[1] Mean diameter and its coefficient of variation (CV) as reported by the manufacturer.

[2] Peak of size distribution by number as measured by a Zetasizer nano S (Malvern Instruments). The uncertainty is estimated from half-width at half-maximum of the size distribution by number.

[3] As reported by a Zetasizer nano Z (Malvern Instruments).

[4] White Sulfate Latex, surfactant free, diluted in 1:1 mixture of di-ionized water and 100mM borate buffer 8.3 PH.

[5] Diluted in di-ionized water. See text for harvest and purification method.

CCMV virions were harvested and purified from previously infected California cowpea plants (Vigna Ungiculata) by homogenization of infected leaves followed by chloroform extraction and successive sucrose density gradient velocity sedimentation steps [S1]. Virion purity was then confirmed by four independent techniques: (i) UV-vis spectrophotometry gave a 260 nm / 280 nm absorbance ratio of 1.6, consistent with pure CCMV. (ii) SDS-PAGE gel electrophoresis revealed the absence of contaminant protein. (iii) Native 1% agarose gel electrophoresis stained with ethidium bromide gave a sharp band of signal with the expected electrophoretic mobility. (iv) Negative-stain transmission electron microscopy confirmed the monodispersity of the purified virions (Figure S2).
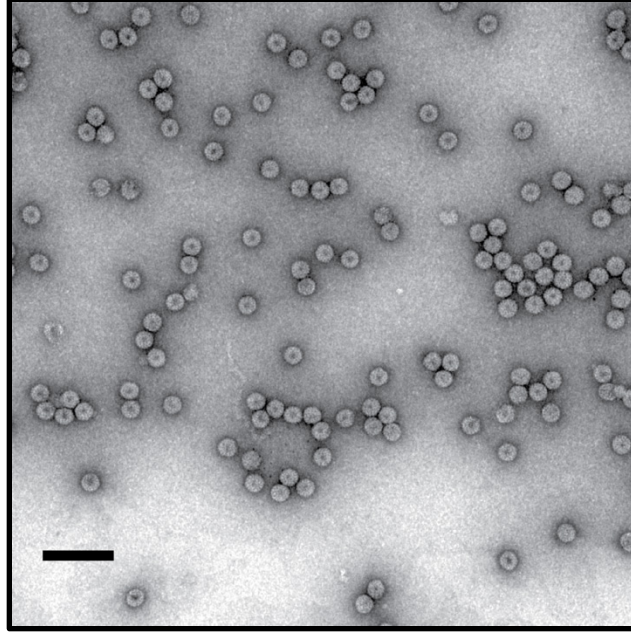
Figure S2. Transmission electron micrograph of CCMV stained with uranyl acetate. Scale bar shows 100 nm.

C. *The tracking algorithm*

Although there are several widely-used particle tracking codes, we found that they did not always correctly track the smallest particles in our experiments, which have a small extent (two to three pixels), relatively large intensity fluctuations (a consequence of the mode distribution in the fiber), and low signal-to-noise ratios. Thus we implemented a custom code, tailored for tracking these small particles. As detailed below, the main difference between our algorithm and previous ones is that ours estimates the position of each of the particles in each of the frames using information from both preceding and following frames. Where possible (for example, for the larger particles) we compared the results of our tracking algorithm to those from a MATLAB implementation of the Crocker-Weeks-Grier algorithm (http://www.physics.emory.edu/faculty/weeks//idl/) based on codes written by Daniel Blair and Eric Dufrene. For these larger particles, the agreement between the results obtained from these two routines was excellent.

The code we have developed extracts the track of each particle from a sequence of images of the type shown in figure 1e and 1f of the main text. The final tracks list the coordinates of each particle along the fiber axis and the scattered intensity as a function of time.

The tracking algorithm works in three stages: (1) Image restoration and noise removal; (2) locating the starting time and starting position of each of the tracks, and (3) following each track from beginning to end throughout the frames.

1. Image restoration and noise removal

The raw data, captured by our CMOS camera, consist of a sequence of frames with a large aspect ratio – usually the full width of the camera frame along the fiber axis, and only 12 pixels across

it. The large aspect ratio enables us to maintain a high frame rate while collecting all the light scattered in the fiber across the field of view.

The captured images contain a constant background which is due to scattering from the interface between the channel and the fiber core. This constant background is evaluated by calculating the median value for each of the pixels in the image across the entire sequence of frames. The background is then subtracted from all the images in the sequence. The mean value of all the pixel medians $\bar{I}_{med}$ is saved as well, and is used later to normalize the apparent intensity of each of the particles, as explained below.
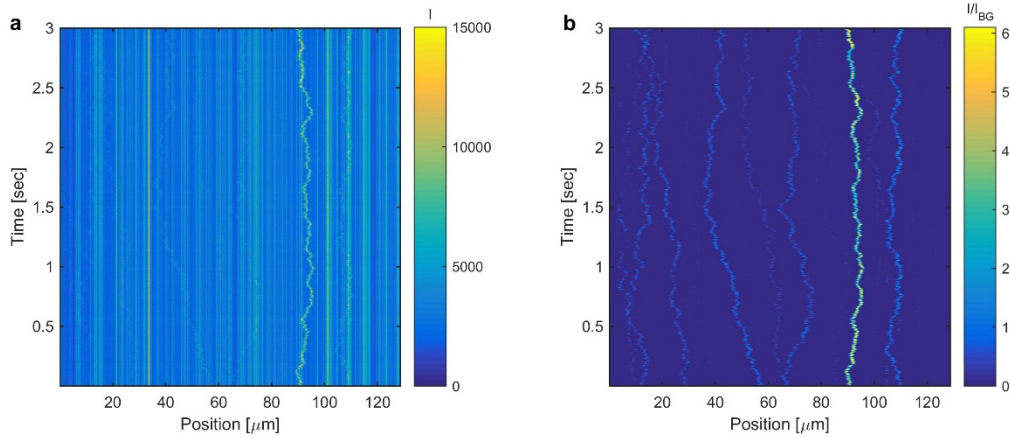

Figure S3. Background subtraction and noise removal example. a) Raw data, where each frame is reduced to a horizontal line in the plot. b) Same data after background subtraction and spatial filtering. The strongly scattering particle is 35 nm in diameter, and is visible also in the raw data. The weakly scattering particles are 19 nm in diameter. They are obscured by the background and noise in (a) but can clearly be seen in the background-corrected data in (b).

In addition to the constant background, each image contains electronic noise ("dark counts"). This noise usually has a spatial correlation length of a single pixel and a temporal correlation length of a single frame. In contrast, the scattering signal from the particles has longer spatial and temporal correlation lengths (typically 4-6 pixels and 10 or more frames). The electronic noise is reduced using two steps. First each image is convolved with a Gaussian surface of revolution with a half-width of 1 pixel, which reduces the pixel-noise without blurring the image. Next, since we are generally interested only in the position of the particle along the axis of the fiber, we project the two-dimensional images to a line parallel to the fiber axis. In this line, each pixel is a sum over the 12 transverse pixels of the raw image. Projecting the data reduces the pixel noise while leaving the scattering signal almost unchanged. All the lines, representing all the frames in the sequence, are then collected into a single matrix as shown in Figure S3.

When there is significant drift in the particle locations due to flow in the channel, the algorithm can fail to track the particles. In such cases we correct for the drift by first choosing one of the longer tracks, evaluating the distance the particle travels across the fiber, and dividing this value by the length of the track in frames. We then correct the data set by applying this linear shift to all the frames.

For the smallest particles we have tracked, the signal can become comparable to the noise, even after background subtraction and noise removal. Thus it is hard to correctly identify the locations of the particles in each frame; simply taking the highest values above some threshold in each frame leads to errors. We therefore choose a different approach, in which information about the location of the particle in both the previous frame and the subsequent frames is used to define a region in which the particle is most likely to be located in the current frame. The local brightness maximum *within that region* is identified as the candidate particle location.

The particle track – that is, its position along the fiber and its apparent brightness as a function of time – is determined in steps 2 and 3.

2. Finding the starting position of each track

To find the position of the particle in the first frame the algorithm first takes an average of the first 20 frames. In this averaged, one-dimensional frame, the different particles appear as blurred regions with a mean intensity that is higher than the surrounding noise. The maximum within each blurred region is identified using Matlab's built in function "findpeaks.m" (see http://www.mathworks.com/help/signal/ref/findpeaks.html). This function finds local maxima in a vector of numbers, and accepts two parameters. The first is the minimal distance between peaks, which should be chosen to select only the highest peak from peaks that are very close to each other (due to noise, for example). This parameter is determined manually from the data to correspond to the typical width of a blurred region (typically 8-10 pixels for the 51-nm particles and 10-12 pixels for the smaller particles). The second parameter is a threshold value, used to avoid identifying peaks in the noise. This threshold is determined manually from the data set. It is selected to be higher than the noise but lower than all the maxima of the blurred regions. The local maxima found by the routine are then saved as candidate particle locations.

Each of the candidate locations found above is taken to be the center of an interval with width w, where w is assigned a value that corresponds to the apparent width of a particle (typically 3-5 pixels). We then return to the first frame and calculate the particle locations in that frame using the centers of intensity in each interval. We store this position as the first element of the track, and we also store the apparent intensity of the particle, which is taken to be the sum of intensity over this interval normalized by $\bar{I}_{med}$.

3. Tracking

Starting with the second frame and the brightest particle found in the previous step, for each frame N the position in the frame $N+1$ is found using the following steps:
- An interval of width a is taken around the last known position of the particle, where a is chosen to be bigger than the apparent translation of the particle during the time interval, but smaller than the typical distance between particles. Typical values are 4-5 pixels for the 51-nm particles and 7-9 pixels for the 19-nm particles.
- For each pixel in this interval, the intensity values in frames $N+1$ through $N+20$ are averaged. The intensity maximum in the averaged interval is taken to be the candidate particle location in frame $N+1$.
- In frame $N+1$, an interval of width w is taken around the new candidate location. Then, as in step 2, the refined location is calculated from the center of intensity within that interval.

- In case the particle is not found (because, for example, the signal temporarily fluctuated to levels below the noise) the algorithm assigns a null value to that location, but keeps the last known location of the particle to try to find it again in the next frames. If the particle is not found for more than *M* consecutive frames, where *M* is entered as a parameter to the algorithm, the track is terminated. Typical values for *M* are 3 to 20 frames. The null values are not included in the analysis described in section F for determining the particle's diffusion coefficient and size.
- The apparent intensity of the particle at that location is taken to be the sum of intensity over this interval normalized by $\bar{I}_{med}$.
- The algorithm then proceeds to frame *N+2* and repeats the preceding steps. The algorithm stops tracking the particle when the signal goes below the acceptable threshold or when the frame sequence ends or when the particle was not found for more than *M* frames.

After the brightest particle is tracked, the next brightest particle is tracked, starting from the first frame and using the steps above, but with one additional step: if the position found for the current particle overlaps with one from a previously found track, the algorithm searches for the next highest maximum within the searching region. Thus if tracks cross one other, the algorithm will continue to track the particle with the same apparent intensity as the particle tracked until that point.

After all the tracks starting in the first frame are found, the same procedure repeats for other starting times along the sequence (typically every 100 frames). This ensures that particles that do not have a signal in the beginning of the sequence – for example, particles that come in to the field of view during the run, or particles that were not detected in the first frames due to noise – are tracked.

D. *Calculating the particle's diffusion coefficient and size:*

For each tracked particle the displacement histogram for each time interval in a range of 1-15 ms was collected. For the 51-nm particles, the displacement histograms were Gaussian down to the smallest lag times. We found that for the smaller particles and for the viruses, all displacement histograms above a lag time of 5 ms were Gaussian, while for shorter intervals the distributions were closer to exponential. We attribute this observation to the higher localization error in the case of smaller particles. The corresponding one-dimensional mean-square displacement (MSD) is calculated by taking the variance of the displacement distribution. To determine the diffusion coefficient D, we fit the MSD curve to $\mathrm{MSD}(t) = 2Dt$. The fit was performed for all time intervals for the larger particles, and for lag times larger than 5 ms for the smaller particles. The particle hydrodynamic diameter is then calculated using the Stokes-Einstein relation $R = k_B T / 6\pi\eta D$, where we took the temperature T=293 K and the viscosity of water to be $\eta = 1.00 \times 10^{-3} \, Pa \cdot s$. As an example, we show measurements and analysis of the diffusion of a 51-nm latex particle in water in Fig. S4.
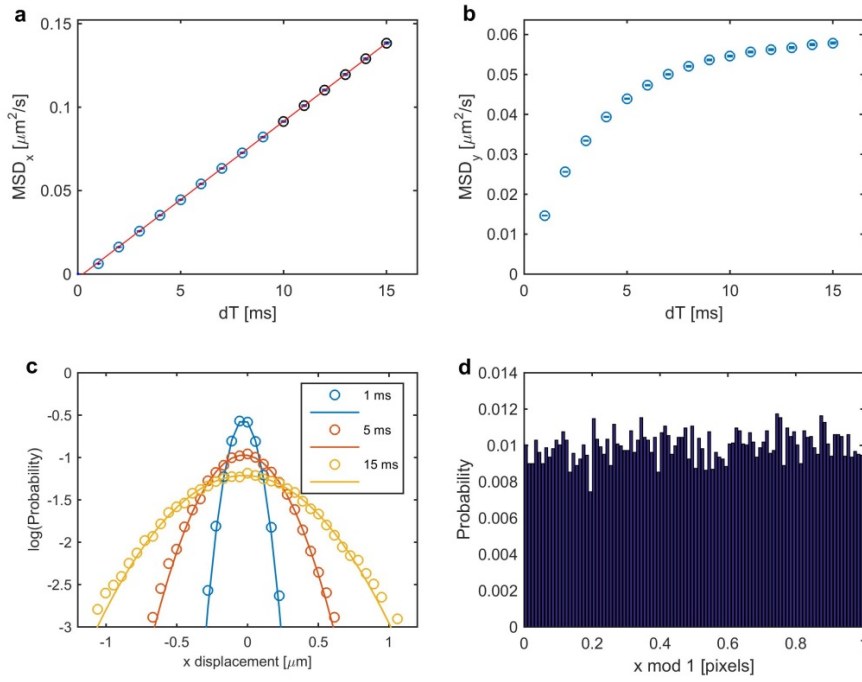
Figure S4. Example showing the tracking analysis. (a) The mean square displacement along the fiber axis, *x*. (b) The mean square displacement perpendicular (*y*) to the fiber axis. Here the displacement is limited by the channel boundaries. c) Histograms of displacement (shown on logarithmic scale) along the fiber axis for three time intervals. Solid curves show fits to Gaussian functions. d) Analysis of the pixel bias in our measurement. Flat histogram indicates no bias.

E.   *The effect of the fiber mode profile*

The apparent fluctuations of the measured intensities in each track as a function of time are related to the non-uniform mode profile inside the fiber. We therefore examine the correlation between the apparent brightness and the off-axis position of the particles inside the channel. Figure S5a depicts the two dimensional track of a single particle. In Fig S5b, we show the probability density function (based on 19,500 frames) of detected intensities as a function of position within the channel. The scattered intensity shows a minimum near the center of the channel, in agreement with numerical optical mode simulations performed by COMSOL Multiphysics.
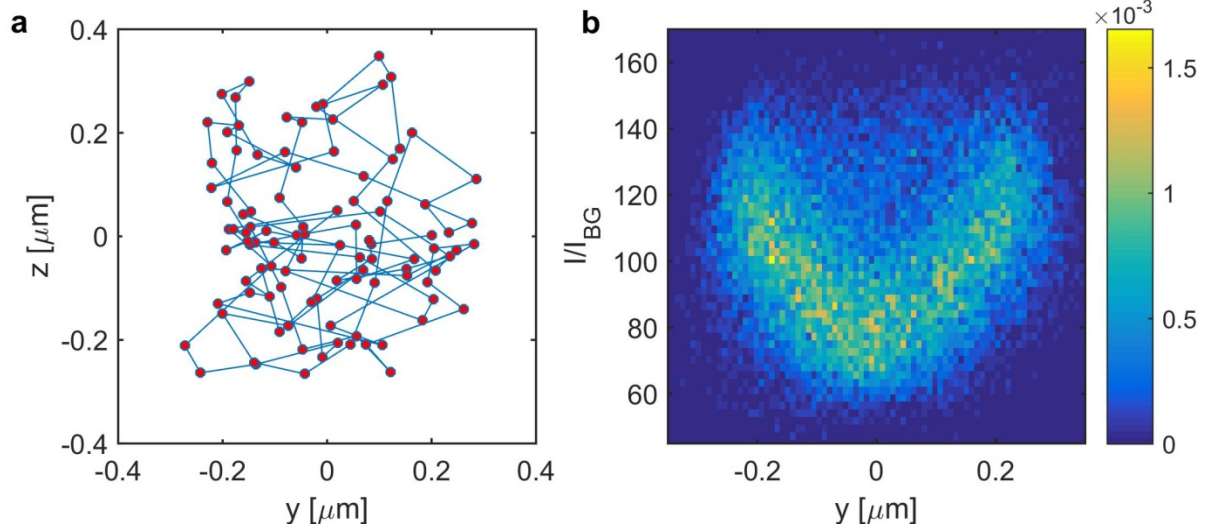
Fig S5: (a) The projected track of a single 51 nm latex particle in y-z plane (orthogonal to the imaging optical axis) for 100 frames. (b) Probability density function (color scale) of the scattered intensity (normalized by the mean background intensity) for the same particle at different projected-radial positions maps the mode profile inside the nano-channel. The larger scattered intensities seen when the particle is further from the axis indicate a concave-shaped mode profile, as expected.

F. *Calculation of the size distribution for CCMV virions based on tracking*

To calculate the size distribution of CCMV virions based on our measurements, we used 19 long tracks and averaged the detected intensity over each track. In the Rayleigh regime, the scattered intensity is proportional to the square of the polarizability, which in turn is proportional to the mass of the particle. Thus, by calculating the standard deviation over the mean of the square roots of the intensities, we estimate the upper bound on the mass polydispersity of the measured viruses to be 7.4%.

G. *Relation between intensity-enhancement and separation in the two-particle measurement*

To recover the interparticle distances from intensity measurements, we calculate the statistical relation between the two quantities considering the experimental conditions. Mathematically, the relation between these two quantitites is not one-to-one (the enhancement for multiple locations of the moving particle can be of the same value), but given the illumination and detection geometry and the well-defined optical mode profile inside the fiber core, we can map the intensity to a narrow, monomodal distribution of distances for interparticle separations smaller than half the propagation wavelength inside the fiber.

This statistical relation depends on the diameter of the channel and the position of the fixed particle, which can be independently determined using the localization and tracking algorithm of the previous section. For the presented data, the fixed particle was confirmed to be located close the detection optical axis.

With this information, we first calculate the joint probability distribution $P(F, r)$, with F the enhancement factor as defined in the text, and r the inter-particle distance, using Monte Carlo numerical integration. For this integration we assume two identical particles with diameters of 50 nm, one fixed on the inner wall of the channel on the detection optical axis and a second one inside the channel with uniform probability at physically allowed positions (hard-sphere interaction). We use the calculated mode profile for the fiber with an inner channel diameter of 400 nm at 670 nm illumination wavelength and consider a uniform phase in the plane orthogonal to the fiber axis. The total scattered intensity is calculated by integrating over the solid angle corresponding to the numerical aperture of the objective.
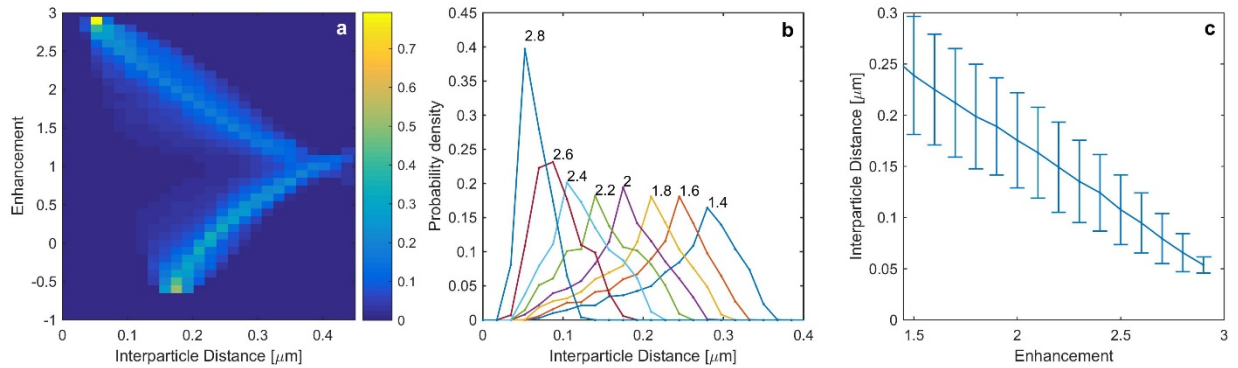

Fig S6: (a) Joint probability distribution of finding the mobile particle at a given distance from the fixed particle with a certain enhancement factor. The sum of probabilities for each enhancement value is normalized to 1. (b) The same probability distribution plotted in linear scale for enhancement values larger than 1.4. Numerical labels indicate the enhancement value for each curve (c) The mean interparticle distance plotted as a function of enhancement, with the error bars depicting the standard deviation around the mean value.

This joint probability distribution is depicted in Fig. S6a, and the probability distributions as a function of separation for enhancement values $F > 1.5$ are plotted in Fig. S6b. From these distributions, we determine the highest probable distance for each enhancement and its uncertainty. This result is depicted in Fig. S6c and is used for determining the interparticle interaction in Fig. 4 of the main text.

H. *Extracting the interaction potential from separation measurements*

Assuming Boltzmann statistics, the probability density of finding two particles at a distance *r>2a* is given by

$$P(r) \sim r^2 e^{-V(r)/k_\mathrm{B}T}$$

Assuming the potential has a finite range; $V(r_{\text{large}}) = 0$ we determine the potential at close distances from the histogram of the particle position distribution:

$$V(r) = -k_{\text{B}}T\ln \frac{P(r)r_{\text{large}}^2}{P(r_{\text{large}})r^2}$$

For our measurements we take $r_{\text{large}} = 250$ nm.

References

S1. Rao, A. L. N., R. Duggal, F. Lahser, and T. C. Hall. (1994) Analysis of RNA replication in plant viruses, pp. 216–236. In K. W. Adolph (ed.), Methods in Molecular Genetics: Molecular Virology Techniques, vol. 4. Academic Press, Orlando, Florida.

S2. Crocker, J. C. and Grier, D. G. Methods of Digital Video Microscopy for Colloidal Studies, Journal of Colloid and Interface Science, 179, 298-310 (1996)