

High throughput analysis of rare nanoparticles with deep-enhanced sensitivity via unsupervised denoising

Corresponding Author: Dr Sadao Ota

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have developed an optofluidic device and an unsupervised deep learning-based denoising method to create a nanoparticle analyzer with unprecedented throughput given its sensitivity. The technology, named "Deep Nanometry" (DNM), is able to detect 27-nm polystyrene beads at a rate of 96,000 events/s. For comparison, a few commercially available state-of-the-art flow cytometers can detect ~70 nm polystyrene beads at a rate of ~30,000 events/s (in practice). Machine learning can resolve dim signals from the noise where humans cannot even spot the signals. Machine learning is probably able to accomplish denoising because noise sources and optical signals originating from particles passing through the laser beam both have characteristic behavior in time. Nevertheless, a 6 dB improvement in the signal-to-noise ratio is impressive and significant when aiming to detect dim nanoparticles. The authors applied DNM to measure the concentration of extracellular vesicles (EVs) in serum from cancer patients and healthy donors. Although I have some doubts about the usefulness of measuring EVs with the used labels in serum (instead of plasma) and establishing percentage ratios of EVs relative to all particles, I think that EVs are a great candidate for DNM. All in all, I find this a great piece of technological work. Please find attached my major and minor comments.

MAJOR COMMENTS

1. At Pg. 2 the authors write: "the model removes instrument- and environment-specific noises". I agree with the authors. However, since the key novelty of this manuscript is the removal of noise, it would be nice to discuss noise sources of flow cytometer-based devices into more detail. According to my understanding, the instrument specific noises that the authors are correcting for involve the RIN-noise from the laser, the dark noise from the detector, and several sources of electronic noise. The environment-specific noise is the fluidics noise, i.e. actual particles in the buffer solution and the sheath fluid that cause a signal event. The main noise source that is left over and that cannot be corrected for is the shot-noise, also called photon-noise, correct? How close to shot-noise limited detection does the DNM become after denoising?
2. If you can detect the light scattering signals from single 26-nm polystyrene beads, you can detect 50-nm EVs, which is impressive. Currently there is no consensus about the diameter of the smallest EVs in plasma or serum. It might be that DNM is able to detect the smallest EVs in serum. However, in case the smallest EVs in serum are 30 nm, even the DNM would miss them. To improve the reproducibility of EV concentration measurements, it is good practice to determine the detection limits of each used detector (See: A compendium of single extracellular vesicle flow cytometry, Welsh et al., J. Extracell. Vesicles 2023). Preferably, this detection limit is provided in terms of EV diameters, but as Rayleigh theory for core-shell particles is not easily available and Mie theory takes considerable time to master and develop, a detection limit statement in terms of a polystyrene bead diameter or scattering cross section (nm²) would fulfill. Ideally, also detection limits for the fluorescence detectors are provided, but since suitable reference particles (e.g. sufficiently dim MESF beads) are unavailable, the authors can get away with this by dropping some lines to the discussion.
3. The light scattering theory at Pg. 3 of the supplementary information is questionable for four reasons. First, the authors use the Rayleigh approximation for scattering of unpolarized light. However, the used lasers are most likely polarized. Second, equation 2 contains a double integral but a single differential element, namely $d\theta$, which is inappropriate. What happened to the azimuthal angle? Third, the authors integrate over a square projected onto the unit circle, whereas a lens is spherical. Fourth, the authors aim to derive an equation for the photon flux collected by the lens, but in the manuscript (figure 3d) they do not report the photon flux. Therefore, equation 6 is abundant and fitting the data in figure 3d with the equation $y = a \cdot x^6$ is

sufficient. Please correct, clarify, or omit the light scattering theory.

4. It is very nice to see that the authors added an application to their manuscript. Two experimental choices, however, are questionable from a hemostatic and physiological perspective.

First, the authors acquired serum and labelled EVs with CD9 and CD147. Serum is fluid that remains after blood has clotted, formed by removing the clotting factors and blood cells from plasma. During clotting, platelets activate and release high numbers of EVs in a virtually uncontrolled way. During activation, platelets express CD9 and CD147 and it is likely that EVs released by platelets during activation also express CD9 and CD147. My impression is that the authors are rather studying this in vitro clotting process than in vivo concentrations of CD9 and CD147 EVs.

Second, the authors report the ratio of CD9+CD147+ particles to all particles. However, the fraction of CD9+CD147+ particles in serum seems irrelevant to me because the total concentration of particles in serum is dominated by lipoproteins. All lipoproteins together have a concentration of 2×10^{16} particles per mL, but HDL and LDL cannot be detected with the DNM. Still, the remaining lipoprotein concentration is high and depends on the diet and time between the last meal and venipuncture. Hence, the ratio of CD9+CD147+ particles to all particles rather tells something about the diet and energy balance of the donors (and the production of serum) than about the in vivo situation.

5. The conclusions are too general, lack quantitative statements and even contains subjective, unscientific statements like: "The ability of DNM to analyse massive numbers of individual particles rapidly and accurately" and "This approach simultaneously demonstrates notable sensitivity, throughput, and scalability, thereby enabling the rapid analysis of rare nanoparticles". Seven value judgements in two sentences! Given the impressive throughput, sensitivity, sensitivity improvements owing to machine learning, and potentially interesting clinical findings, I highly recommend basing the conclusions on the data and emphasize the key outcomes to substantiate the conclusions. Please remove all value judgements.

MINOR COMMENTS

1. At Pg. 7 the authors write: "(e) PR curve for evaluating the peak detection accuracy". However, figure 2e shows the "Precision" versus Recall. Measurement accuracy and precision are different things. Please select the right term and correct the manuscript accordingly.

2. Regarding the PR curve, did the authors verify that without ND filter in the attenuated detection path, the mAP was approximately 1?

3. At Pg. 9 the authors introduce the use of a commercially available single-photon counting flow cytometer (SP-FCM). Please add the product type and manufacturer of the SP-FCM to the manuscript to allow improved data interpretation and reproducibility.

4. For all figure legends: please add all used acronyms to the legends to improve quick interpretation of the figures.

5. At Pg. 9, the authors write: "In our method, the temporal width of each detected scattering peak is 1 μ s, theoretically allowing a throughput of up to 1,000,000 events per second, assuming equal distances between particles". According to the supplementary materials, the authors are aware that in flow cytometry-based instruments, particles arrive at random times in the interrogation volume and therefore the theoretical throughput without an acceptable number of multiplets is not the inverse of the acquisition time per particle. The claim of the authors is therefore misleading and should be omitted or rephrased.

6. At Pg. 9, the authors write: "The throughput we experimentally record is 96,000 events/second". Congratulations, because this is a higher throughput than the typical operational throughputs of standard flow cytometers. However, the authors do not provide data based upon which the reader can confirm this claim. Please substantiate this claim with data and indicate the fraction of multiplets (e.g. doublets, triplets) that were present during this experiment.

7. At Pg. 9, the authors write: "we conclude that the samples are detectable at concentrations of at least 2.82×10^3 particles/mL". Why did the authors not dilute further than 10^{-7} -fold?

8. Figure 4b is unnecessarily difficult to interpret and require improvement. The vertical axis of figure 4b shows the counts per minute. However, the reader is not aware of the flow rate and even if the flow rate would be defined, the reader needs to calculate how the counts relate to the concentration. Flow cytometer-based instruments have the (quite) unique capability to determine absolute particle concentrations. To improve data interpretation and reproducibility, therefore please report particle concentrations whenever possible.

9. The previous minor comment also applies to panel figure 4c, figure 5 and figure 6. It should be a sin to report percentage fractions of flow cytometry data without showing the absolute concentrations. Reported fractions, as shown in figure 4c are very difficult to reproduce due to differences between detection limits of flow cytometers and possible systematic errors of the derived particle concentrations. Moreover, the fraction of CD9+ particles in serum seems an irrelevant measure to me because the total concentration of particles in serum is dominated by lipoproteins. In turn, the concentration of lipoproteins strongly depends on the diet and time between the last meal and venipuncture. Hence, does the ratio of CD9+ particles to all particles tell something about EVs or rather something about the diet and energy balance of the donor? I strongly recommend plotting the concentration (/mL) of CD9+ particles on the vertical axis.

10. In figure 4c, the ratio of CD9+ particles to all particles is a sample property and should therefore be independent of the dilution of the sample. Why is the fraction of CD9+ particles affected by the dilution of the sample?

11. At Pg. 9 the authors write "The detection limit derived from blank PBS (background level of $+3\sigma$ for 4-minute measurements) is 0.01 detections/minute or 6.8×10^2 particles/mL". The authors used the buffer-only control to establish the concentration detection limit of their instrument. However, when considering the entire assay, the antibodies in buffer control would be a more suitable measure of the concentration detection limit. Please also indicate the background $+3 \times$ S.D. of the antibodies in buffer control in Figure 4b.

12. At Pg. 11 the authors claim: "To our knowledge, this is the first study to count EVs in unpurified serum at single-particle resolution accurately". However, in the Methods the authors write: "The supernatant was ultracentrifuged to prepare EVs at 49,500 rpm (210,000 $\times g$) using a TLS-55 rotor for 35 min at 4°C. The pellets were washed with 2 mL PBS by ultracentrifugation at 49,500 rpm (210,000 $\times g$) using a TLS-55 rotor for 35 min at 4°C and resuspended in PBS". The latter

implies that EVs were purified by ultracentrifugation and washing. Please clarify. In addition, there are many examples in literature where EVs from plasma or serum were directly measured in stained diluted plasma and serum. The main difference between those studies and the work of the authors, however, is that the smallest detectable EVs in the published studies are generally >100 nm.

13. At Pg. 12, the authors write: "The EV samples are diluted 50-fold with PBS as a recommended optimal condition for the reference SP-FCM". The interrogation volume and particle velocity differ between the SP-FCM and the DNM. The optimal dilution, which is the dilution with the maximum count rate without having swarm detection, therefore might differ between the SP-FCM and DNM. Generally, it is recommended to perform a dilution series to verify that after log-transformation, the slope between the measured concentration and dilution factor is -1 (See: A compendium of single extracellular vesicle flow cytometry, Welsh et al., J. Extracell. Vesicles 2023). Why the authors are sure that 50-fold dilution is sufficient?

14. I found two minor typos in the figures and equations:

a. S8a: measurements--> measurements

b. Extended figure 1a: training --> training

c. Equation 1 at Pg. 17: mean square error --> mean square error

15. At Pg. 16, the authors write: "To thoroughly remove cellular debris, the supernatant was filtered through a 0.22- μ m filter (Millipore)". Please specify the used filter. Please notice that if the authors used a PES filter, that the pore diameter of the filter is poorly defined and has little to do with "0.22 μ m". Please check manuscripts with scanning electron microscopy images of PES filters.

16. Please specify the antibody concentrations that were used during labelling. In addition, please make clear whether the authors used a procedure to remove possible aggregates from the stock of antibodies.

17. At pg. 16 the authors write: "To remove the unbound antibodies, EVs were then transferred to ultrafiltration spin-columns (300 kDa cut off, APRO Science) and centrifuged at 14,000 g at room temperature, followed by washing twice with PBS." If I understand this procedure well, unbound antibodies slip through the filter whereas EVs stay behind. Please make clear how you release the EVs from the filter membrane.

18. In fig. S5a, why the lasers for fluorescence excitation are not elliptically focused?

(Remarks on code availability)

I did not yet have time to review the code, but since it's on Github, I will check it out coming months.

Reviewer #2

(Remarks to the Author)

Please see the attached report.

(Remarks on code availability)

Reviewer #4

(Remarks to the Author)

Recommendation: reconsider after major revisions noted.

In the manuscript titled "High throughput analysis of rare nanoparticles with deep-enhanced sensitivity via unsupervised denoising", the authors introduce a computational pipeline that leverages on machine learning to denoise 1D time traces. Specifically, the authors apply the algorithm to the detection and subsequent analysis of nanoparticles via flow cytometry. The scope of the work is focussed on the flow cytometry and nanoparticle characterisation community and addresses an existing challenge to detect rare subpopulations within a heterogeneous sample mixture, which is only possible when high sensitivity is combined with high throughput. To put this work into a larger context of high throughput detection of single nanoparticles or highly sensitive detection of biological nanoparticles, similar approaches of the optofluidic platform exist either in the form of nanoFCM and single particle imaging FCM exist, as well as deep-learning implementations of particle detection, such as anomaly detection. The significance and novelty of this work comes down to demonstrating the use of unsupervised learning to push both the measurement throughput and detection sensitivity. The authors provide proof of validation experiments for detecting low abundance populations of cancer markers in a sample of EVs.

The manuscript is well written and structured and offers a promising implementation to access high throughput. Nevertheless, given that the manuscript describes a technical method, and the authors claim that this approach could potentially be a new standard for the analysis of nanoparticles using flow cytometry, it is this referee's opinion that both the description of the methods, formulation of the noise model, claims of multiparametric nanoparticle analysis and some of the conclusions stated in the manuscript could be further improved prior to acceptance. As such the authors should address the concerns listed below before this work is considered for publication:

Major

1. Regarding the unsupervised denoising model. In the formulation of the model, the authors assume that the signal detected is the sum of a noise free signal and signal-independent background noise. We strongly encourage the authors to reassess this statement. It is not clear how the authors can assume that shot noise, an intrinsic source of noise to any optical read out technique is a signal-independent background noise. Furthermore, when experiments are performed in complex sample mixtures such as serum, there are other background noise sources that should be included, such as those given by the presence of smaller nanoparticles such as proteins, which will lead to variable background scattering noise contributions.

2. Regarding the random sampling in the consensus model. The authors describe a consensus model where 100 denoised signals are sampled and the point-wise median is taken as the consensus. Is it possible to quantify or comment on the confidence or uncertainty in choosing such an approach. Namely, what is the confidence that such a particular peak is assigned to noise vs signal. The authors should also briefly comment on the reasoning behind choosing the median to represent whether a portion of the data is assigned to signal or to background noise. For example in figure 3a, the time trace before and after denoising is shown, yet some peaks that look like possible signals are assigned to noise, this begs the question of what is the uncertainty or rather confidence probability of such assignment.

3. Regarding the quantitative nature of the denoised time traces. In Figure 3d show the retrieved scattering amplitude signals after denoising and argue that these are quantitatively valid because these follow a size^6 scaling according to Rayleigh scattering theory. However the authors fail to convincingly show whether such denoising skews the retrieved scattering amplitude distribution. For instance figure 2c shows the distributions of the ground truth, attenuated vs denoised of the same nanoparticle sample, yet it is visible that the shape of the distribution of the scattering height is affected between ground truth and denoised. Furthermore the authors should explain the source of the uncertainties in the scattering amplitude scaling shown in Figure 3d; and test whether the denoising algorithm has any affect whatsoever. In the context of imaging, machine learning has been shown to be extremely good at solving segmentation, classification and detection problems, whilst showing mixed results for quantitative extraction of information. The authors have yet to show whether the quantitative aspect of the peak amplitude (scattering and fluorescence) is conserved by this algorithm and whether this affects all types of signals equally.

4. Regarding the training set. The training set for denoising corresponds to a 1s time-series of ultrapure water. This seems like a valid approach for training the model on the noise contributions that are sample-independent. The authors should discuss how this extends to more complex samples where sample-dependent noise cannot be ignored.

Regarding the applicability between scattering and fluorescence signals. The authors should comment if there is any reason why the denoising algorithm would be affected whether the detected signal is from scattering or fluorescence. Specifically the authors state in the SI that the “noise model is autoregressive because the scattering noise is temporally correlated” as shown in figure S2. Given that the fluorescence signal is both a spatially and temporally incoherent light source, would this model still be suitable?

Minor

The manuscript could use a quick polish to amend typos found in the figure: concensus vs consensus.

The author should comment on the conditions that lead to the setup running at 1e5 events/s and state whether this same throughput rate was generally maintained throughout the different measurements shown in the figures in the manuscript. For instance, Figure 4b claims a 4 minute trace. If the setup were running at 1e5 events/s this would lead to about 6e6 events/min, yet only about 1e4 EVs are detected at the highest serum concentration. Would it thus be correct that only 1% of the events detected correspond to EVs. If so, what do the other events correspond to? Perhaps, background signal.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have developed an unsupervised deep learning-based denoising method to realize a nanoparticle analyzer with unprecedented throughput given its sensitivity. The technology, named “Deep Nanometry” (DNM), is able to detect 30-nm polystyrene spheres at a rate of 100,000 events/s. For comparison, a few commercially available state-of-the-art flow cytometers can detect ~70 nm polystyrene beads at a rate of ~30,000 events/s (in practice). Machine learning can resolve dim signals from the noise where humans cannot even spot the signals. Machine learning is probably able to accomplish denoising because noise sources and optical signals originating from particles passing through the laser beam both have characteristic behavior in time. Nevertheless, a 6 dB improvement in the signal-to-noise ratio is impressive and significant when aiming to detect dim nanoparticles. The authors applied DNM to measure the concentration of extracellular vesicles (EVs) in serum from cancer patients and healthy donors. In my opinion, the authors have well-addressed my previous major and minor comments and I particularly appreciate the added noise analysis. I have just a few more minor comments left, please see below.

MINOR COMMENTS

1. In Figure S10 the authors investigate the detection limit of DNM. For that they use 40-nm polystyrene beads, with a specified median diameter of 40 nm and a standard deviation of 6.7 nm. In contrast to the standard deviation, only the specified median diameter is traceably determined, i.e. with an uncertainty statement. The standard deviation is rather the maximum possible standard deviation according to the manufacturer. NIST-traceable beads from ThermoFisher typically have a smaller standard deviation than specified. It seems that your dataset confirms my statement. However, no action

needed here, as it is difficult to obtain the “true” standard deviation.

Below 30 nm, the data does not follow the specified curve, hence the authors define the detection limit as 30 nm for polystyrene spheres. I agree with that approach. Next, however, the authors claim “DNM is able to analyze 30 nm polystyrene beads at over 100,000 events/second, scaling to millions of events.” and “DNM can analyze 30 nm polystyrene beads at a throughput of over 100,000 events/second”. I do not agree with the latter statements, because 30-nm polystyrene beads also have a considerable standard deviation and given that the detection limit is 30 nm for a polystyrene sphere, half the population of 30-nm polystyrene beads would fall below the detection limit. Technically, DNM is therefore able to detect 40 nm polystyrene beads and not 30 nm polystyrene beads. I guess this is a matter of phrasing the text well to avoid an invalid claim.

2. The authors used rainbow calibration particles to calibrate MESF values for FITC, Alexa488 and Tide FluorTM245 5WS fluorescence. However, rainbow beads do not have FITC, Alexa488 and Tide FluorTM245 5WS molecules and the emission spectrum of rainbow beads therefore differs from the fluorescent molecules used by the authors to label EVs. Possibly assigned MESF values of the manufacturer to the rainbow beads are therefore only valid if the authors used the same excitation laser and spectral filters as the manufacturer, which is unlikely the case. Hence the question: how did the authors perform the MESF calibration? Did the authors consider possible differences between the fluorescence optics of the manufacturer and the authors?

3. In the Discussion, the authors ambitiously wrote: “An important milestone for future work will be developing a method capable of additionally removing signal-dependent shot noise”. Shot noise originates from the fundamental quantum nature of light, specifically the discrete and probabilistic arrival of photons. This noise is inherent to electromagnetic radiation and the detection thereof and cannot be removed. Experimental limitations induced by shot noise can in general be mitigated by increasing the number of collected photons. In my opinion, the milestone for future work to “removing signal-dependent shot noise” seems therefore impossible.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)

The authors have provided a thorough response to the questions raised in the initial review. The latest manuscript shows clear improvements and has addressed all my concerns. I recommend the publication of the manuscript.

(Remarks on code availability)

The code is well organized and instructions are provided.

Reviewer #4

(Remarks to the Author)

I'd like to commend the authors for their detailed, thorough and clear responses to all comments and suggestions raised not only by me but by the other reviewers. The revised manuscript has vastly improved and better articulates the structure, advantages and current limitations of the algorithm. It is this reviewer's opinion that the manuscript is ready for publication. I very much look forward to reading their future work, and implementing these algorithms in my own research.

(Remarks on code availability)

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Reviewer #1:

Overall comment

The authors have developed an optofluidic device and an unsupervised deep learning-based denoising method to create a nanoparticle analyzer with unprecedented throughput given its sensitivity. The technology, named “Deep Nanometry” (DNM), is able to detect 27-nm polystyrene beads at a rate of 96,000 events/s. For comparison, a few commercially available state-of-the-art flow cytometers can detect ~70 nm polystyrene beads at a rate of ~30,000 events/s (in practice). Machine learning can resolve dim signals from the noise where humans cannot even spot the signals. Machine learning is probably able to accomplish denoising because noise sources and optical signals originating from particles passing through the laser beam both have characteristic behavior in time. Nevertheless, a 6 dB improvement in the signal-to-noise ratio is impressive and significant when aiming to detect dim nanoparticles. The authors applied DNM to measure the concentration of extracellular vesicles (EVs) in serum from cancer patients and healthy donors. Although I have some doubts about the usefulness of measuring EVs with the used labels in serum (instead of plasma) and establishing percentage ratios of EVs relative to all particles, I think that EVs are a great candidate for DNM. All in all, I find this a great piece of technological work. Please find attached my major and minor comments.

Response to overall comment

We thank the reviewer for critical and constructive comments on our manuscript. We also appreciate their recognition of the potential value of our technology, Deep Nanometry (DNM), and their positive feedback on our research. The comments on our methodology and EV analysis provided valuable insight and enabled us to clearly explain the capabilities and applications of our technology. We have carefully considered all the comments and addressed the raised concerns.

Major Comment 1

At Pg. 2 the authors write: “the model removes instrument- and environment-specific noises”. I agree with the authors. However, since the key novelty of this manuscript is the removal of noise, it would be nice to discuss noise sources of flow cytometer-based devices into more detail. According to my understanding, the instrument specific noises that the authors are correcting for involve the RIN-noise from the laser, the dark noise from the detector, and several sources of electronic noise. The environment-specific noise is the fluidics noise, i.e. actual particles in the buffer solution and the sheath fluid that cause a signal event. The main noise source that is left over and that cannot be corrected for is the shot-noise, also called photon-noise, correct? How close to shot-noise limited detection does the DNM become after denoising?

Response (major) 1

We appreciate the reviewer pointing out the importance of discussing the noise sources of our flow cytometer-based devices in more detail. We agree that understanding these noise sources is crucial for clearly evaluating our denoising approach. To address this comment, we have performed frequency analysis on the noises and investigated the variance of signals according to the signal intensities as follows.

We first perform a frequency analysis on the instrument-specific noise before applying the denoising method. Herein, RIN, dark noise from the detector, and fluid noise from the PBS are considered instrument-specific noise (**Figure R1_1**). We measured each noise as follows: (i) we measured the RIN noise by directing the laser input into the PMT after attenuating the laser power, (ii) the dark noise with the laser turned off, and (iii) the fluidic noise with particle-free PBS flowing in the channel as a sample stream. As a result, the typical frequency peaks seen in the power spectrum of RIN noise and dark noise do not appear in that of the fluidic noise, suggesting that the fluidic noise is dominantly large compared to the others in the DNM (**Figure R1_1**).

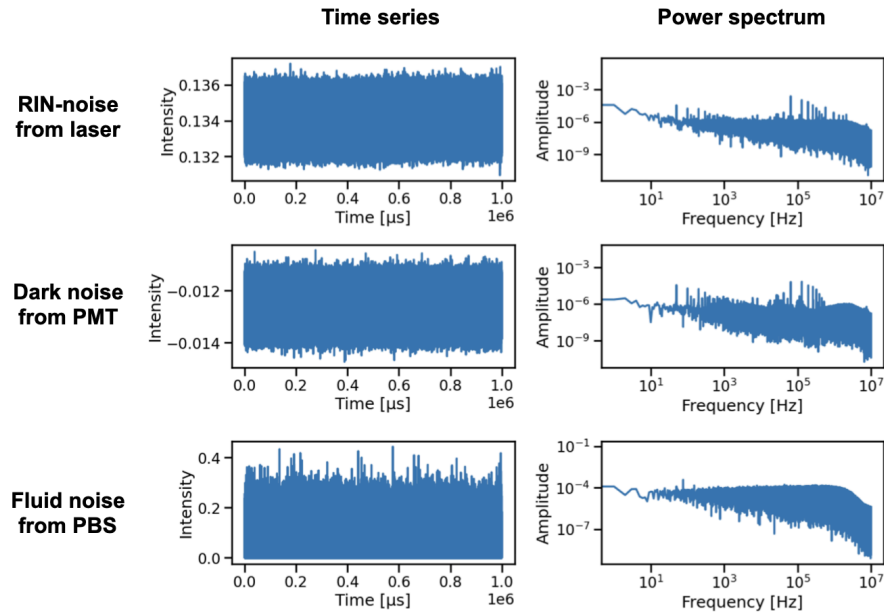


Figure R1_1:
Scattering time series and frequency power spectrum for each type of noise (RIN noise, dark noise, flow noise).

Next, we further analyzed the variance of the attenuated signal intensity before and after denoising as a function of the corresponding ground-truth signal intensity. Consequently, as the reviewer noted, we show that, assuming that the shot noise limit is the state where only signal-dependent noise remains, shot noise (photon noise) is the main noise source after denoising.

In the analysis, we used a set of corresponding ground-truth, attenuated, and denoised time series from **Figure 2** of the original manuscript. In the analysis, we first max normalized the ground-truth time series, then binned these normalized ground-truth signal intensities into 100 intervals. For each bin, we calculated the variance of the attenuated signal at corresponding time points.

As a result, as shown in **Figure 2(f)**, before denoising, the variance of the signal from the attenuated arm increases approximately linearly with ground-truth signal intensity, matching the behavior of shot noise. The y-intercept of the line – the variance of the noise when the ground-truth signal intensity is zero – is approximately equal to the variance of the PBS-derived fluidic noise. Since the variance of the sum of two independent random variables is additive, this indicates that the noise in the attenuated time series is the sum of the PBS-derived fluidic noise and shot noise. To make this clear, we subtracted the variance of the PBS-derived fluidic noise from that of the attenuated time series and plotted it as the

dashed line in **Figure R1_1**. When this analysis is repeated on the variance of the denoised signal, its slope becomes close to the slope before denoising, indicating that the shot noise indeed remains. However, the y-intercept is now close to zero, indicating the removal of the PBS-derived fluidic noise. We believe that deviations in the linearity of the denoised signal variance are due to errors in the denoising. Overall, this analysis confirms that the signal-independent noise component is largely removed.

Accordingly, we have revised the main text to include these analyses in lines 211-231 on page 6:

“Additionally, we analyzed the attenuated signal variance before and after denoising as a function of the ground-truth signal intensity to evaluate how closely DNM approaches the shot noise limit. We consider the shot noise limit to be the state in which there is only noise that depends on the optical signal and whose variance is proportional to the ground-truth signal intensity. In the analysis, we first max normalized the ground-truth time series, then binned these normalized ground-truth signal intensities into 100 intervals. For each bin, we calculated the variance of the attenuated signal at corresponding time points. We plot the results of this analysis in **Figure 2(f) (R1_2 in this response)**.

Before denoising, the variance of the signal from the attenuated arm increases approximately linearly with ground-truth signal intensity, matching the behavior of shot noise. The y-intercept of the line – the variance of the noise when the ground-truth signal intensity is zero – is approximately equal to the variance of the PBS-derived fluidic noise. Since the variance of the sum of two independent random variables is additive³⁶, this indicates that the noise in the attenuated time series is the sum of the PBS-derived fluidic noise and shot noise. To make this clear, we subtracted the variance of the PBS-derived fluidic noise from that of the attenuated time series and plotted it as the dashed line in **Figure 2(f) (R1_2 in this response)**.

We then repeated this analysis on the variance of the denoised signal. The slope of the variance here is close to the slope before denoising, indicating that the shot noise indeed remains. However, the y-intercept is now close to zero, indicating the removal of the PBS-derived fluidic noise. The curve slightly deviates from the ideal linear shape due to imperfections in the denoising result.”

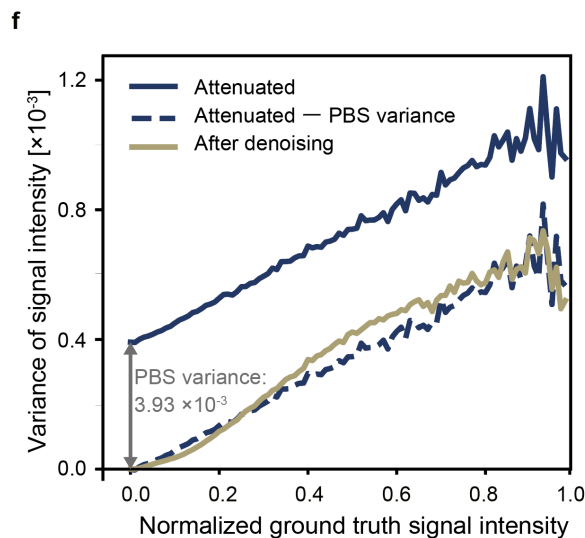


Figure 2(f), Figure R1_2 in this response:

The variance of the attenuated signal before denoising (blue line) and after denoising (olive line) as a function of the corresponding normalized ground truth signal intensity. The dashed blue line is the variance of the noisy attenuated signal minus the variance of signal-independent PBS-derived fluidic noise.

“[36] Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P. & Meester, L. E. Covariance and correlation. 135–150 (2005) doi:10.1007/1-84628-168-7_10”

We appreciate this insightful question, helping us to improve the clarity and depth of our discussion on noise sources and the effectiveness of our denoising model.

Major Comment 2

If you can detect the light scattering signals from single 26-nm polystyrene beads, you can detect 50-nm EVs, which is impressive. Currently there is no consensus about the diameter of the smallest EVs in plasma or serum. It might be that DNM is able to detect the smallest EVs in serum. However, in case the smallest EVs in serum are 30 nm, even the DNM would miss them. To improve the reproducibility of EV concentration measurements, it is good practice to determine the detection limits of each used detector (See: A compendium of single extracellular vesicle flow cytometry, Welsh et al., J. Extracell. Vesicles 2023). Preferably, this detection limit is provided in terms of EV diameters, but as Rayleigh theory for core-shell particles is not easily available and Mie theory takes considerable time to master and develop, a detection limit statement in terms of a polystyrene bead diameter or scattering cross section (nm²) would fulfill. Ideally, also detection limits for the fluorescence detectors are provided, but since suitable reference particles (e.g. sufficiently dim MESF beads) are unavailable, the authors can get away with this by dropping some lines to the discussion.

Response (major) 2

We thank the reviewers for their valuable comments and suggestions. We have consulted the suggested comprehensive reference¹ (A compendium of single extracellular vesicle flow cytometry, Welsh et al. Vesicles 2023) and here provide the following discussions for each scattering and fluorescence detection.

Scattering detection

We agree that it is good practice to determine the detection limits of each detector and the DNM misses 30 nm EVs. Although it is preferable to provide the detectable EV diameters, because of the unknown EV size distribution in our sample, we will in the following discuss the detection limit using polystyrene (PS) beads, for which the approximate true size distribution has been made available by the manufacturer.

In this analysis, we determined the scattering detection limit by fitting and comparing the size distribution of 40-nm PS beads obtained by experimental measurements and denoising with the size distribution of the same beads estimated from the spec sheet.

We first estimated the particle size distribution, shown as an olive histogram in **Figure R1_3(a)**, by experimentally measuring the scattering height of the detected 40-nm particles and converting the height into the particle diameter using the Rayleigh scattering theory². In the theory, the scattering amplitude for small particles is expressed as

$$\begin{aligned} \text{Scattering amplitude } (d) &= A \times \sigma_{\text{scatt}}(d, n_{\text{med}}, n_{\text{particle}}, \lambda) \\ &= A \times \frac{2\pi^5 d^6 n_{\text{med}}^4}{3\lambda^4} \left| \frac{m^2 - 1}{m^2 + 2} \right|^2 = cd^6, \quad (1) \end{aligned}$$

where σ_{scatt} is scattering cross section, $m = \frac{n_{\text{particle}}}{n_{\text{med}}}$, d is the particle size, n_{med} is the refractive index of medium, n_{particle} is the refractive index of the particle, λ is the excitation

wavelength. The scattering amplitude is proportional to the scattering cross section, scaled by a factor A . All of these parameters, except d , are dependent of the experimental setup and can be reduced to the constant factor c . In the fitting, using the medians of the experimentally obtained scattering height for 40-nm and 60-nm particles, respectively, we obtained $c = 3.5 \times 10^{-11}$ in **Equation (1)**. We describe this Rayleigh scattering theory and how we determined the factor c in **Supplementary Note 4**.

Then, we also calculated the theoretical distribution of the 40-nm particles, shown as a black line in **Figure R1_3(a)**. Here we assumed that the 40-nm PS bead size follows a normal distribution with mean and variance listed on the spec sheet³ as 40 ± 6.7 nm.

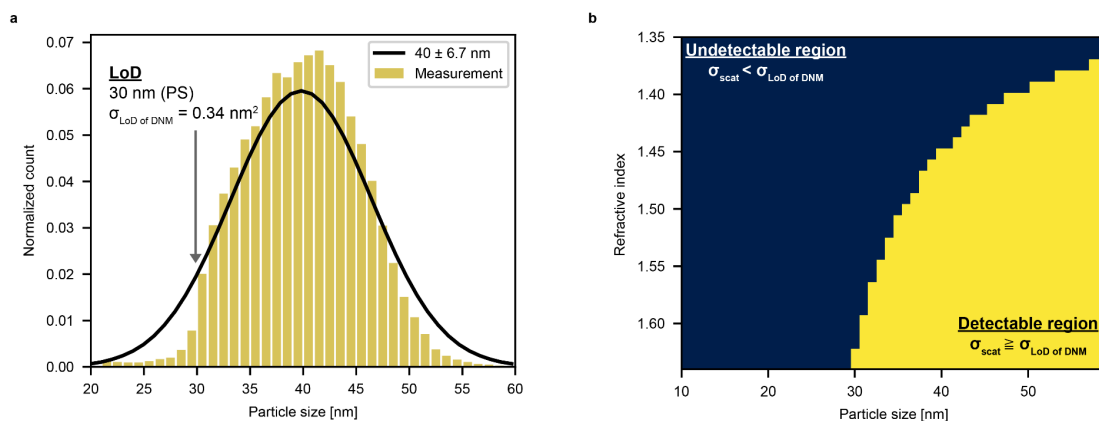
We normalized the experimentally measured histogram by dividing the counts by double the number of particles detected as 40 nm or larger, accounting for the fact that we are more likely to miss very small particles, but not larger ones. Here we assume that the total number of particles in the entire population is double the number of particles detected as 40 nm or higher. When assuming the symmetric particle size distribution around 40 nm, the true number of particles with a size less than 40 nm should be equal to the number of particles with a size greater than 40 nm.

Collectively, by comparing the experimentally measured size histogram with the theoretically estimated distribution, we can obtain the ratio of the number of detected particles to that anticipated from the spec sheet. We defined the LoD as the smallest size at which the particle count ratio does not fall below 50%^{4,5,6}, and thereby obtained an LoD of 30 nm ($\sigma_{\text{LoD of DNM}} = 0.34 \text{ nm}^2$).

In addition, we calculated the particle size and refractive index combinations that are detectable by DNM based on the obtained $\sigma_{\text{LoD of DNM}}$ ($= 0.34 \text{ nm}^2$). We calculated scattering cross sections for each particle size and refractive index pair based on **equation (1)**, and determined the region where the obtained scattering cross sections were smaller than $\sigma_{\text{LoD of DNM}}$ were considered undetectable regions, and those that were larger were considered detectable regions (**Figure R1_3(b)**).

Note that these results are only approximations since the conversion from scattering height to particle size is based on Rayleigh theory, which is not perfect as seen in the deviation of the distributions in **Figure R1_3(a)**. If Mie theory is adopted, the intensity dependence, which is proportional to the fourth power of the particle size, would result in a wider distribution of experimentally detected particle size histogram, further pushing down the LoD. However, we dare to adopt the Rayleigh theory for more stringent estimation.

Based on this analysis, we have added **Figure R1_3** as **Supplementary Figure S10** as follows:



Supplementary Figure S10, Figure R1_3 in this response:

The estimation of the limit of scattering detection (LoD) (a) Comparison of the size distribution using the denoised data of $40 \text{ nm} \pm 6.7 \text{ nm}$ PS beads (Thermo Fisher, F8795) for estimating the LoD. First, we define the detection threshold value below which the 99.9% of 1-second PBS scattering time series values are included. Learning from the previous research¹⁰, we compare the experimentally estimated size distribution of PS beads with that estimated based on the spec sheet provided by a manufacturer (as a ground truth distribution). To estimate the experimental size distribution, we converted the scatter height to particle size using Rayleigh scattering theory⁹ and fitting (**Supplementary Note 4**). Herein, we normalized the experimentally measured histogram by dividing the counts by double the number of particles detected as 40 nm or larger. We defined the LoD as the smallest size at which the particle count ratio does not fall below 50%¹¹⁻¹³, and thereby obtained the LoD of 30 nm ($\sigma_{\text{LoD of DNM}} = 0.34 \text{ nm}^2$). **(b)** Detectable size and refractive index range calculated based on scattering cross section of LoD. We calculated the particle size and refractive index combinations that are detectable by DNM based on the obtained $\sigma_{\text{LoD of DNM}}$ ($= 0.34 \text{ nm}^2$). We calculated scattering cross sections for each particle size and refractive index pair based on **equation (1)**, and determined the region where the obtained scattering cross sections were smaller than $\sigma_{\text{LoD of DNM}}$ were considered undetectable (blue) regions, and those that were larger were considered detectable (yellow) region.

We have also added the following sentence in line 12-13 on page 1 of the main text (abstract):

“DNM is able to analyze 30 nm polystyrene beads at over 100,000 events/second, scaling to millions of events.”

in line 54-56 on page 2 of the main text:

“DNM can analyze 30 nm polystyrene beads at a throughput of over 100,000 events/second (**Supplementary Figure S1, Supplementary Figure S10**) and scale up to millions of events.”

in line 474-476 on page 17 of the main text:

DNM demonstrated the sensitive detection of 30 nm polystyrene nanoparticles, achieving a high throughput detection of over 100,000 events per second and scalability of over a million event analyses.

in line 349-350 on page 12 of the main text:

“See **Supplementary Figure S10, S11** for a detailed discussion about the limit of scattering and fluorescence detection³⁷.”

“[9] Bohren, Craig F., and Donald R. Huffman. Absorption and scattering of light by small particles (John Wiley & Sons, 2008). Cited as [2] in this Response (major) 2.

[10] Welsh, Joshua A., et al. "A compendium of single extracellular vesicle flow cytometry." Journal of extracellular vesicles 12.2 (2023): e12299; Cited as [37] in the modified main text; Cited as [1] in this Response (major) 2.

[11] Pan, Lei, et al. "Determination of X-ray detection limit and applications in perovskite X-ray detectors." Nature communications 12.1 (2021): 5258. Cited as [4] in this Response (major) 2.

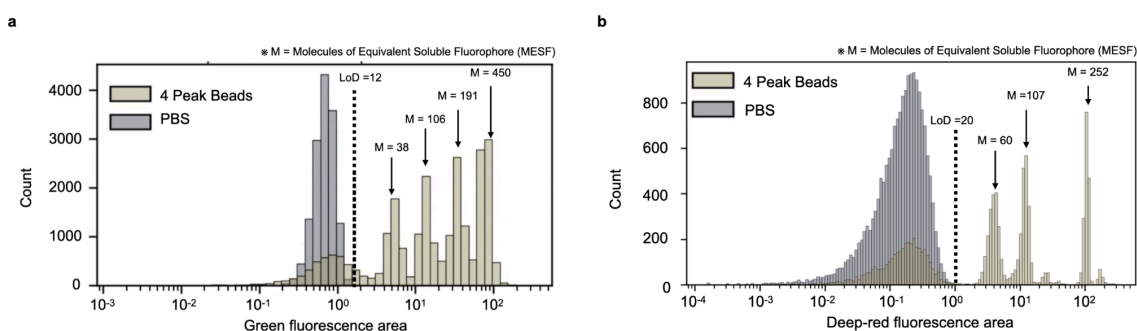
[12] Currie, Lloyd A. "Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995)." Pure and applied chemistry 67.10 (1995): 1699-1723. Cited as [5] in this Response (major) 2.

[13] Long, Gary L., and James D. Winefordner. "Limit of detection. A closer look at the IUPAC definition." Analytical chemistry 55.7 (1983): 712A-724A. Cited as [6] in this Response (major) 2.”

Fluorescence detection

For the green fluorescence detection, we measured and calculated the Mean Equivalent Soluble Fluorophore (MESF) using rainbow calibration particles (Spherotech, RCP-05-5) and compared it to the blank (PBS) data. We thereby determined the LoD of FITC fluorescence to be 12 MESF (QE: Quantum Efficiency = 0.92) and the LoD of Alexa488 to be 12 MESF (QE = 0.92) (**Figure R1_4(a)**), below which 99.9% of the blank PBS time series signals are included. For the deep-red fluorescence detection, we did the same experiment. As a result, we determined the LoD of APC fluorescence to be 20 MESF (QE: Quantum Efficiency = 0.68) and the LoD of Tide FluorTM5WS to be 55 MESF (QE=0.25) (**Figure R1_4(b)**).

Based on this analysis, we have added **Figure R1_4** as **Supplementary Figure S11** as follows:



Supplementary Figure S11, Figure R1_4 in this response:

The estimation of the limit of (a) green and (b) deep-red fluorescence detection (LoD) For fluorescence detection, we measured rainbow calibration particles (Spherotech, RCP-05-5). We calculate the Mean Equivalent Soluble Fluorophore (MESF)¹⁴ for green fluorescence, compared it to the PBS data and determined a LoD of FITC fluorescence equal to 12 MESF (Quantum Efficiency (QE) = 0.92), LoD of Alexa488 equal to 12 MESF (QE = 0.92) (**Supplementary Figure S11(a)**). For the deep-red fluorescence detection, we did the same experiment. As a result, we determined the LoD of APC fluorescence to be 20 MESF (QE = 0.68) and the LoD of Tide FluorTM5WS to be 55 MESF (QE = 0.25) .

[14] Maia, J. *et al.* Employing Flow Cytometry to Extracellular Vesicles Sample Microvolume Analysis and Quality Control. *Front Cell Dev Biol* **8**, 593750 (2020).

We have also added the following sentence in line 349-350 on page 12 of the main text:

“See **Supplementary Figure S10, S11** for a detailed discussion about the limit of scattering and fluorescence detection³⁷.”

References

3. <https://www.thermofisher.com/order/catalog/product/jp/ja/F8795>

Major Comment 3

The light scattering theory at Pg. 3 of the supplementary information is questionable for four reasons. First, the authors use the Rayleigh approximation for scattering of unpolarized light. However, the used lasers are most likely polarized. Second, equation 2 contains a double integral but a single differential element, namely $d\theta$, which is inappropriate. What happened to the azimuthal angle? Third, the authors integrate over a square projected onto the unit circle, whereas a lens is spherical. Fourth, the authors aim to derive an equation for the photon flux collected by the lens, but in the manuscript (Figure 3d) they do not report the photon flux. Therefore, equation 6 is abundant and fitting the data in figure 3d with the equation $y = a \cdot x^6$ is sufficient. Please correct, clarify, or omit the light scattering theory.

Response (major) 3

We thank the reviewer for their clear and detailed comments regarding the light scattering theory presented on page 3 of our supplementary information. We appreciate the reviewer's suggestion and agree that fitting $y = ax^6$ is sufficient for our purposes. In response, we have removed **Supplementary Note 2**, but we added another explanation as **Supplementary Note 4** to answer another reviewer's request in parallel. In this note, we added the following relevant description for Rayleigh scattering theory¹ and fitting strategy.

“Note 4: Rayleigh scattering theory and fitting strategy

Using the Rayleigh theory⁹, the scattering amplitude for small particles is expressed as

$$\begin{aligned} \text{Scattering amplitude } (d) &= A \times \sigma_{\text{scat}}(d, n_{\text{med}}, n_{\text{particle}}, \lambda) \\ &= A \times \frac{2\pi^5 d^6 n_{\text{med}}^4}{3\lambda^4} \left| \frac{m^2 - 1}{m^2 + 2} \right|^2 = cd^6, \quad (1) \end{aligned}$$

where σ_{scat} is scattering cross section, $m = \frac{n_{\text{particle}}}{n_{\text{med}}}$, d is the particle size, n_{med} is the refractive index of the medium, n_{particle} is the refractive index of the particle, and λ is the excitation wavelength.

The scattering amplitude is proportional to the scattering cross section, scaled by a factor A . All of these, except d , are parameters of the experimental setup and can be reduced to the constant factor c . In the fitting, using the medians of the experimentally obtained scattering height for 40 nm and 60 nm particles, respectively, we obtained $c = 3.4 \times 10^{-11}$ in the **equation (1)**.”

“[9] Bohren, Craig F., and Donald R. Huffman. Absorption and scattering of light by small particles (John Wiley & Sons, 2008). Cited as [1] in this Response (major) 3.”

Major Comment 4

It is very nice to see that the authors added an application to their manuscript. Two experimental choices, however, are questionable from a hemostatic and physiological perspective.

First, the authors acquired serum and labeled EVs with CD9 and CD147. Serum is fluid that remains after blood has clotted, formed by removing the clotting factors and blood cells from plasma. During clotting, platelets activate and release high numbers of EVs in a virtually uncontrolled way. During activation, platelets express CD9 and CD147 and it is likely that EVs released by platelets during activation also express CD9 and CD147. My impression is

that the authors are rather studying this in vitro clotting process than in vivo concentrations of CD9 and CD147 EVs.

Second, the authors report the ratio of CD9+CD147+ particles to all particles. However, the fraction of CD9+CD147+ particles in serum seems irrelevant to me because the total concentration of particles in serum is dominated by lipoproteins. All lipoproteins together have a concentration of 2E16 particles per mL, but HDL and LDL cannot be detected with the DNM. Still, the remaining lipoprotein concentration is high and depends on the diet and time between the last meal and venipuncture. Hence, the ratio of CD9+CD147+ particles to all particles rather tells something about the diet and energy balance of the donors (and the production of serum) than about the in vivo situation.

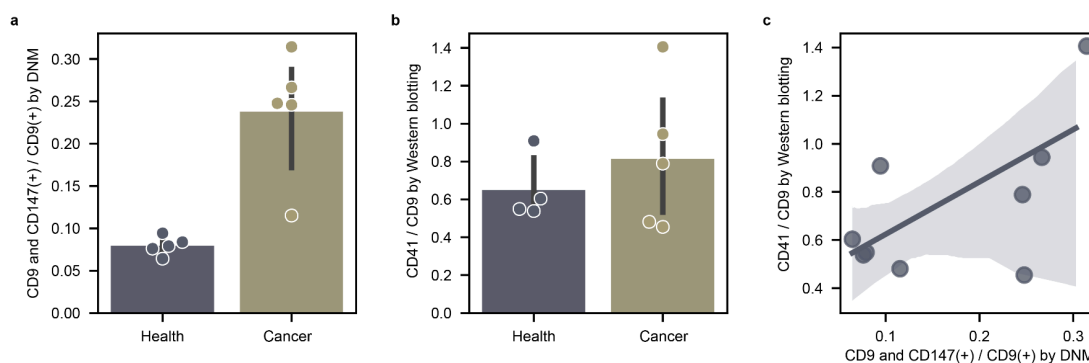
Response (major) 4

We thank the reviewer for their comments. Their attention to detail and observations regarding EV liquid biopsy are important and relevant to our methodology, providing critical insights into the broader field of EV analysis in liquid biopsies.

The influence of platelets

To investigate the influence of EVs from platelets, we performed Western blotting on the same serum-derived EV samples used in our experiments to determine the expression level of CD41, a platelet marker (**Figure R1_5**). Our analysis shows that the ratio of CD41 to CD9, an EV marker, weakly correlates with that of CD9 and CD147 double-positive EVs to CD9 EVs measured by DNM ($R^2 = 0.61$). This result suggests that platelet-derived EVs alone cannot explain CD9 and CD147 double-positive EVs in serum and that at least some are associated with cancer-derived EVs.

Based on this analysis, we added **Figure R1_5** as **Supplementary Figure S13** in the modified manuscript:



Supplementary Figure S13, Figure R1_5 in this response:

The analysis of platelets-derived EV influence in the serum-derived EV samples (a) Bar plots of the proportion of CD9 and CD147 double-positive particles measured by DNM and **(b)** the ratio of CD41 to CD9 by Western blotting. Error bars represent the standard deviations (n=5 for cancer patients, n=5 for healthy controls, biological replicates). **(c)** The relationship between the proportion of CD9 and CD147 double-positive particles to CD9 particles by DNM and the ratio of CD41 to CD9 by Western blotting. A correlation between the x- and y-axes ($R^2 = 0.61$) indicates that platelet-derived EVs alone cannot explain CD9 and CD147 double-positive EVs in serum and that at least some are associated with cancer-derived EVs.

Regarding the CD9+CD147+ particle ratio:

As the reviewer pointed out, lipoproteins constitute the largest fraction of particles in serum. To address this concern, in the original manuscript, we used CD9 as an EV marker and

quantified the number of CD9⁺CD147⁺ particles relative to the number of CD9⁺ particles in the diagnostic demonstration (**Figure 5(d)**). This approach minimizes the impact of lipoproteins on our analysis by focusing solely on EVs.

Considering the analysis above, we have revised the main text in line 426-427 as follows:

“We focused only on the CD9-positive EVs to minimize the impact of lipoproteins on the analysis.”

, and lines 452-458 on pages 14-15 as follows:

“In addition, we assess the potential influence of EVs from platelets by performing Western blotting on the same serum-derived EV samples to measure the expression level of CD41, a known platelet marker. Consequently, our analysis shows that the ratio of CD41 expression-levels to CD9 ones weakly correlates with that of CD9 and CD147 double-positive EVs to CD9-positive EVs measured by DNM ($R^2 = 0.61$) (**Supplementary Figure S12**). This result suggests that platelet-derived EVs alone cannot explain CD9 and CD147 double-positive EVs in serum and that at least some are associated with cancer-derived EVs.”

Major Comment 5

The conclusions are too general, lack quantitative statements and even contains subjective, unscientific statements like: “The ability of DNM to analyse massive numbers of individual particles rapidly and accurately” and “This approach simultaneously demonstrates notable sensitivity, throughput, and scalability, thereby enabling the rapid analysis of rare nanoparticles”. Seven value judgements in two sentences! Given the impressive throughput, sensitivity, sensitivity improvements owing to machine learning, and potentially interesting clinical findings, I highly recommend basing the conclusions on the data and emphasize the key outcomes to substantiate the conclusions. Please remove all value judgements.

Response (major) 5

We appreciate the reviewer's critical but constructive feedback on our manuscript's conclusions. We have thoroughly reviewed and revised the conclusion section to ensure it is based on quantitative data and critical outcomes. We have removed subjective statements and value judgments. Instead, we have emphasized the quantitative results and necessary outcomes, such as the throughput and sensitivity improvements observed in our experiments. We also added the current limitation of our technique and future direction and the potential clinical implications of our development. The revised main text details these revisions in the line 473-506 on page 17 as follows:

“In summary, we developed a technique termed Deep Nanometry, which combines unsupervised deep learning for noise reduction with optofluidic technologies specialized for nanoparticle detection. DNM demonstrated the sensitive detection of 30 nm polystyrene nanoparticles, achieving a high throughput detection of nearly 100,000 events per second and scalability of over a million event analyses. The scalable measurement of DNM enabled the purification-free detection of rare CD9 and CD147 double-positive EVs that account for 0.002% of all confident detections in body fluids; it also enabled the cancer diagnostic detection of CD9 and CD147 double-positive EVs among CD9 positive EVs, accounting for only 0.93% and 0.17% of the cancer patients and healthy control serum, respectively, with significance ($p < 0.05$).

We foresee that the unsupervised 1D denoising technique we have developed can impact a wide range of technological fields where detecting weak 1D signals is crucial. For instance, reducing instrument-

or environment-specific background noise is highly beneficial for flow cytometers in general, from the traditional intensity-based to the image information-based⁵¹. Moreover, there is an increasing need to integrate new data modalities, such as Raman spectroscopy⁵² and refractive index measurements, with high-throughput measurements to offer deeper insights into the molecular compositions and structures of the particles. The significance of denoising will be further emphasized as we advance in such measurement technology, given that the signals derived from these modalities are inherently weak.

An important milestone for future work will be developing a method capable of additionally removing signal-dependent shot noise. While in this work, the sensitivity was significantly improved by removing only the signal-independent noise as it was dominant in the scattered signal, we also observed that the signal-dependent noise in fluorescence data, for example, was too intense to ignore. To this end, a promising solution could be to extend another version⁵³ of our denoising approach, originally developed for removing signal-dependent noise in images.

The sensitive, rapid, and scalable DNM may find diagnostic and therapeutic applications such as the multiparametric detections of EV-based biomarkers and the development and quality assurance of viral or drug delivery particles in pharmaceuticals and vaccines. Similarly, in research, DNM may provide a comprehensive and high-resolution understanding of EV populations in body fluids, opening new avenues for dissecting and deciphering EV heterogeneity and the clinical significance of rare EVs.”

“[51] Ugawa, M. *et al.* In silico-labeled ghost cytometry. *Elife* **10**, (2021).

[52] Lindquist, N. C., de Albuquerque, C. D. L., Sobral-Filho, R. G., Paci, I. & Brolo, A. G. High-speed imaging of surface-enhanced Raman scattering fluctuations from individual nanoparticles. *Nature Nanotechnology* 2019 14:10 **14**, 981–987 (2019)

[53] Salmon, B. & Krull, A. Unsupervised Denoising for Signal-Dependent and Row-Correlated Imaging Noise. (2023).”

We hope these revisions address the reviewer's concerns and enhance our conclusions' clarity and scientific rigor.

=====

Minor Comment 1

At Pg. 7 the authors write: “(e) PR curve for evaluating the peak detection accuracy”. However, figure 2e shows the “Precision” versus Recall. Measurement accuracy and precision are different things. Please select the right term and correct the manuscript accordingly.

Response (minor) 1

We thank the reviewer for their careful review and for pointing out the issue with the terminology in our manuscript. We agree with the reviewer that measurement accuracy and precision are distinct concepts, and we appreciate the opportunity to correct this oversight.

We have revised the caption of **Figure 2(e)** to correctly describe the precision-recall (PR) curve in line 249-251 on page 9 as follows:

“(e) PR curve for evaluating peak detection precision and recall. The colors indicate the time series generated before filtering, after the optimized Gaussian filter, and after denoising. The mean Average Precision (mAP) is the area under the PR curve.”

Minor Comment 2

Regarding the PR curve, did the authors verify that without ND filter in the attenuated detection path, the mAP was approximately 1?

Response (minor) 2

We thank the reviewer for the question. To address this comment, we performed an additional verification experiment, wherein we removed the ND filter from the attenuated detection path and re-evaluated the mAP between the two paths. The results show that, in the absence of the ND filter, the mAP is approximately 1, verifying our measurement system and analysis method (**Figure R1_6**).

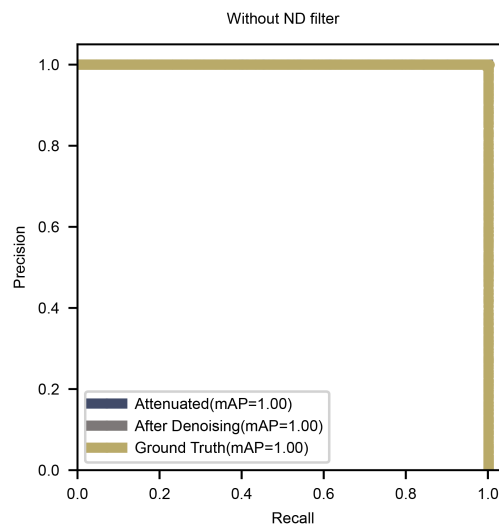


Figure R1_6: The PR-curve obtained when we removed the ND filter in the attenuated detection path. The mAP is the area under the PR curve.

Minor Comment 3

At Pg. 9 the authors introduce the use of a commercially available single-photon counting flow cytometer (SP-FCM). Please add the product type and manufacturer of the SP-FCM to the manuscript to allow improved data interpretation and reproducibility.

Response (minor) 3

We thank the reviewer for the suggestion. Accordingly, we have updated the main text to include the product type and manufacturer of the single-photon counting flow cytometer (SP-FCM) in lines 313-315 on page 11 as follows:

“As a reference, we used a commercially available single-photon counting flow cytometer (SP-FCM), the NanoAnalyzer (NanoFCM Inc.), which exhibits high scattering and fluorescence sensitivities.”

Minor Comment 4

For all figure legends: please add all used acronyms to the legends to improve quick interpretation of the figures.

Response (minor) 4

We thank the reviewer for the suggestion and apologize for this. In response, we have revised all figure legends to include the relevant acronyms, enhancing the clarity and readability of our figures.

Minor Comment 5

At Pg. 9, the authors write: "In our method, the temporal width of each detected scattering peak is 1 μ s, theoretically allowing a throughput of up to 1,000,000 events per second, assuming equal distances between particles". According to the supplementary materials, the authors are aware that in flow cytometry-based instruments, particles arrive at random times in the interrogation volume, and therefore, the theoretical throughput without an acceptable number of multiplets is not the inverse of the acquisition time per particle. The claim of the authors is therefore misleading and should be omitted or rephrased.

Response (minor) 5

We thank the reviewer's comment and understand the concern. The simplified phrase (theoretically allowing a throughput of up to 1,000,000 events per second, assuming equal distances between particles) is often used in the field of imaging flow cytometry^{1,2,3}. The field, including us, is aware that in experiments particles arrive randomly, but estimates the theoretically possible maximum throughput assuming equal distances between particles.

Nevertheless, we completely agree that the phrase "theoretical throughput" is a vague expression without a detailed explanation. Thus, for clarity, we have added a more detailed experimental condition to the main text in lines 315-321 of page 11 as follows:

"In our method, the temporal width of each detected scattering peak is approximately 1 μ s, theoretically allowing a maximum throughput of up to 1,000,000 events per second, assuming equal distances between particles. When we dilute the particle solution down to a concentration where the particles arrive every 10 μ s on average, considering the random arrival times of particles and the need to minimize coincident events, our system can detect up to 100,000 events per second with minimal aborts (Figure 3(a))."

References

1. Diebold, Eric D., et al. "Digitally synthesized beat frequency multiplexing for sub-millisecond fluorescence microscopy." *Nature Photonics* 7.10 (2013): 806-810.
2. Han, Yuanyuan, et al. "Imaging technologies for flow cytometry." *Lab on a Chip* 16.24 (2016): 4639-4647.
3. Han, Yuanyuan, and Yu-Hwa Lo. "Imaging cells in flow cytometer using spatial-temporal transformation." *Scientific reports* 5.1 (2015): 13267.

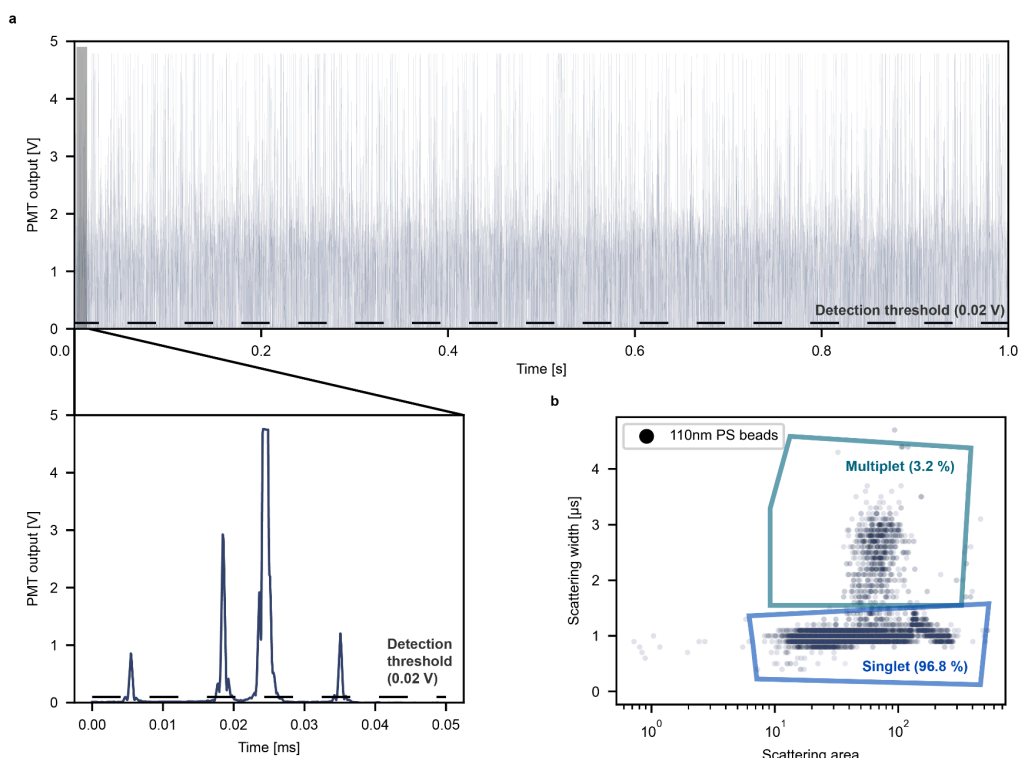
Minor Comment 6

At Pg. 9, the authors write: "The throughput we experimentally record is 96,000 events/second". Congratulations, because this is a higher throughput than the typical operational throughputs of standard flow cytometers. However, the authors do not provide data based upon which the reader can confirm this claim. Please substantiate this claim with data and indicate the fraction of multiplets (e.g. doublets, triplets) that were present during this experiment.

Response (minor) 6

We thank the reviewer for their valuable comment highlighting the need to substantiate our throughput claim with supporting data.

To claim DNM's throughput, we again measured the scattering signal of 110 ± 8.7 nm particles¹ (Thermofisher, F8803) throughout 1 second. In **Figure R1_7(a)**, we show the scattering time series, wherein we identified the singlet and multiplet detections by gating the plot of scattering width and area, as is commonly done in conventional flow cytometers. Here we call the detection of individual particles a “singlet” detection and the simultaneous detection of two or more particles or aggregates of particles a “multiplet” detection. As a result, this analysis shows that the DNM system detected 241,510 events for one second, of which 96.8% were singlets and 0.6 % were multiplets.



Supplementary Figure S1, Figure R1_7 in this response:

Detection of 110nm polystyrene (PS) beads by DNM (a) Time series of scattering signals of the PS beads measured for 1 second. The zoomed-up panel shows the time series scattering signal of the beads measured for 0.05 ms. (b) The scatter plot of scattering width and scattering area of the beads, obtained by applying a constant trigger threshold to the time series. In this plot, we identified the singlet and multiplet detections by gating the plot of scattering width and area, where we call the detection of individual particles a “singlet” detection and the simultaneous detection of two or more particles or aggregates of particles a “multiplet” detection. From this analysis, we obtained a detection throughput of the DNM as 241,510 events for one second, of which 96.8% were singlets and 0.6 % were multiplets.

Also, we have added **Supplementary Figure S1**, and revised the main text in lines 322-324 on page 11 as follows:

“Experimentally, we achieved a throughput of 241,510 events per second by flowing a suspension of 110 nm PS beads and gating singlet events, which accounted for 96.8% of the total events in the

scattering width versus area plot (**Supplementary Figure S1**). Note that this detection throughput varies depending on the input concentration of particles."

We appreciate the reviewer's feedback, which has helped us provide more detailed data to support our claim.

References

1. <https://www.thermofisher.com/order/catalog/product/jp/ja/F8803>

Minor Comment 7

At Pg. 9, the authors write: "we conclude that the samples are detectable at concentrations of at least 2.82×10^3 particles/mL". Why did the authors not dilute further than 10^{-7} -fold?

Response (minor) 7

We thank the reviewer for their comment.

We recognize that if samples are measured at a dilution rate of 10^{-7} or more, our system can confirm its further superiority by demonstrating even lower detection limits. However, we opted not to perform further dilutions as it requires significant measurement time to detect a number of particles at such a low concentration.

In detail, given the sample flow rate of 1 $\mu\text{L}/\text{min}$, detecting 100 particles for reliable statistics from a concentration of 2.82×10^2 particles/mL would require approximately 355 minutes (nearly 6 hours) per sample for DNM. For the state of the art single-photon counting flow cytometer, the NanoAnalyzer from NanoFCM Inc., it takes 1848 hours and, in fact, does not permit measurements longer than minutes. We therefore decided to limit the dilution to 10^{-7} -fold.

If the reviewer requests a detailed explanation in the manuscript or additional experiments, we can do it accordingly. We hope the reviewer understands the situation.

Minor Comment 8

Figure 4b is unnecessarily difficult to interpret and require improvement. The vertical axis of figure 4b shows the counts per minute. However, the reader is not aware of the flow rate and even if the flow rate would be defined, the reader needs to calculate how the counts relate to the concentration. Flow cytometer-based instruments have the (quite) unique capability to determine absolute particle concentrations. To improve data interpretation and reproducibility, therefore please report particle concentrations whenever possible.

Response (minor) 8

We appreciate the reviewer's comment. We agree that it is difficult for readers to translate the vertical axis of **Figure 4(b)** in the original manuscript counts per minute, to particle concentrations. In response, we have revised **Figure 4(b)**(**Figure R1_8**), to show the concentration of CD9⁺ particles in y-axis. The concentrations were calculated with the throughput of particle detections and the flow rate specified in **Supplementary Figure S9(a)** (**Figure R1_9**) that we also added in this revision. Accordingly, we revised the relevant main text and caption in lines 337-338 of page 11 and in **Supplementary Figure S9**, respectively.

“Among the detections defined by the scattering and fluorescence intensities, we assessed the counts and concentration of CD9-positive EVs in DNM and observe that they are proportional to the dilution rate of 10^{-5} (**Figure 4(b)** and **Supplementary Figure S9**).”

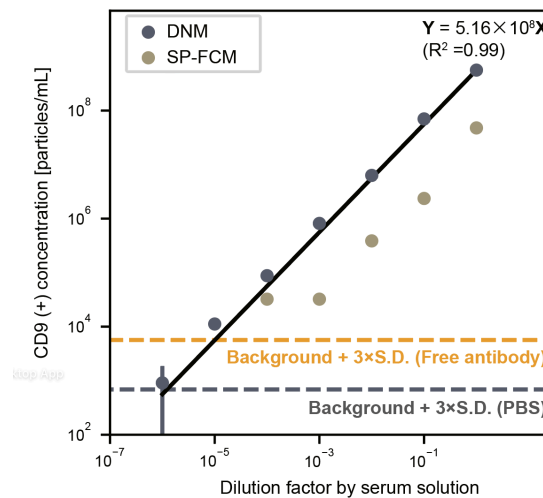
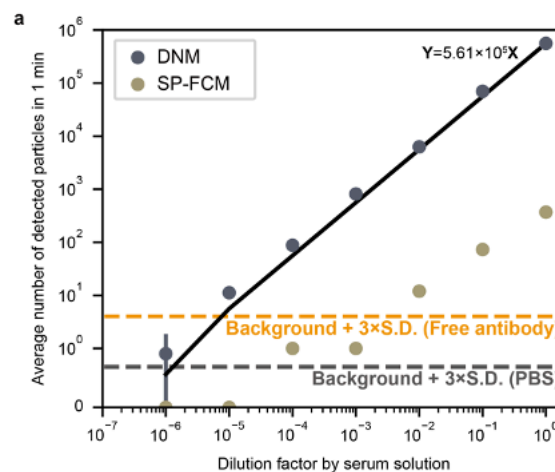


Figure 4(b), Figure R1_8 in this response:

Scatter plot of the concentration of detected CD9 positive EVs and a factor of diluting them with unstained serum (10^{-1} to 10^{-6}). The y-axis is a log scale. The blue and olive points indicate the results of 4-minute DNM and 1-minute SP-FCM measurements, respectively. The black line indicates the linear regression result of the DNM measurements ($n = 3$, mean $\pm$ S.D., technical replicate). The black and orange dashed horizontal lines represent



Supplementary Figure S9(a), Figure R1_9 in this response:

(a) Scatter plot of the average detection throughput of CD9(+) EVs and the factor of diluting them with unstained serum (10^{-1} to 10^{-6}). The y-axis is symmetric log scale. The blue and olive colors indicate the result of 3-minute DNM and 1-minute SP-FCM measurements, respectively. The black line indicates the linear regression result produced for DNM measurements ($n=3$, technical replicate). The black and orange dashed horizontal lines represent the PBS background mean + $3 \times$ S.D. and the free antibody background mean + $3 \times$ S.D., respectively: Limit of detection values for DNM. Sample flow rate of DNM is 1 μ L / min and that of SP-FCM is 3.2 nL/min.

On the other hand, we would like to emphasize that the absolute count of detected particles is important for reliable data assessment. According to our simulation, fewer than 23 total detections are below the limit of quantitation (**Supplementary Figure S15**). In **Figure 5(e)**, NanoFCM's insufficient counts of CD9 and CD147 double-positive EVs render the calculated positivity rate unreliable, making it difficult to distinguish between healthy controls and cancer patients.

Minor Comment 9

The previous minor comment also applies to panel figure 4c, figure 5 and figure 6. It should be a sin to report percentage fractions of flow cytometry data without showing the absolute concentrations. Reported fractions, as shown in figure 4c are very difficult to reproduce due to differences between detection limits of flow cytometers and possible systematic errors of the derived particle concentrations. Moreover, the fraction of CD9+ particles in serum seems an irrelevant measure to me because the total concentration of particles in serum is dominated by lipoproteins. In turn, the concentration of lipoproteins strongly depends on the diet and time between the last meal and venipuncture. Hence, does the ratio of CD9+ particles to all particles tell something about EVs or rather something about the diet and energy balance of the donor? I strongly recommend plotting the concentration (/mL) of CD9+ particles on the vertical axis.

Response (minor) 9

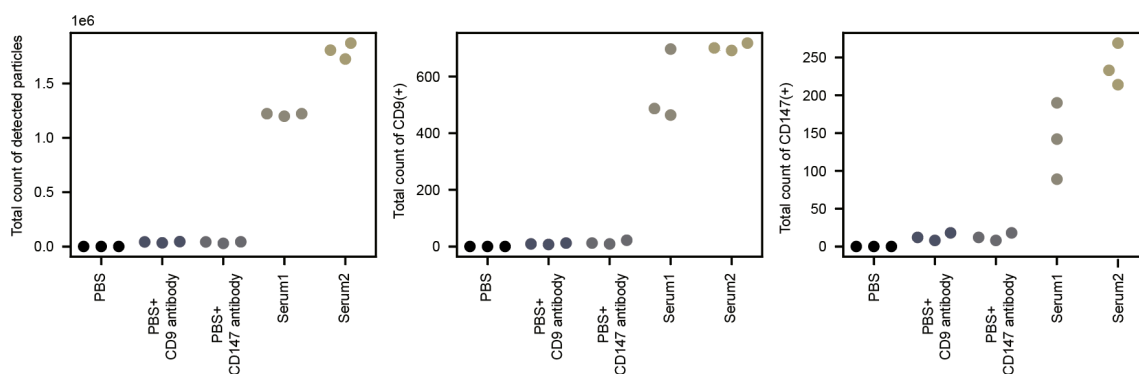
We thank the reviewer for their careful review and valuable comments regarding the presentation of our flow cytometry data. In response, we first discuss each Figure and its relevant revisions, and then discuss the fraction of CD9+ particles in serum.

Revision of each Figure

Regarding **Figure 4(c)**, which presents the CD9 positivity rates based on data shown in Figure 4(b), as we have already revised **Figure 4(b)** to show the concentration of CD9+ particles (/mL) on the vertical axis in **Response (minor) 8**, we believe that further modification on **Figure 4(c)** is unnecessary. To avoid potential confusion, we revised the caption of **Figure 4(c)** in lines 370-371 of page 13 to clearly state that **Figure 4(c)** is derived from the same dataset as **Figure 4(b)**:

“(c) Scatter plot of the average percentage of CD9 positive particles in the unstained serum and the dilution factor (dilution from 10^{-1} - 10^{-6}), obtained from the same data used in **Figure 4(b)**. ”

Regarding **Figure 5**, we originally reported the absolute concentrations of CD9+CD147+ particles (/mL) in **Figure 5b**, followed by the percentage fractions in **Figure 5c**. To further improve clarity, we have added the flow rate ($\mu\text{L}/\text{min}$) and the particle detection throughput (particle detections/min) in **Supplementary Figure S12 (Figure R1_10)** as follows, allowing readers to calculate the particle concentrations (/mL) independently.



Supplementary Figure S12(c), Figure R1_10 in this response:

(c) The total count of particles, CD9(+) particles, and CD147(+) particles detected from control samples and non-purified serum, obtained by DNM measurement. Sample flow rate of DNM is 1 $\mu\text{L}/\text{min}$.

Following the reviewer's suggestion, we modified **Figure 6 (Figure R1_11)** to show both the the number of detected particles (particle detections/min) in **Figure 6(b)** and the corresponding concentration of detected particles (particles/mL) in **Figure 6(c)**. To further improve clarity, we have also added the flow rate ($\mu\text{L}/\text{min}$) in the caption of in **Figures 6(c) and 6(d)** as follows, providing the necessary data for concentration calculations.

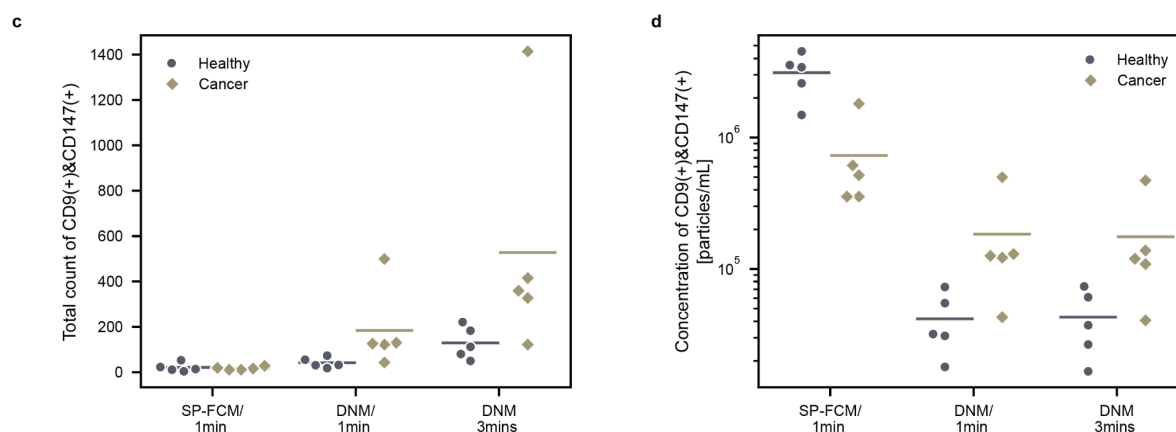


Figure 6(c) and 6(d), Figure R1_11 in this response:

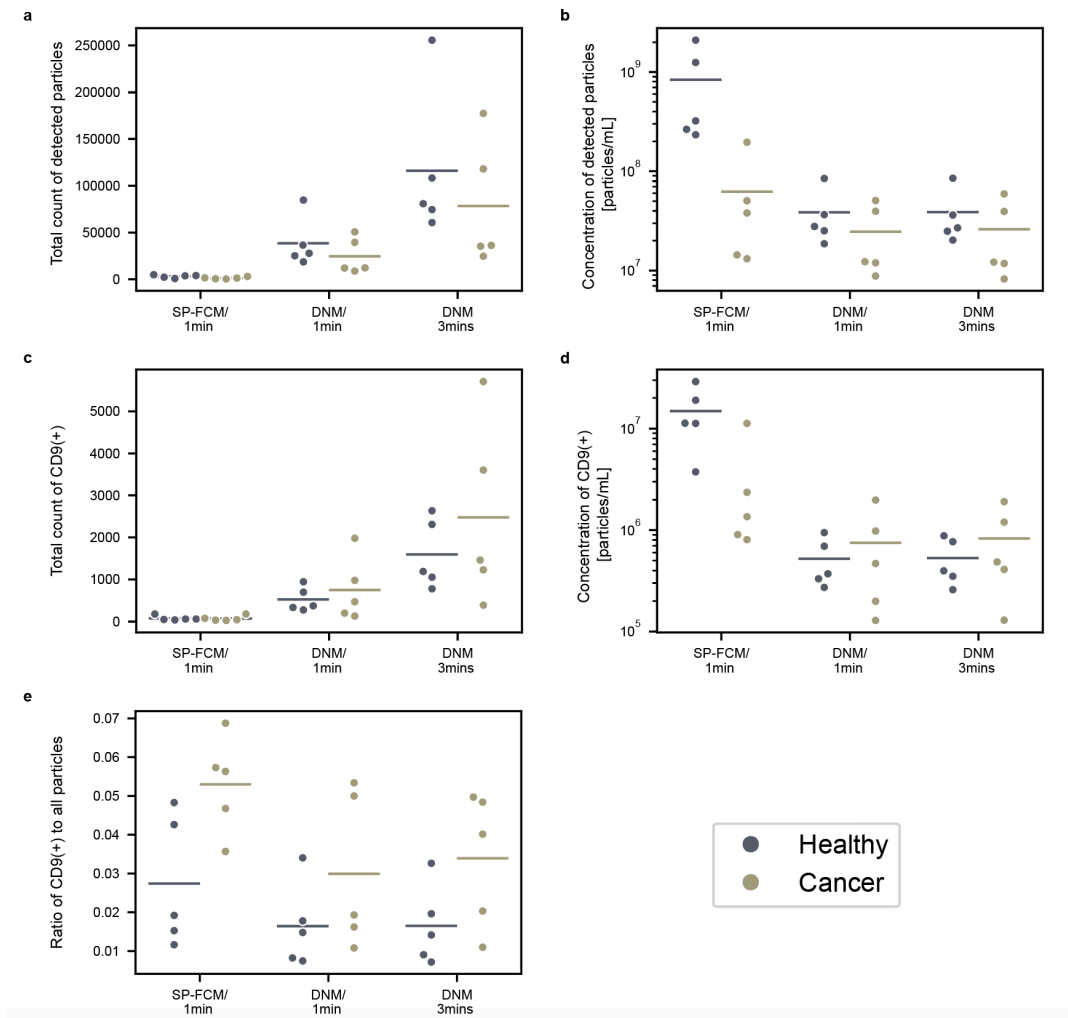
(c) The total count and (d) the concentration of CD9- and CD147-positive EVs from cancer patients and healthy controls, obtained with the SP-FCM measurement for 1 minute and DNM measurement for 1 minute and 3 minutes, respectively. The bar shows mean ($n = 5$ for cancer patients, $n = 5$ for healthy controls, biological replicates). Sample flow rate of DNM is 1 $\mu\text{L}/\text{min}$ and that of SP-FCM is 3.2 nL/min.

Discussion on the fraction of CD9+ particles in serum

To clarify, the materials used and the intent of each figure is as follows:

- **Figures 4(b) and 4(c)** analyze EVs from cultured cell supernatants.
- **Figures 5 and 6** aim to demonstrate the capability of DNM to detect rare particles.
- **Figure 6** further examines the ratio of CD9 and CD147 double-positive particles relative to CD9-positive particles.

Therefore, our primary objective was not to assess CD9+ particle fractions specifically for EV-based diagnostics. However, we agree that investigating CD9+ particle concentrations (/mL), given the predominance of lipoproteins in total particle concentration, could be of interest in understanding aspects of diet and energy balance in donors. To address this, we have added **Supplementary Figure S14 (Figure R1_12)**, which includes both absolute concentrations of CD9+ particles (/mL) and CD9+ particle percentages (relative to total detections). Unfortunately, detailed information on donor diet and energy balance is not available. We hope that this addition provides a more comprehensive view, allowing readers to interpret the data from various perspectives.



Supplementary Figure S14, Figure R1_12 in this response:

Detection of EVs from cancer patients and healthy donors for colorectal cancer diagnosis. We performed SP-FCM measurement for 1 minute and DNM measurement for 1 minute and 3 minutes, respectively for each sample (n=5 for cancer patients, n=5 for healthy controls, biological replicates). Plots of **(a)** the total count of detected particles, **(b)** the concentration of detected particles (/mL), **(c)** the total count of CD9-positive EVs, **(d)** the concentration of CD9-positive EVs (/mL), and **(e)** the ratio of CD9-positive particles in all detected particles.

[Minor Comment 10](#)

In figure 4c, the ratio of CD9+ particles to all particles is a sample property and should therefore be independent of the dilution of the sample. Why is the fraction of CD9+ particles affected by the dilution of the sample?

Response (minor) 10

We thank the reviewer for pointing out this with the careful review of our work. We apologize for the unclear description in the manuscript.

The reviewer is correct that, if the diluent does not contain particles, the ratio of CD9+ particles to all particles should be independent of the sample dilution, as it is an intrinsic property of the sample.

However, in our experiment, the dilution process involved adding additional serum to the sample rather than diluting it with a buffer. This means that as we diluted the sample, we introduced more serum-derived particles, such as lipoproteins, into the mixture. Consequently, the proportion of CD9+ particles decreased with increasing dilution due to the increased presence of particles that do not express CD9.

To address this issue, we have revised the caption to clarify the dilution procedure and its impact on the ratio of CD9+ particles. The revised text now explicitly states that the dilution was performed using serum, which alters the particle composition and affects the fraction of CD9+ particles. We have also added an explanation in the caption of **Figure 4(c)** (in lines 370-374 of page 13) to clarify why the fraction of CD9+ particles decreases with dilution.

“(c) Scatter plot of the average percentage of CD9 positive particles in the unstained serum and the dilution factor (dilution from 10^{-1} - 10^{-6}). As dilution increases, the proportion of CD9-positive particles decreases due to the increased presence of particles in the unstained serum. The scale of the y-axis is a symmetric log scale: the linear scale is from 0 to 0.1 and the log scale starts from 0.1. The blue and olive points indicate the results of the DNM and SP-FCM measurements, respectively.”

Minor Comment 11

At Pg. 9 the authors write “The detection limit derived from blank PBS (background level of $+3\sigma$ for 4-minute measurements) is 0.01 detections/minute or 6.8×10^2 particles/mL”. The authors used the buffer-only control to establish the concentration detection limit of their instrument. However, when considering the entire assay, the antibodies in buffer control would be a more suitable measure of the concentration detection limit. Please also indicate the background $+ 3 \times$ S.D. of the antibodies in buffer control in Figure 4b.

Response (minor) 11

We appreciate the reviewer’s comment about the choice of the control sample to establish the concentration detection limit. We have conducted additional measurements using the antibodies in buffer control to provide a measure of the detection limit. The measurement yielded a detection limit of 0.26 detections per minute, equivalent to 4.0×10^3 particles/mL. Following the reviewer's suggestion, we have updated **Figure 4(b)** (**Figure R1_8**) to reflect the background $+ 3\sigma$ of the antibodies in buffer control and revised the main text in line 337-345 on pages 11-12 to reflect this updated value as follows:

“Among the detections defined by the scattering and fluorescence intensities, we assessed the counts and concentration of CD9-positive EVs in DNM and observe that they are proportional to the dilution rate of 10^{-5} (**Figure 4(b)** and **Supplementary Figure S9**). This dilution rate is equal to the CD9-positive concentration is 1.1×10^4 particles/mL and the detection rate of 11 particles/min; the

CD9-positive rate is estimated to be 0.04% (**Figure 4(c)**). The detection limit derived from the antibodies in buffer control (background level of $+3\sigma$ for 10-minute measurements) is 4 detections per minute or 4.0×10^3 particles/mL. The detection limit derived from blank PBS (background level of $+3\sigma$ for 10-minute measurements) is 0.68 detections/minute or 6.8×10^2 particles/mL. ”

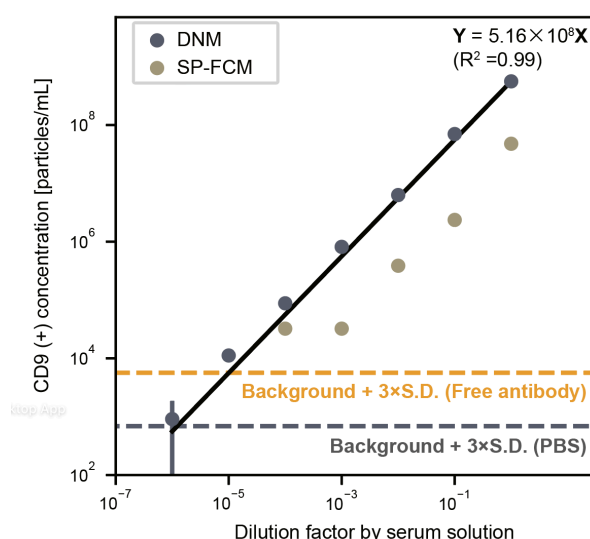


Figure 4(b), Figure R1_8 in this response. The same figure used in **Response (minor) 8**:

Scatter plot of the concentration of detected CD9 positive EVs and a factor of diluting them with unstained serum (10^{-1} to 10^{-6}). The y-axis is a log scale. The blue and olive points indicate the results of 4-minute DNM and 1-minute SP-FCM measurements, respectively. The black line indicates the linear regression result of the DNM measurements ($n = 3$, mean $\pm$ S.D., technical replicate). The black and orange dashed horizontal lines represent

Moreover, we have added the explanation of how we prepared antibodies in buffer control to the method section in lines 586-588 on pages 19 as follows:

“After antibodies was added to a PBS at a concentration used for labeling particles, the same centrifugation and filtering procedure was applied to this solution to prepare the antibodies in buffer control.”

Minor Comment 12

At Pg. 11 the authors claim: “To our knowledge, this is the first study to count EVs in unpurified serum at single-particle resolution accurately”. However, in the Methods the authors write: “The supernatant was ultracentrifuged to prepare EVs at 49,500 rpm (210,000 $\times g$) using a TLS-55 rotor for 35 min at 4°C. The pellets were washed with 2 mL PBS by ultracentrifugation at 49,500 rpm (210,000 $\times g$) using a TLS-55 rotor for 35 min at 4°C and resuspended in PBS”. The latter implies that EVs were purified by ultracentrifugation and washing. Please clarify. In addition, there are many examples in literature where EVs from plasma or serum were directly measured in stained diluted plasma and serum. The main difference between those studies and the work of the authors, however, is that the smallest detectable EVs in the published studies are generally >100 nm.

Reponse (minor) 12

We thank the reviewer for pointing out the inconsistency between our claim and the methods described in our manuscript. We agree that this clarification is crucial for enhancing the clarity and reproducibility of our work.

To clarify, our study involves two distinct experiments using serum samples:

- **In Figure 5**, we analyzed EVs **contained in unpurified serum** that had not undergone ultracentrifugation or antibody removal column.
- **In Figure 6**, we analyzed EVs **purified from serum** using ultracentrifugation and an antibody removal column.

We have revised the Methods section to clearly differentiate between these two experimental approaches. The revised main text in the line 570-572 on page 19 now reads:

“Extracellular vesicles contained in unpurified serum

All serum samples from healthy donors were purchased from BizCom Japan (Tokyo, Japan). Serum samples were centrifuged at $10,000 \times g$ for 10 min at 4 °C to remove cells and debris."

and in the line 573-578 on the page 19:

“Extracellular vesicles purified from serum

All serum samples from healthy donors and colorectal cancer patients were purchased from BizCom Japan (Tokyo, Japan). Serum samples were centrifuged at $10,000 \times g$ for 10 min at 4°C to remove cells and debris. The supernatant was ultracentrifuged to prepare EVs at 49,500 rpm ($210,000 \times g$) using a TLS-55 rotor for 35 min at 4°C. The pellets were washed with 2 mL PBS by ultracentrifugation at 49,500 rpm ($210,000 \times g$) using a TLS-55 rotor for 35 min at 4°C and resuspended in PBS."

We have also corrected our claim regarding the detection of EVs contained in unpurified serum. We acknowledge that there are previous studies where EVs larger than 100 nm were directly measured in stained plasma and serum, detecting EVs. Our study extends this by accurately counting EVs in unpurified serum at single-particle resolution down to sizes much smaller than 100 nm.

The revised statement in the main text in the lines 393-394 on page 14 now reads:

"To our knowledge, this is the first study to accurately count EVs smaller than 100 nm in unpurified serum at single-particle resolution."

Minor Comment 13

At Pg. 12, the authors write: "The EV samples are diluted 50-fold with PBS as a recommended optimal condition for the reference SP-FCM". The interrogation volume and particle velocity differ between the SP-FCM and DNM. The optimal dilution, which is the dilution with the maximum count rate without having swarm detection, therefore might differ between the SP-FCM and DNM. Generally, it is recommended to perform a dilution series to verify that after log-transformation, the slope between the measured concentration and dilution factor is -1 (See: A compendium of single extracellular vesicle flow cytometry, Welsh et al., J. Extracell. Vesicles 2023). Why the authors are sure that 50-fold dilution is sufficient?

Response (minor) 13

We thank the reviewer for this valuable comment.

We recognize that the optimal dilution may differ between the SP-FCM and DNM due to their differing interrogation volumes and particle velocities. However, for consistency in our comparative analysis of different instruments, we used the same 50-fold dilution factor for both systems.

The choice of the 50-fold dilution for the EV samples was made empirically, ensuring reliable detection within the measurement range of both SP-FCM and DNM. Considering the limited sample volume for the donor-derived serum, this dilution factor was not based on a dilution series analysis to verify that the slope between the measured concentration and dilution factor is -1, as suggested by the reviewer. Instead, we selected the 50-fold dilution to balance between avoiding possible issues in SP-FCM such as swarm detection (from overly concentrated samples) and insufficient event counts (from overly diluted samples). We would like to note that DNM detected 526 events per second, which are orders of magnitude lower than the 100,000 events per second, the maximum detection throughput of DNM.

To avoid any misunderstanding, we have revised the main text in lines 427-430 on page 15 as follows :

“The EV samples are diluted 50-fold with PBS as an empirical optimal condition for the reference SP-FCM to avoid swarm detection and we used the same dilution factor for consistency in our comparative analysis.”

We appreciate the reviewer's question, which has helped us to clarify the rationale behind our selection of the dilution factor and to avoid any potential confusion.

Minor Comment 14

I found two minor typos in the figures and equations:

- a. S8a: measurents--> measurements
- b. Extended figure 1a: traning --> training
- c. Equation 1 at Pg. 17: mean squre error --> mean square error

Response (minor) 14

We thank the reviewer for identifying these minor typos in our figures and equations and apologize for these. We have corrected them accordingly.

Minor Comment 15

At Pg. 16, the authors write: “To thoroughly remove cellular debris, the supernatant was filtered through a 0.22-µm filter (Millipore)”. Please specify the used filter. Please notice that if the authors used a PES filter, that the pore diameter of the filter is poorly defined and has little to do with “0.22 µm”. Please check manuscripts with scanning electron microscopy images of PES filters.

Response (minor) 15

We thank the reviewer for their sharp comment regarding the filter specifications used in our study.

The reviewer is correct that the pore size of filters, particularly polyethersulfone (PES) filters, can be variable, and that simply describing them as “0.22 µm” may be misleading. We agree

that this could cause confusion and acknowledge that filtration does not result in perfect size exclusion. We have revised the main text to include the specific product name and material of the filter used to improve the transparency and reproducibility of our experimental methods, and removed the specification about filter size to avoid unnecessary confusion. WE The updated sentence in the line 564-566 on the page 19 as follows:

“After incubation for 48 h, the CM was collected and centrifuged at 2,000 ×g for 10 min at 4°C. To thoroughly remove cellular debris, the supernatant was filtered through a polyethersulfone membrane filter (Millipore, Stericup Quick Release-GP Sterile Vacuum Filtration System).”

We again appreciate the reviewer's helpful suggestion.

Minor Comment 16

Please specify the antibody concentrations that were used during labeling. In addition, please make clear whether the authors used a procedure to remove possible aggregates from the stock of antibodies.

Response (minor) 16

We thank the reviewer for pointing out the need to specify the antibody concentrations used during labeling and to clarify our procedures regarding potential antibody aggregates.

The antibody concentration used during labeling was 1 mg/mL. We did not perform any specific procedures to remove potential aggregates from the antibody stock before use. To address the comment, we have updated the Methods section of the main text in the line 580-583 on page 19 to include this information:

“To analyze EVs by DNM, EVs derived from CM or serum were incubated with anti-human CD9 antibodies-conjugated Tide Fluor™5WS (Cosmo Bio, clone: 12A12) and anti-human CD147 antibodies-conjugated Alexa488 (abcam, clone: MEM-M6/1) for 2 hours at room temperature in the dark. The antibody concentration used during labeling was 1 mg/mL.”

Minor Comment 17

At pg. 16 the authors write: “To remove the unbound antibodies, EVs were then transferred to ultrafiltration spin-columns (300 kDa cut off, APRO Science) and centrifuged at 14,000 g at room temperature, followed by washing twice with PBS.” If I understand this procedure well, unbound antibodies slip through the filter whereas EVs stay behind. Please make clear how you release the EVs from the filter membrane.

Response (minor) 17

We appreciate the reviewer's comment regarding the recovery of EVs from the ultrafiltration spin-columns.

As the reviewer pointed out, unbound antibodies pass through the filter membrane while EVs are retained. To release the EVs from the filter membrane, we carefully pipetted the retained volume multiple times, ensuring that the EVs were thoroughly resuspended and collected into solution.

We have updated the Methods section of the main text to clarify this procedure. The revised sentences in the line 583-586 on the page 19 as follows:

“To remove the unbound antibodies, EVs were transferred to ultrafiltration spin-columns (300 kDa cutoff, APRO Science) and centrifuged at $14,000 \times g$ at room temperature, followed by washing twice with PBS. The EVs were then recovered from the filter membrane by carefully pipetting the retained volume multiple times to resuspend and collect the EVs into solution.”

Minor Comment 18

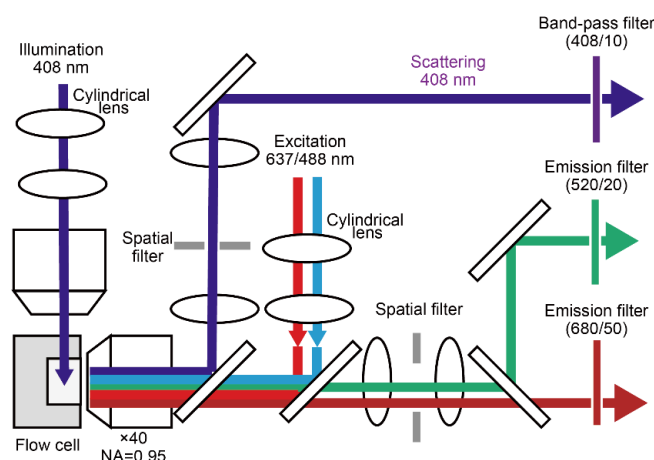
In fig. S5a, why the lasers for fluorescence excitation are not elliptically focused?

Response (minor) 18

We thank the reviewer for highlighting the discrepancy regarding the beam shape in **Figure S5(a)**.

We apologize for any confusion caused by the original figure. In our experimental setup, the excitation lasers are indeed elliptically focused to create a sheet-like illumination at the focal plane. This is achieved by introducing a cylindrical lens into the optical path. To address the reviewer's comment, we have updated **Supplementary Figure S8(a)(Figure R1_13)** which is corresponding to **Figure S5(a)** in the original manuscript, to include the cylindrical lens, clearly illustrating the elliptical focusing of the lasers. Additionally, we have revised the Methods section of the main text to reflect this detail. The revised sentences in the line 518-520 on the page 18 as follows:

“All excitation lasers were elliptically focused using a cylindrical lens to generate a sheet-like beam profile at the focal plane with typical dimensions of $2 \mu\text{m} \times 5 \mu\text{m}$, $2 \mu\text{m} \times 10 \mu\text{m}$ and $2 \mu\text{m} \times 5 \mu\text{m}$ for the 408-, 488- and 637-nm laser beams, respectively.”



Supplementary Figure S8 (a), Figure R1_13 in this response:
Optical configuration for EV detection.

=====

We have revised the manuscript accordingly and believe that these changes effectively address the reviewer's concerns. It is hoped that the responses provided are satisfactory, and we look forward to further evaluation.

We thank once again for the time and effort dedicated to reviewing our work.

Best regards,
Authors

Reviewer #2:

Overall comment

The paper presents a new technique termed DNM combining an unsupervised signal denoising algorithm with SP-FCM and demonstrating enhanced sensitivity in detecting and measuring various physical characteristics of protein biomarkers in serum compared to conventional SP-FCM. Despite existing neural network methods that have been proposed for single-photon flow data analysis (classification) and for imaging flow image denoising, there is now still a lack of effective one-dimensional flow data denoising. The author modifies and validates an unsupervised learning(neural network-based) signal denoising method for this field, which constitutes a major contribution. We appreciate predictably potential biomedicine applications of DNM, especially noninvasive detection and assessment of early cancer. Still, I have concerns about the algorithms and readability of the paper.

Response to overall comment

We sincerely appreciate the thorough review of our manuscript and the recognition of the novelty and potential applications of our technique, Deep Nanometry (DNM). We are also grateful for the acknowledgment of our contribution to developing an unsupervised learning-based signal-denoising method for one-dimensional flow data. The reviewer's insightful comments have helped guide us to enhance the manuscript, especially in providing a more detailed analysis of the algorithms and improving readability.

Major Comment 1

The author first employs a CNN to learn and model the probability distribution of background noise data collected by imaging particle-free samples like ultrapure water, and then employs a VAE to remove noise with the learned probability distributions from practical imaging data(beads/patient's serum/in control). However, in PMT imaging of practical samples, signal-dependence shot noise (Poisson noise) dominates and must be well considered. However, in the first stage, only signal-independence noise is taken into account. This leads to a problem, which as my core concern, that the probability distribution learned in the first stage is not consistent with that in practice. It can result in inadequate denoising performance and weak generalization. Despite the effectiveness of the proposed algorithm in Results, there is a lack of sufficient comparison with other algorithms, especially with supervised deep learning methods. From an algorithmic perspective, the algorithm described by the author may remove part of shot noise that displays similar characteristics to environment/instrument/sensor noise in imaging of particle-free samples. However, the theoretical gap should be fully and analytically discussed.

Response (major) 1

We appreciate the reviewer's valuable comments and concerns about whether the algorithms can remove the signal-dependent shot noise. We apologize for any confusion caused by our original statement. We have revised the relevant sections of the manuscript to improve the clarity of noise composition, signal-independent and signal-dependent noises, and the effect of noise removal models on each along with additional analysis.

We first would clarify the description of the signal composition and the effect of noise removal models on each. It is overly simplistic to view the measured signal as the sum of a clean signal and signal-independent noise. Instead, the measured signal should be more accurately described as the sum of the signal originating from light scattered by the

particles, which is affected by signal-dependent shot noise, and signal-independent background noise. In our denoising process, we can only remove signal-independent background noise using a noise model pre-trained with data from flowing PBS. On the other hand, signal-dependent noise, such as photon shot noise, is not captured by our noise model and thus cannot be removed. In other words, noise sources that are not present in the PBS, e.g. smaller nanoparticles, or proteins, are not modeled and therefore not removed. We have clarified this with the following text, which is included on lines 94-102 of page 3 of the revised main text:

“Our denoising approach assumes that a measured noisy time series, $x = s + n$, $x \in R^T$, is the sum of the measured signal intensity originating from the light scattered by particles (particle signal), $s \in R^T$, and signal-independent background noise, $n \in R^T$, where T is the time series length. The particle signal s is affected by shot noise with a variance that depends linearly on the unknown underlying clean signal. The background noise n has a fixed variance and is independent of the particle signal³⁰, and we therefore describe it as sampled from probability distribution $p(n)$. Our denoising method removes signal-independent noise to recover the particle signal. Even though it is currently not able to remove the remaining shot noise, we experimentally show in the following section that this approach leads to improved sensitivity.”

“[30] de Rond, L. *et al.* Quantification of Light Scattering Detection Efficiency and Background in Flow Cytometry. *Cytometry* **99**, 671 (2021).”

Furthermore, to experimentally investigate the aforementioned signal composition and effectiveness of our denoising model, we further conducted an additional analysis viewing the variance of the attenuated signal intensity before and after denoising as a function of the corresponding ground-truth signal intensity. To avoid redundancy, we here explain the details of the analysis by including the content we added from line 211 to line 231 on page 6 of the revised main text as it is. The results show that photon-derived shot noise was the main component of residual noise after noise removal.

“Additionally, we analyzed the attenuated signal variance before and after denoising as a function of the ground-truth signal intensity to evaluate how closely DNM approaches the shot noise limit. We consider the shot noise limit to be the state in which there is only noise that depends on the optical signal and whose variance is proportional to the ground-truth signal intensity. In the analysis, we first max-normalized the ground-truth time series, then binned these normalized ground-truth signal intensities into 100 intervals. For each bin, we calculated the variance of the attenuated signal at corresponding time points. We plot the results of this analysis in **Figure 2(f) (R2_1 in this response)**.

Before denoising, the variance of the signal from the attenuated arm increases approximately linearly with ground-truth signal intensity, matching the behavior of shot noise. The y-intercept of the line – the variance of the noise when the ground-truth signal intensity is zero – is approximately equal to the variance of the PBS-derived fluidic noise. Since the variance of the sum of two independent random variables is additive³⁶, this indicates that the noise in the attenuated time series is the sum of the PBS-derived fluidic noise and shot noise. To make this clear, we subtracted the variance of the PBS-derived fluidic noise from that of the attenuated time series and plotted it as the dashed line in **Figure 2(f) (R2_1 in this response)**.

We then repeated this analysis on the variance of the denoised signal. The slope of the variance here is close to the slope before denoising, indicating that the shot noise indeed remains. However, the y-intercept is now close to zero, indicating the removal of the PBS-derived fluidic noise. The curve slightly deviates from the ideal linear shape due to imperfections in the denoising result.”

“[36] Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P. & Meester, L. E. Covariance and correlation. 135–150 (2005) doi:10.1007/1-84628-168-7_10”

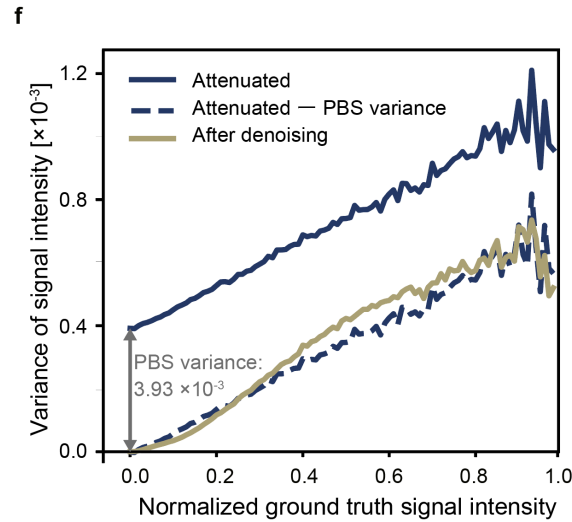


Figure 2(f), Figure R2_1 in this response:

The variance of the attenuated signal before denoising (blue line) and after denoising (olive line) as a function of the corresponding normalized ground truth signal intensity. The dashed blue line is the variance of the noisy attenuated signal minus the variance of signal-independent PBS-derived fluidic noise.

We are also grateful to the reviewer for recommending a comparison to supervised deep learning methods. We present this analysis in **Response (major) 2**.

Major Comment 2

The algorithm proposed is trained and executed in an unsupervised manner. Owing to incomplete overconfident modeling of signals and noise, unsupervised denoising algorithms often display a degenerate performance (validation PSNR metrics for example) not as good as those in a supervised manner. For this reason, in most papers involving unsupervised denoising algorithms, results of unsupervised models and supervised models are carefully and fairly compared (on the same dataset, w/wo ground-truth labels). It is recommended to provide such an experimental analysis.

Response (major) 2

We appreciate the recommendation to compare the unsupervised denoising algorithm to a supervised model. We would like to emphasize that a supervised method is challenging to apply in practice because it is difficult to experimentally obtain the training pairs of noisy data and noise-free ground truth data. Nevertheless, in response, we have added the following experimental analysis for the comparison as **Supplementary Note 2** and **Supplementary Figure S6**. Please note that the results figure has been updated from the original **Figure 2** since we used a different test dataset; a subset of the original test dataset, which was withheld from training the supervised denoiser. In **Supplementary Note 2**, both methods (supervised and unsupervised) are evaluated on this set.

“Note 2: Comparison to a supervised deep learning approach

To benchmark the performance of the unsupervised method in DNM, we compare it to a supervised deep learning-based denoiser. This supervised denoiser is a network trained using the attenuated data as input and the clean ground-truth data as target. In practice, such training data constituting both noisy data and noise-free data is much more difficult to obtain than the data required for the

unsupervised method in DNM. DNM only requires a time series from particle-free ultrapure water, collected separately from the noisy time series that is to be denoised.

For the supervised denoiser to also predict the point-wise median of possible solutions (**Apparatus and denoising workflow for DNM**), it is optimized by minimizing the L1 norm of the error between prediction and target. For a fair comparison, the architecture of the supervised network is the same as the architecture of the VAE in the unsupervised DNM, except stochastic layers are made deterministic by replacing random samples from the learned Gaussian distributions with the learned mean of the distributions. This makes the network a U-Net⁷. Unlike the unsupervised DNM, a denoiser trained in a supervised manner should not be evaluated using its training data, so we withheld 10% of the total dataset for evaluation and used it to evaluate both methods. The results in **Figure S6 (R2_2 in this response)** were obtained from this subset.

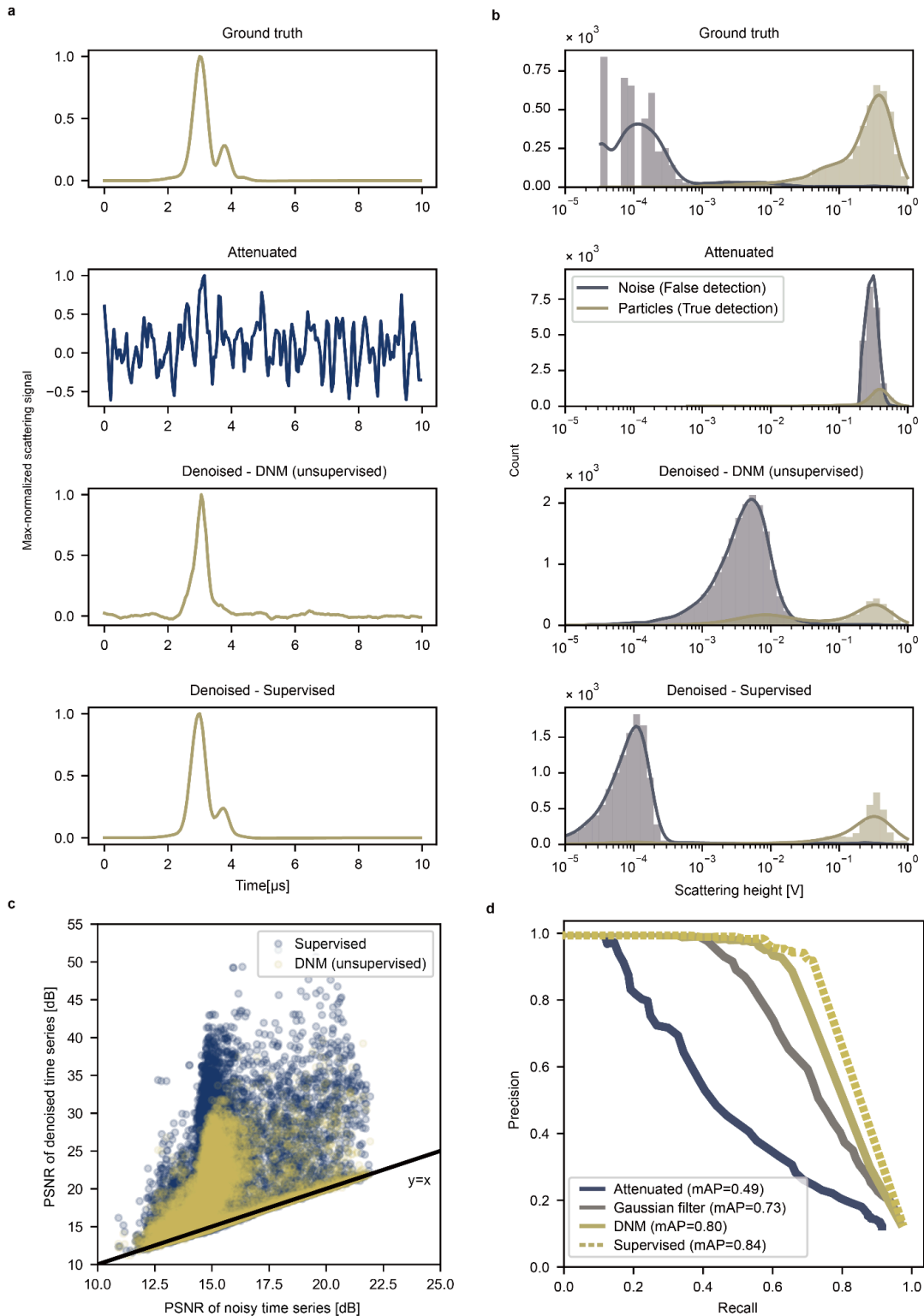
First, **Figure S6(a) (R2_2 in this response)** shows a max-normalized ground-truth time series, the corresponding attenuated time series before denoising, and the attenuated time series after denoising by both the supervised denoiser and the unsupervised denoiser in DNM. The supervised denoiser produced a result more closely resembling the shape of the ground truth peak. In the histograms of max-normalized scattering height distributions in **Figure S6(b) (R2_2 in this response)**, the supervised denoiser distinguishes more particle detections from noise than the unsupervised denoiser in DNM, which has some overlap between noise and particle scattering heights below 0.05.

Next, **Figure S6(c) (R2_2 in this response)** shows the scale-invariant peak signal-to-noise ratio (PSNR)⁸ of the attenuated time series before and after denoising by both the supervised denoiser and the unsupervised denoiser of DNM. This was obtained by dividing all time series equally into windows of equal length of 200 sample points and employing the ground truth to disregard windows containing no particle detections. We find that the supervised denoiser improves scale-invariant PSNR in 79.6% of cases, compared to 76.0 % of cases denoised by the unsupervised DNM. The supervised denoiser achieves a median scale-invariant PSNR value of 15.6 dB, while the unsupervised DNM achieves a median scale-invariant PSNR value of 15.2 dB. The raw attenuated data has a median scale-invariant PSNR value of 14.7 dB. Finally, the precision-recall (PR) curve in **Figure S6(d) (R2_2 in this response)** was generated using the same procedure as **Figure 2(e)** in the original manuscript. The two denoisers produce output with a similar precision up to a recall of 0.6. We observe that the precision of the supervised denoiser's output sharply drops at a recall of 0.8, compared to the precision of DNM's output sharply dropping at a recall of 0.7. The mean average precision (mAP) of the supervised denoiser's output is 0.04 higher than the unsupervised DNM's output.

In conclusion, the supervised method outperforms DNM across metrics. We believe that the main reason for this gap is the fact that the supervised approach can learn to remove signal independent noise as well as signal dependent shot noise, while our method is currently trained solely to remove signal independent noise. However, it should be again noted that the ground truth data required for supervised methods is rarely available in practice, and our unsupervised method did provide a substantial improvement."

"[7] Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention 2015 9351 (2015)

[8] Luo, Yi, and Nima Mesgarani. "Tasnet: time-domain audio separation network for real-time, single-channel speech separation." 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018."



Supplementary Figure S6, Figure R2_2 in this response:

Comparison of the supervised and unsupervised denoising results (a) Acquired max-normalized time-series scattering signals derived from 110 nm beads. The ground-truth signal, the attenuated signal before denoising, the attenuated signal after denoising by the unsupervised denoiser (DNM), and the attenuated signal after denoising by the supervised denoiser, respectively. **(b)** Histograms of the max-normalized scattering peak heights. In the ground truth times series, peak heights exceeding a threshold of 0.01 are considered true nanoparticles, and those lower than the threshold are noise. In the attenuated time series, peaks with a height exceeding 0.05 are considered detections (positives),

and only those that coincide with the positions of the ground truth peaks (true nanoparticles) are considered correct detections (true positives). In the denoised time series, peaks at the same location as the positive peaks in the attenuated time series are considered as detections. (c) Scatter plot comparing PSNRs before and after supervised and unsupervised (DNM) denoising illustrates the quality of the time series relative to the ground truth series. Points above the line of $y = x$ show improved PSNR, indicating an improved ability to detect particles. In contrast, points below this line represent a decrease in PSNR, suggesting that the particle signals are erroneously treated as noise and flattened. (d) Precision and recall (PR) curve for evaluating precision and recall of peak detection in time series denoised by the supervised model, added to **Figure 2(d)** in the main text as a dashed olive line. The mAP is the area under the PR curve.

In addition, we have added the following sentence in lines 207-209 on page 6 of the main text:

“See **Supplementary Note 2** for a detailed discussion of a comparison to a supervised deep learning approach.”

Major Comment 3

The authors claim that DNM achieves a high throughput of 100,000 events/second (Line 255-258). DNM is based on VAE, and its denoising execution requires multiple samples (Line 105). However, Is the denoising performed in realtime? How long does it take to process events generated within one second? What is the specific configuration (CPU, CPU memories, GPUs, system version,python/C++ version) of the workstation/computing server used in this study? Addressing these issues will make the algorithm part easy to follow.

Response (major) 3

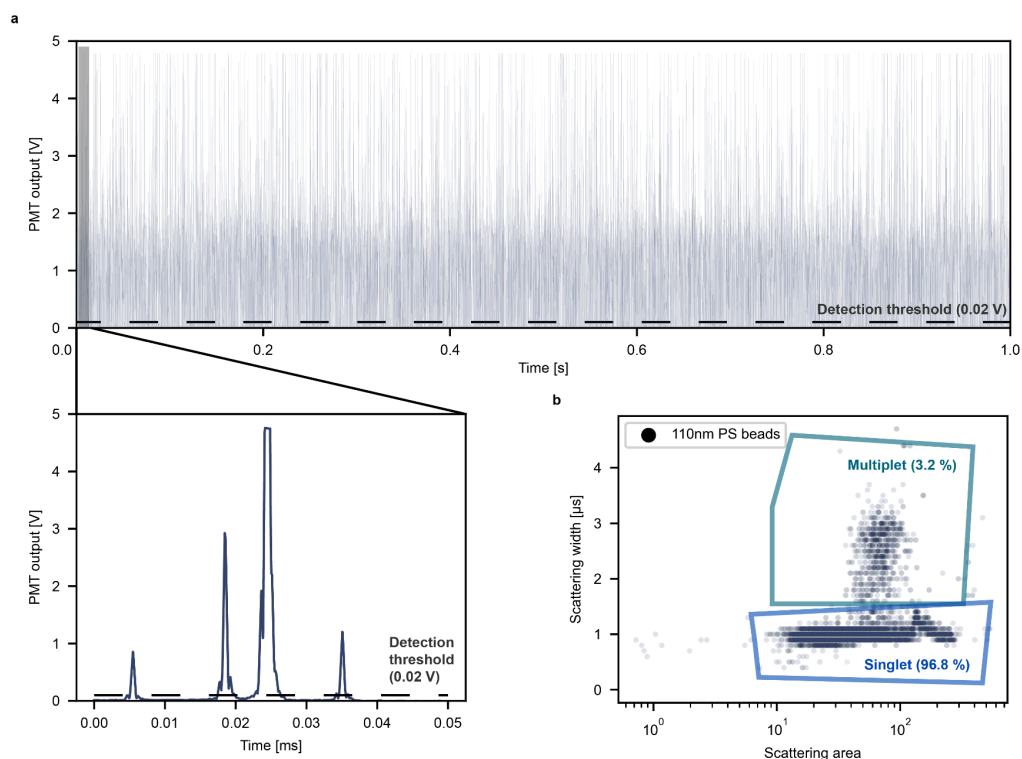
We appreciate the insightful comments and the request for further practical details. In the original manuscript, 100,000 events/second was intended as the detection throughput, and we apologize for any confusion caused by the misleading statement. As the reviewer pointed out, denoising analysis at the throughput described in the original manuscript is not feasible with our current implementation. To provide clarity, we have concretely redefined throughput into two categories: detection throughput and denoising throughput.

Detection Throughput:

Detection throughput refers to the maximum number of particles that can be optically detected per unit of time. The minimum separable time width of the scattering peak determines the theoretical maximum throughput. When particle spacing is uniform, the maximum throughput equals the reciprocal of the scattering peak width (1 μ s), allowing up to 100,000 events per second in DNM^{1,2,3}. However, in reality, since particle arrival is random and follows a Poisson distribution, there is a risk of particles arriving too close together and being detected as a single event. To mitigate this, the particle solution was diluted to ensure an average arrival interval of 10 μ s, enabling detection rates of up to 100,000 events per second.

In addition, we further experimentally assessed throughput using 110 ± 8.7 nm particles⁴ (Thermofisher, F8803) to answer another reviewer's request. The scattering signals were recorded over 1 second, with peak detection and peak height analyzed offline. In **Figure R2_3**, we visualized the scattering time series and histogram. Based on the plot of scattering width versus scattering area, we categorized the signals into singlets and multiplets, defining

multiplets as signals with a typical scattering width of 1.5 μs or more. A "multiplet" indicates multiple particles or aggregates detected simultaneously. Experimental results revealed that the DNM system detected 241,510 events per second, with 96.8% being singlets and 0.6% multiplets.



Supplementary Figure S1, Figure R2_3 in this response:

Detection of 110nm polystyrene (PS) beads by DNM (a) Time series of scattering signals of the PS beads measured for 1 second. The zoomed-up panel shows the time series scattering signal of the beads measured for 0.05 ms. (b) The scatter plot of scattering width and scattering area of the beads, obtained by applying a constant trigger threshold to the time series. In this plot, we identified the singlet and multiplet detections by gating the plot of scattering width and area, where we call the detection of individual particles a “singlet” detection and the simultaneous detection of two or more particles or aggregates of particles a “multiplet” detection. From this analysis, we obtained a detection throughput of the DNM as 241,510 events for one second, of which 96.8% were singlets and 0.6 % were multiplets.

Based on the discussion and analysis, we revised the main text in lines 315-321 on page 11 now states:

“In our method, the temporal width of each detected scattering peak is approximately 1 μs , theoretically allowing a maximum throughput of up to 1,000,000 events per second, assuming equal distances between particles. When we dilute the particle solution to a concentration where the particles arrive every 10 μs on average, considering the random arrival times of particles and the need to minimize coincident events, our system can confidently detect up to 100,000 events per second (Figure 3(a)).”

Also, we have added **Supplementary Figure S1**, and revised the main text in lines 322-324 on page 11 as follows:

“Experimentally, we achieved a throughput of 241,510 events per second by flowing a suspension of 110 nm PS beads and gating singlet events, which accounted for 96.8% of the total events in the scattering width versus area plot (**Supplementary Figure S1**).”

Denoising throughput

Denoising throughput refers to the actual computation time required to train the autoregressive noise model and VAE on a fixed amount of time-series data, followed by the denoising process. The training time is influenced by the computational hardware and the efficiency of the code, while the denoising time depends on these factors as well as the number of candidate time series generated. In our computational setup, training the noise model and VAE using 1 second of data took a total of 2.5 hours. For denoising with the trained model, processing 1 second of data, while randomly sampling 100 candidate time series from the estimated signal distribution and computing the median of them, took 8.5 minutes.

We have added the following description of denoising throughput, the computational hardware and software that we used, and the time that it took to train and apply our denoiser to **Supplementary Note 3**.

“Note 3: Training and evaluation time

Our workstation consists of a 36 core Intel(R) Core(TM) i9-10980XE CPU @ 3.00GHz with 134GB of RAM running Ubuntu 22.04.4, Python 3.11.5 and pytorch-lightning 2.2.1. The GPU used for training is a NVIDIA GeForce RTX 3090 with 24GB of VRAM. The minimum VRAM required for training the model is 2GB. In our environment, total training time for the noise model and the VAE, using 1 second of data for each, takes 2.5 hours. Once the model is trained, denoising a 1 second time series by randomly sampling 100 denoised time series from the estimated signal distribution and computing the median of them takes 8.5 minutes.”

We recognize that real-time denoising is a crucial step for further development of the sorting system and the implementation of DNM. Our efforts to implement real-time denoising are indeed underway, and we want to report on them in future updates.

Once again, we appreciate the valuable feedback. These revisions and additional details will improve the clarity and practicality of our manuscript.

References

1. Diebold, Eric D., et al. "Digitally synthesized beat frequency multiplexing for sub-millisecond fluorescence microscopy." *Nature Photonics* 7.10 (2013): 806-810.
2. Han, Yuanyuan, et al. "Imaging technologies for flow cytometry." *Lab on a Chip* 16.24 (2016): 4639-4647.
3. Han, Yuanyuan, and Yu-Hwa Lo. "Imaging cells in flow cytometer using spatial-temporal transformation." *Scientific reports* 5.1 (2015): 13267.
4. <https://www.thermofisher.com/order/catalog/product/jp/ja/F8803>

Major Comment 4

Line 498-501 and Line 507-514. Loss function configurations of the two neural network models are ambiguous.

Response (major) 4

We thank the reviewer for making us aware of this. The explicit loss function used to train the noise model has been added to Line 21 of **Supplementary Note 1** as follows,

“The training loss (being minimized) for this model is defined as the negative log-likelihood of the training data,

$$L(\eta; n) := -\log p_{\eta}(n) = -\log p_{\eta}(n_1) - \sum_{t=2}^T \log p_{\eta}(n_t | n_{<t}),$$

where n is a time series of noise, n_t is the noise realization at time t , $n_{<t}$ denotes the noise realizations before time t , T is the data dimensionality and η are the parameters of a CNN built from shifted convolutions as explained in **Extended Figure 2.**”

The loss function used to train the VAE has been broken up into two major parts, the regularization loss and the reconstruction loss. These changes are made to Line 40 of **Supplementary Note 1** as follows,

“In this process, a Ladder VAE³ is trained to maximize a lower bound on the logarithm of the likelihood of noisy data $x \in R^T$, denoted as $\log p_{\theta}(x)$. This is known as the Evidence Lower Bound (ELBO). It is computed by first encoding an input x as a learned distribution over latent variables z , $q_{\phi}(z|x)$, using a CNN known as the encoder. A sample from this distribution is drawn and passed to a CNN known as the decoder, which outputs a denoised time series $\hat{s} = g_{\theta}(z)$.

The VAE training loss is defined as the negative ELBO, given by

$$L(\theta, \phi; x) := E_{q_{\phi}(z|x)}[-\log p_{\eta}(x - g_{\theta}(z))] + D_{KL}[q_{\phi}(z|x) \parallel p_{\theta}(z)],$$

where the expectation $E_{q_{\phi}(z|x)}$ is approximated using random sampling and the reparameterization trick⁵, D_{KL} is the Kullback-Leibler divergence, $p_{\theta}(z)$ is a prior distribution over the latent variables and p_{η} is the noise model from the previous section with frozen parameters. The parameters ϕ and θ are jointly optimized to minimize the loss function $L(\theta, \phi; x)$.

The loss function can be broken up into two summed terms, called the regularization loss, $L_{KL}(\theta, \phi; x) := D_{KL}[q_{\phi}(z|x) \parallel p_{\theta}(z)]$ and the reconstruction loss, $L_{Rec}(\theta, \phi; x) := E_{q_{\phi}(z|x)}[-\log p_{\eta}(x - g_{\theta}(z))]$. We will first describe the regularization loss.

Since we use a Ladder VAE, the latent variable is partitioned into hierarchically dependent layers, $z = \{z_1, \dots, z_L\}$. The joint distributions over the latent variables are autoregressive, such that they factorize as,

$$p_{\theta}(z) = p_{\theta}(z_1) \prod_{l=2}^L p_{\theta}(z_l | z_{<l}),$$

$$q_{\phi}(z|x) = q_{\phi}(z_1|x) \prod_{l=2}^L q_{\phi}(z_l | z_{<l}, x).$$

At the first level of the hierarchy, the prior $p_{\theta}(z_1)$ is a unit Gaussian, and at subsequent layers the prior $p_{\theta}(z_l | z_{<l})$ is a Gaussian with mean and standard deviation as a function of the sample from the

previous layer. The same is true for the encoding distribution $q_\phi(z|x)$, but Gaussian parameters at all levels are also a function of x .

The regularization loss is therefore decomposed as,

$$L_{KL}(\theta, \phi; x) = D_{KL}[q_\phi(z_1|x) \parallel p_\theta(z_1)] + \sum_{l=2}^L E_{q_\phi(z_{<l}|x)}[D_{KL}[q_\phi(z_l|z_{<l}, x) \parallel p_\theta(z_l|z_{<l})],$$

where $q_\phi(z_{<l}|x) := \prod_{i=1}^{l-1} q_\phi(z_i|z_{<i}, x)$ ⁶ and the expectation $E_{q_\phi(z_{<l}|x)}$ is again estimated using sampling and the reparameterization trick⁵.

To calculate the reconstruction loss, a latent variable is sampled from the final layer of the encoding distribution and deterministically mapped to an estimate of the denoised signal, $\hat{s} = g_\theta(z)$, via a learned function, g_θ , which is the decoder. Since the noisy input is the sum of a particle signal and noise, $x = s + n$, an estimate of the noise, $\hat{n}$, in the input is obtained by subtracting the estimate of the denoised signal from the input. The noise model trained in the previous section is used to evaluate the log-likelihood of this noise estimate.”

“[3] Sønderby, C. K., Raiko, T., Maaløe, L., Sønderby, S. K. & Winther, O. Ladder Variational Autoencoders. *Adv Neural Inf Process Syst* **29**, (2016).

[5] Kingma, D. P. & Welling, M. Auto-Encoding Variational Bayes. *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings* (2013).

[6] Vahdat, A., & Kautz, J. NVAE: A deep hierarchical variational autoencoder. *Advances in neural information processing systems*, 33, 19667-19679 (2020).”

Major Comment 5

The algorithm proposed is central to the paper but I cannot find any available code/software attached. Is the code/software and demon available?

Response (major) 5

We appreciate the reviewer's interest in our algorithm. The code and sample data have been prepared and are currently hosted on GitHub. Although the repository is private until the paper is published, we share the small dataset and our full code with the editor and ask the editor to share it with the reviewers. Reviewers can also access the code and related materials using the following link:

[<https://drive.google.com/drive/folders/1Nmkl04phUMeCnvoF5ksQ4ES6BwXueWG8?usp=sharing>]

Please do not hesitate to contact us if you have any problems accessing the repository. We appreciate the feedback and hope that this access will facilitate a thorough review.

Major Comment 6

In the implementation, the noise model is trained with data obtained by recording with particle-free ultrapure water. Therefore, the noise distribution learned by the neural network

should be instrument-specific and signal-independent. As a consequence, the denoising VAE seems to be incompetent in dealing with signal-independent noise.

Response (major) 6

We believe there may have been a minor typo in the reviewer's comment, and that they intended to refer to *signal-dependent* rather than signal-independent noise. We agree that the VAE implemented in the DNM is currently unsuitable for dealing with signal-dependent noise. This means that sensitivity in this work was increased by removing only signal-independent noise. We would like to make this clear by making the changes detailed in **Response (major) 1** and revising lines 111-116 of page 4 of the main text as follows:

“Once the whole system is trained, we feed noisy measurements x into our VAE and randomly sample denoised signals from the signal model, $q_{\phi, \theta}(s|x)$. (**Extended Figure 1 (a)-(c), and Supplementary Note 1**) Since the noise model assumes the signal intensity in its training data is zero throughout, it does not describe signal-dependent noise. Therefore, samples from the signal model will be shot noise-affected particle signals.”

Extending our method to also remove signal-dependent noise would be a highly impactful future work. Recently, it was shown that signal-dependent noise in images can be removed with a similar unsupervised approach¹. We are currently trying to determine whether it is possible to adapt this method to time series data.

Based on this, we also revised lines 494-499 of page 17 of the main text:

“An important milestone for future work will be developing a method capable of additionally removing signal-dependent shot noise. While in this work, the sensitivity was significantly improved by removing only the signal-independent noise as it was dominant in the scattered signal, we also observed that the signal-dependent noise in fluorescence data, for example, was too intense to ignore. To this end, a promising solution could be to extend another version⁵³ of our denoising approach, originally developed for removing signal-dependent noise in images.”

“[53] Salmon, B. & Krull, A. Unsupervised Denoising for Signal-Dependent and Row-Correlated Imaging Noise. *arXiv preprint [arXiv:2310.07887v2](https://arxiv.org/abs/2310.07887v2)*, (2024)”. Cited as [1] in this Response (major) 6.

Major Comment 7

In DNM, the one-dimensional noise is learned by a CNN, which is commonly used to deal with 2-D image data. Considering that scattering noise is temporally correlated, recurrent neural networks(e.g. LSTM) seems to be more suitable to capture the temporal correlation in time series.

Response (major) 7

We appreciate this insightful comment regarding the choice of a Convolutional Neural Network (CNN) for learning one-dimensional noise in DNM. We agree that an LSTM network could be a suitable solution and has been used in works on autoregressive models¹.

We chose to use CNNs instead, primarily due to their efficiency, as the entire time series can be processed in a single forward pass. In practice, we pass 1000 time points at once. This significantly increases the speed of training compared to recurrent neural networks, which require time points to be processed one at a time. CNNs are also very effective in handling short-range correlations, which are characteristic of the noise and signal in our data. This is evidenced by works such as PixelVAE², which leverages this capability in image processing.

In our implementation, we employ size 11 kernels, which are substantially larger than the typical 3 or 5 size kernels used in image modeling, enhancing the CNN's ability to capture the temporal correlations present in our time-series data. This choice provides a balance between computational efficiency and effectively modeling relevant correlations.

References

1. Van Den Oord, Aaron, Nal Kalchbrenner, and Koray Kavukcuoglu. "Pixel recurrent neural networks." *International conference on machine learning*. PMLR, 2016.
2. Gulrajani, Ishaan, et al. "Pixelvae: A latent variable model for natural images." *arXiv preprint arXiv:1611.05013* (2016).

Major Comment 8

More details about the training objective of the noise model should be provided.

Response (major) 8

We agree and have provided the explicit loss function used to optimize the noise model in **Supplementary Note 1**. We added a relevant description in **Response (major) 4**.

=====

Minor Comment 1

In terms of the "voting" mechanism of the 100 samplings from the denoising VAE, why choose "median" instead of other statistical indicators such as "mean average"? And I wonder how diverse are the different samplings from the VAE?

Response (minor) 1

We thank the reviewer for their insightful question regarding why we chose the median instead of the mean and the uncertainty and confidence when selecting the median.

One of the reasons why we chose the median as a consensus statistic is its robustness against outliers. By using the median of 100 denoised samples, we aim to capture the most consistent behavior of the signal while minimizing the influence of random fluctuations caused by noise. The median ensures that the result represents the central tendency of the data, and will be stable as long as the number of outliers lies below 50%. In contrast to the median, the mean of samples is more sensitive to outliers. If only a few percent of the denoised samples have a sufficiently high intensity, this may distort the mean and lead to (false) detections even when there is only a negligible predicted probability of a particle being present.

Regarding the confidence in distinguishing signal from noise, the median approach inherently reflects the prediction confidence of the model. For any given time point, if the median of the samples lies above the detection threshold, at least 50% of the predictions suggest the presence of a signal. This provides a probabilistic measure of confidence: the model estimates that it is more likely that there is a particle than not. To quantify uncertainty further, one could adjust the percentile used for the consensus. For example, choosing the 10th percentile would provide more conservative results, requiring 90% of the samples to be above the detection threshold to indicate a signal, thus increasing confidence at the cost of

potentially missing weaker signals. Conversely, using the 90th percentile would lower the confidence threshold, increasing sensitivity but raising the risk of false positives.

Based on the discussion above, we have added the following text to lines 118-134 of page 4 to explain how the median of denoised samples can be interpreted.

“For evaluation, the random denoised samples from a trained signal model must be aggregated into a single consensus solution with which peak detection can be performed. We chose to use the point-wise median of the samples as it can be used to interpret the model’s confidence for this task. At each time point, denoised random samples can be classified as either above or below the peak-detection threshold. If the median at that point in time is above the threshold, at least 50% of the denoised samples must be above. Therefore, DNM has predicted the probability distribution over denoised solutions to have at least 50% of its probability above the detection threshold. In other words, the model believes it is more likely that there is a particle than not.

For more certainty from the model, a lower percentile, for example the 10th, could be used as a consensus. This would demand that the model place 90% of its samples above the detection threshold to reveal a signal peak. Inversely, a higher percentile, for example the 90th, would require that the model place only 10% of its samples above the detection threshold to reveal a signal peak.

In contrast to the median, the mean of samples is more sensitive to outliers. Even if only a few percent of the noise-removed samples have a sufficiently high intensity above the detection threshold, this could distort the mean and lead to false detections.”

Minor Comment 2

Line 111 and Figure 1. Given that, in general, the excitation intensity and beam uniformity and stability of a 408 laser are not as good as those of a 488 laser, but with a higher risk of photodamage, I am confused as to why the authors chose a 408 laser to obtain SSC.

Response (minor) 2

We appreciate the comment regarding our choice of laser for the experiments. We agree that the excitation intensity, beam uniformity, and stability of 408 nm lasers are lower than those of 488 nm lasers, and there may be a higher risk of optical damage. Nevertheless, we chose the 408 nm laser for the SSC measurement because we expected that the signal-to-noise ratio of the scattering from the particles would be higher, leading to the superior detection sensitivity. Additionally, we anticipate that issues such as beam uniformity, stability, and photodamage would be minimal, not significantly impacting our measurements. We will explain these accordingly.

The estimation of sensitivity for the laser selection

The sensitivity, the signal-to-noise ratio (SNR), of scattering detection is determined by the scattering cross-section and the input laser power. The scattering cross-section quantifies how efficiently a particle interacts with and scatters the light. The scattering cross-section can be derived from Rayleigh scattering theory⁸ and is expressed as

$$\sigma_{scatt}(d, n_{med}, n_{particle}, \lambda) = \frac{2\pi^5 d^6 n_{med}^4}{3\lambda^4} \left| \frac{m^2 - 1}{m^2 + 2} \right|^2, \quad (1)$$

where $m = \frac{n_{particle}}{n_{med}}$ and d is the particle size, n_{med} is the refractive index of medium, $n_{particle}$ is the refractive index of the particle, λ is the excitation wavelength. According to this **equation (1)**, the cross-section is inversely proportional to the fourth power of the wavelength. This

means that, as the wavelength becomes smaller, the cross-section increases, making scattering events easier to detect. Thus, the scattering cross-section for polystyrene particles is larger with the 408 nm laser than that with the 488 nm laser (**Table R2_1**).

On the other hand, the input laser power is also a critical factor as the SNR is proportional to the square root of that. For a 250 mW 408 nm laser and a 500 mW 488 nm laser used in our setup, when the same scattering cross sections are assumed, the 488 nm laser achieves an SNR approximately $\sqrt{(500/250)} = 1.4$ times higher than that the 408 nm laser achieves.

Collectively, the relative SNR for the 408 nm laser to the 488 nm laser can be expressed as a function of the scattering cross section and the square root of the relative laser power (**equation (2)**). As a result, for example, the SNR of 40 nm polystyrene particles is greater with the 250 mW 408 nm laser than with the 500 mW 488 nm laser (relative SNR >1), making the 408 nm laser more advantageous for high-sensitivity measurements (**Table R2_1**).

$$\begin{aligned} \text{Relative SNR} &= \text{relative cross section} \times \sqrt{\text{relative laser power}} \\ &= \frac{\text{cross section of particle [nm}^2\text{]}}{\text{cross section of 40nm, 408 nm excitation [nm}^2\text{]}} \times \sqrt{\frac{\text{Laser power [mW]}}{250 \text{ [mW]}}} \quad (2) \end{aligned}$$

Table R2_1 Comparison of sensitivity by using the different wavelength lasers

Laser	Material	Particle RI	Particle size [nm]	Cross section [nm ²]	Excitation Power [mW]	Relative SNR
408 nm	Polystyrene	1.62	40	1.81	250	1.00
488 nm	Polystyrene	1.60	40	0.78	500	0.61

Effect of laser non-uniformity

The scattering intensity results from a multiplication of the scattering cross-section of particles and the input laser intensity at the position of particles. In our system, the width of streamlines of particles is focused to about 1-2 μm , where the 408 nm beam profile spans perpendicular to the stream with 5 μm width at half maximum. Therefore, we assume that the particles pass through approximately the same position of the sufficiently stretched laser, experiencing similar excitation intensities.

To confirm the minimal effect of laser uniformity on detection, we calculated the coefficient of variation (CV) of the experimentally measured scatter height from 110 ± 8.3 nm PS particles¹ (Thermofisher, F8803) using the same data shown in **Figure R2_3** in **Response (major) 3**, where the CV is defined as the standard deviation of the scatter height divided by the mean scatter height. As a result, after removing multiplets and saturated peaks, the CV of the measured scattering height was 42.1%. If the non-uniform laser profile affects our detection, this measured CV will become significantly larger than the ideal CV estimated based on a spec sheet provided by the manufacturer¹.

In converting the spec sheet-derived size to the scatter height, we first created a normal distribution based on the given mean and variance of the particle size distribution, then estimated the scattering cross-section distribution according to **equation (1)**, and finally calculated the CV from the mean and variance of the scatter distribution. As a result, the estimated CV of the scattering height was 45.6%. Herein, the CV of the scatter cross-section and the CV of the scatter signal coincide because the scattering signal is proportional to the scatter cross-section, and the CV is normalized to the mean.

Above all, the CV of the experimentally measured scattering height (42.1%) was close to the CV of the spec sheet-derived scattering height (45.6%), indicating that the contribution of the beam profile to the measurement variability was minimal.

Effect of laser instability

To investigate the effect of the temporal instability of the 408 nm laser on the detection, we perform frequency analysis on the instrument-specific noise before applying the denoising method and discussing the origin of the noise. Herein, RIN, dark noise from the detector, and fluidic noise from the PBS are considered instrument-specific noise. We measured each noise as follows: (i) we measured the RIN noise by directing the laser input into the PMT after attenuating the laser power, (ii) the dark noise with the laser turned off, and (iii) the fluidic noise with particle-free PBS flowing in the channel as a sample stream. As a result, the typical frequency peaks seen in the power spectrum of RIN noise and dark noise did not appear in that of the fluidic noise, suggesting that the fluidic noise was dominantly large compared to the others in the DNM (**Figure R2_4**). Thus, we believe that relative intensity noise (RIN) associated with the less stable 408 nm laser did not significantly contribute to the noise characteristics in this study.

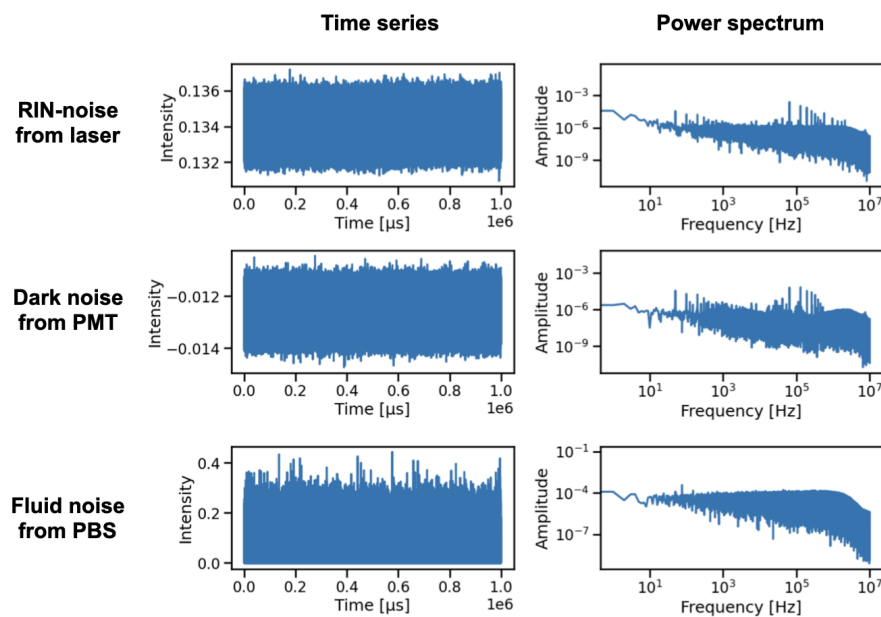


Figure R2_4:
Scattering time series and frequency power spectrum for each type of noise (RIN noise, dark noise, fluidic noise)

Photodamage

We believe that the photodamage due to 408 nm is negligible due to the short exposure time, approximately 1 μ s, during which particles pass through the illumination. The high-speed fluid flow and minimal water absorption at 408 nm also result in an insignificant

temperature effect. Lastly, since the measurement targets—EVs and nanoparticles—are non-living materials, we think that the potential photodamage caused by the 408 nm laser would not be a significant issue.

References

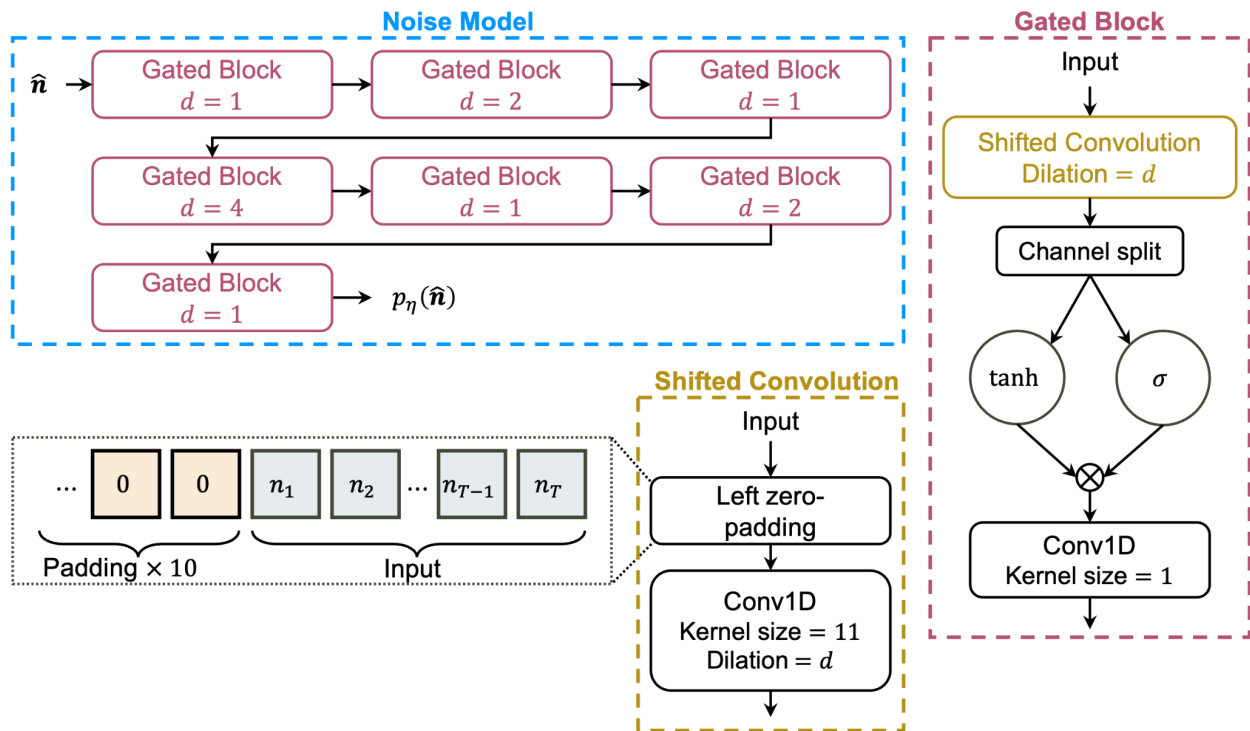
1. <https://www.thermofisher.com/order/catalog/product/jp/ja/F8803>,
2. Bohren, Craig F., and Donald R. Huffman. *Absorption and scattering of light by small particles* (John Wiley & Sons, 2008).

Minor Comment 3

Line 498-501 and Line 507-514. It is strongly recommended to provide a neural network structure/architecture diagram (in an extended or supp figure, for example).

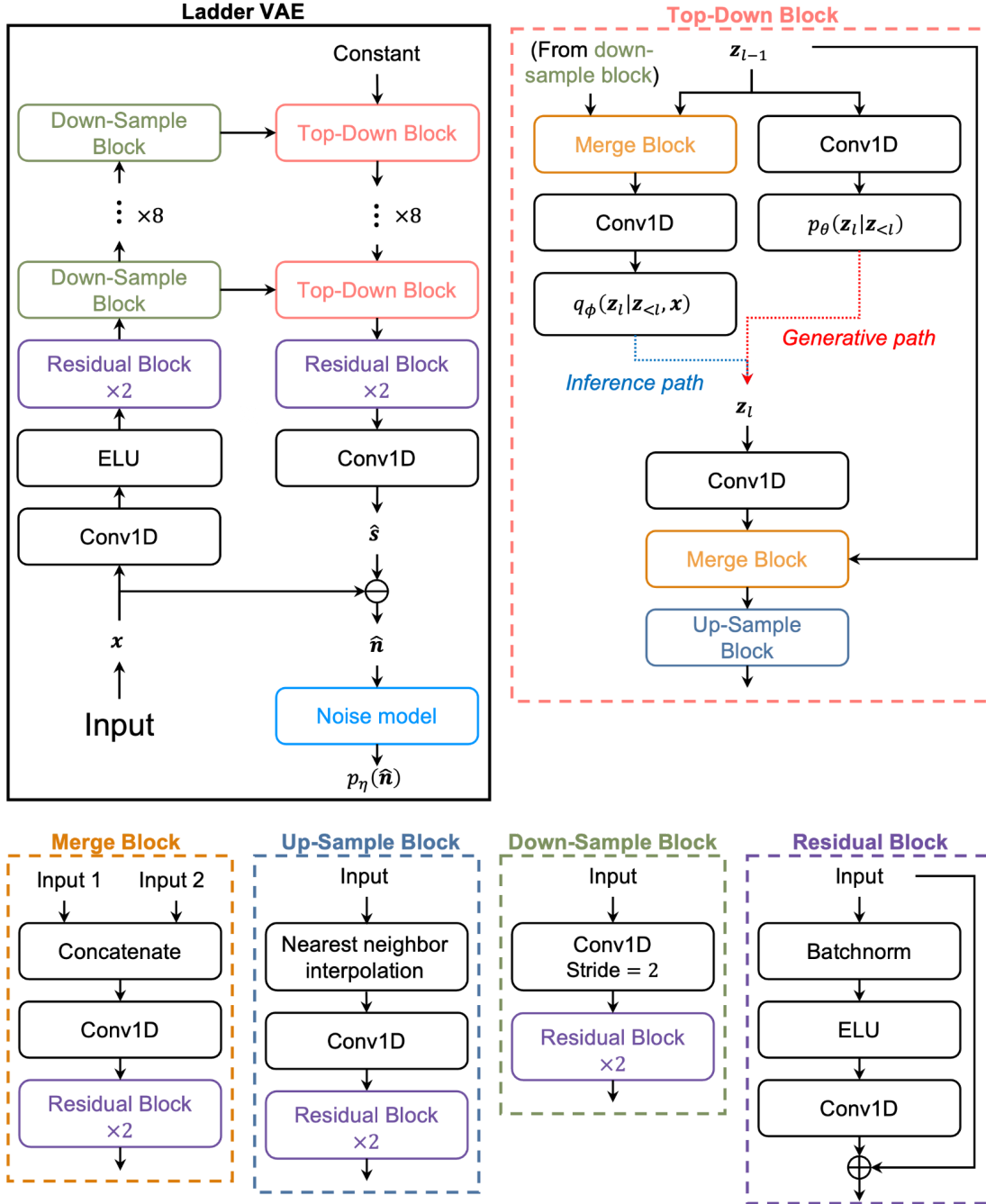
Response (minor) 3

We agree that providing the neural network architecture diagram would improve understanding of our network and we include the following figures as **Extended Figure 2 (Figures R2_5 and R2_6)** in the revised manuscript. Additionally, we would like to stress that the full code including the network architectures will be publicly available upon publication.



Extended Figure 2(a), Figure R2_5 in this response:

Noise model architecture. The noise model is based on the PixelCNN architecture³¹, adapted to one-dimensional data by using one-dimensional convolutions. The Shifted Convolution module enforces the autoregressive structure of the noise model. It pads the start of the input array with $k - 1$ zeros, where k is the size of the convolutional kernel. This means that the t th element of the output array is a function of the first t elements in the input array. The exception to this is in the first Shifted Convolution of the noise model, where the input is padded with k zeros, so that the t th element of the output array is a function of the first $t - 1$ elements in the input array. In the Channel split, the input array is divided in two along the channel axis. The hyperbolic tangent and sigmoid functions are represented by $\tanh$ and σ respectively. $\otimes$ represents element-wise multiplication.



Extended Figure 2(b), Figure R2_6 in this response:

Ladder VAE architecture. This follows the architecture used by Prakash et al.⁵⁸. The constant fed into the first top-down block is an array of zeros with the same dimensions as the data. The probability distributions $q_{\phi}(z_l | z_{<l}, x)$ and $p_{\theta}(z_l | z_{<l})$ are Gaussians. Their means are the first half of channels of the block's input and their standard deviations are the second half, with positivity enforced by a softplus function. All convolutions have a kernel size of 3. $\oplus$ and $\ominus$ represent element-wise addition and subtraction respectively. The concatenate module denotes concatenating arrays along the channel axis. ELU represents the Exponential Linear Unit activation function⁵⁹.

[31] Van Den Oord, Aaron, Nal Kalchbrenner, and Koray Kavukcuoglu. "Pixel recurrent neural networks." *International conference on machine learning*. PMLR, 2016.

[58] Prakash, Mangal, et al. "Interpretable unsupervised diversity denoising and artifact removal." *arXiv preprint arXiv:2104.01374*, (2021).

[59] Clevert, D. A., Unterthiner, T. & Hochreiter, S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *4th International Conference on Learning Representations, ICLR 2016 - Conference Track Proceedings* (2015).

Minor Comment 4

Line 530 and Equation 2. The super/sub-scripts of the sum symbol encompass $n+1$ datapoints instead of n .

Response (minor) 4

We appreciate bringing this to our attention. It has been corrected in the main text in lines 630-633 of page 20 as follows:

“The mean squared error is calculated as follows

$$Mean\ Square\ Error = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

When a total of n data points are obtained, y_i represents the true value (signal intensity at the ground truth) at the i -th measured data point and $\hat{y}_i$ represents the actual measured value (signal intensity before and after denoising).”

Minor Comment 5

Line 88-91. Suggest changing ‘address’ to another word. This is unproven in the paper and seems to be hard to prove/validate that with a trained VAE, solutions obtained by reparametrizing are consistent with the solution space of the original ill-posed problem.

Response (minor) 5

We appreciate this helpful feedback. We have reworded this section to say that we acknowledge the ill-posedness of the problem. Furthermore, we have clarified that we are estimating the distribution over the solution space. The revised description is now on lines 89-93 of page 3 as follows:

“Finally, our method acknowledges the ill-posed-ness of the denoising problem, where infinitely many possible clean signals can explain a given noisy measurement, by estimating the probability distribution over possible denoised signals for a given noisy measurement. This enables us to randomly sample denoised estimates.”

Minor Comment 6

Line 139-142. There are too many references, which may reduce the readability of the paper.

Response (minor) 6

We acknowledge the comment and agree that the number of references could affect readability. We have revised the lines and removed four redundant citations to address this,

reducing the original twelve references to seven. Specifically, in lines 38-39 on page 2, the revised sentence will read as follows:

“For the small EVs, however, conventional high-sensitivity multi-parametric analysis techniques^{15–21} struggle with scalability, leading to limited statistical accuracy in detecting minor subpopulations”

“[15] Naiim, M. *et al.* Multiangle dynamic light scattering for the improvement of multimodal particle size distribution measurements. *Soft Matter* **11**, 28–32 (2014).

[16] Dehghani, M., Gulvin, S. M., Flax, J. & Gaborski, T. R. Systematic Evaluation of PKH Labelling on Extracellular Vesicle Size by Nanoparticle Tracking Analysis. *Scientific Reports* 2020 10:1 **10**, 1–10 (2020).

[17] Filipe, V., Hawe, A. & Jiskoot, W. Critical evaluation of nanoparticle tracking analysis (NTA) by NanoSight for the measurement of nanoparticles and protein aggregates. *Pharm Res* **27**, 796–810 (2010).

[18] Zhu, S. *et al.* Light-scattering detection below the level of single fluorescent molecules for high-resolution characterization of functional nanoparticles. *ACS Nano* **8**, 10998–11006 (2014).

[19] Li, L. *et al.* Development of Spectral Nano-Flow Cytometry for High-Throughput Multiparameter Analysis of Individual Biological Nanoparticles. *Anal Chem* **95**, 3423–3433 (2023).

[20] Fraikin, J. L., Teesalu, T., McKenney, C. M., Ruoslahti, E. & Cleland, A. N. A high-throughput label-free nanoparticle analyser. *Nature Nanotechnology* 2011 6:5 **6**, 308–313 (2011).

[21] Kashkanova, A. D., Blessing, M., Gemeinhardt, A., Soulat, D. & Sandoghdar, V. Precision size and refractive index analysis of weakly scattering nanoparticles in polydispersions. *Nature Methods* 2022 19:5 **19**, 586–593 (2022).”

Minor Comment 7

Figure 3. Suggest replacing the symbol of PBS with a dashed box.

Response (minor) 7

We appreciate the remarks regarding the readability of the figures. Following the reviewer's comment, we have corrected the **Figure 3(b)(Figure R2_7)** as follows:

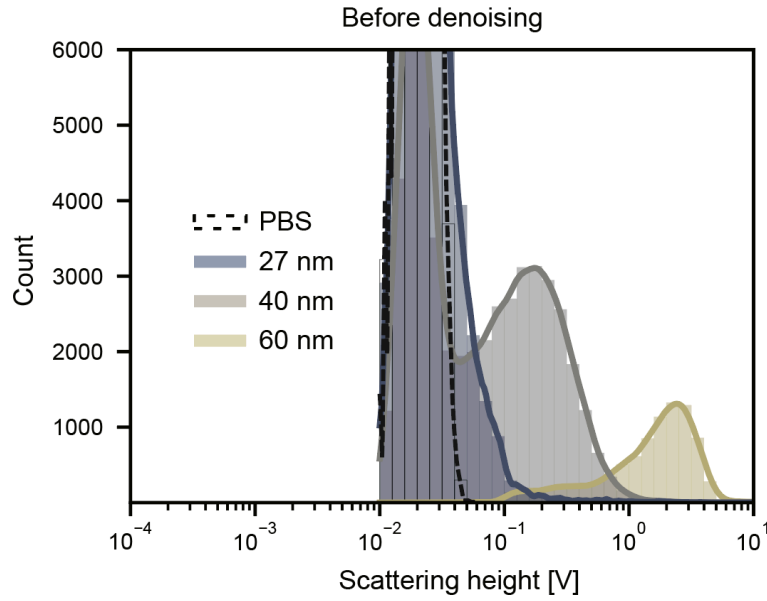


Figure 3(b), Figure R2_7 in this response.

Minor Comment 8

Supp Note 1, Line 38. What is the distribution of $p_{\theta}(z)$?

Response (minor) 8

We appreciate bringing this to our attention, we have clarified in Supplementary Note 1 that our VAE uses a hierarchy of latent variables. Therefore, at the first level of the hierarchy, the prior is a unit Gaussian, and at subsequent levels, the prior is a Gaussian with parameters as a function of the sampled latent variables from previous levels. These changes are on Line 53 of **Supplementary Note 1** as follows,

“Since we use a Ladder VAE, the latent variable is partitioned into hierarchically dependent layers, $z = \{z_1, \dots, z_L\}$. The joint distributions over the latent variables are autoregressive, such that they factorize as,

$$p_{\theta}(z) = p_{\theta}(z_1) \prod_{l=2}^L p_{\theta}(z_l | z_{<l}),$$

$$q_{\phi}(z|x) = q_{\phi}(z_1|x) \prod_{l=2}^L q_{\phi}(z_l | z_{<l}, x).$$

At the first level of the hierarchy, the prior $p_{\theta}(z_1)$ is a unit Gaussian, and at subsequent layers the prior $p_{\theta}(z_l | z_{<l})$ is a Gaussian with mean and standard deviation as a function of the sample from the previous layer. The same is true for the encoding distribution $q_{\phi}(z|x)$, but Gaussian parameters at all levels are also a function of x .

The regularization loss is therefore decomposed as,

$$L_{KL}(\theta, \phi; x) = D_{KL}[q_{\phi}(z_1|x) \parallel p_{\theta}(z_1)] + \sum_{l=2}^L E_{q_{\phi}(z_{<l}|x)}[D_{KL}[q_{\phi}(z_l | z_{<l}, x) \parallel p_{\theta}(z_l | z_{<l})],$$

where $q_{\phi}(z_{<l}|x) := \prod_{i=1}^{l-1} q_{\phi}(z_i|z_{<i}, x)$ ⁶ and the expectation $E_{q_{\phi}(z_{<l}|x)}$ is again estimated using the reparameterization trick⁵.”

“[5] Kingma, D. P. & Welling, M. Auto-Encoding Variational Bayes. *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings* (2013).

[6] Vahdat, A. & Kautz, J. NVAE: A Deep Hierarchical Variational Autoencoder. *Adv Neural Inf Process Syst* **2020-December**, (2020).”

Minor Comment 9

Figure S2. I cannot understand why the autocorrelation curve contains a trough (Lag = 4μs).

Response (minor) 9

Although this measurement seems counterintuitive, in our experience, it is not unusual in electronic signal processing that these negative correlations appear. For example, please see **Figure 3** of a recent paper where a negative spatial correlation is observed in imaging noise¹. We suspect that these effects are due to the design of the detection hardware and regret that discovering why the autocorrelation takes on the values is beyond the scope of our work.

References

1. Salmon, B. & Krull, A. Unsupervised Denoising for Signal-Dependent and Row-Correlated Imaging Noise. *arXiv preprint [arXiv:2310.07887v2](https://arxiv.org/abs/2310.07887v2)*, (2024).

=====

We believe that we have thoroughly addressed the reviewer's concerns. We are grateful for the reviewer's valuable feedback, which has significantly contributed to improving the quality of our work. We hope the reviewer finds the changes satisfactory and look forward to further evaluation.

We sincerely thank the reviewer again for her/his time and effort in reviewing our manuscript.

Best regards,
Authors

Reviewer #4:

Overall comment

In the manuscript titled "High throughput analysis of rare nanoparticles with deep-enhanced sensitivity via unsupervised denoising", the authors introduce a computational pipeline that leverages on machine learning to denoise 1D time traces. Specifically, the authors apply the algorithm to the detection and subsequent analysis of nanoparticles via flow cytometry. The scope of the work is focussed on the flow cytometry and nanoparticle characterisation community and addresses an existing challenge to detect rare subpopulations within a heterogenous sample mixture, which is only possible when high sensitivity is combined with high throughput. To put this work into a larger context of high throughput detection of single nanoparticles or highly sensitive detection of biological nanoparticles, similar approaches of the optofluidic platform exist either in the form of nanoFCM and single particle imaging FCM exist, as well as deep-learning implementations of particle detection, such as anomaly detection. The significance and novelty of this work comes down to demonstrating the use of unsupervised learning to push both the measurement throughput and detection sensitivity. The authors provide proof of validation experiments for detecting low abundance populations of cancer markers in a sample of EVs.

The manuscript is well written and structured and offers a promising implementation to access high throughput. Nevertheless, given that the manuscript describes a technical method, and the authors claim that this approach could potentially be a new standard for the analysis of nanoparticles using flow cytometry, it is this referee's opinion that both the description of the methods, formulation of the noise model, claims of multiparametric nanoparticle analysis and some of the conclusions stated in the manuscript could be further improved prior to acceptance.

Response to overall comment

We appreciate the reviewer's detailed and insightful feedback on our manuscript "High throughput analysis of rare nanoparticles with deep-enhanced sensitivity via unsupervised denoising." We appreciate the reviewer's recognition of our machine learning-based technology's novelty and potential value. The reviewer's comments have been instrumental in guiding us to refine our manuscript. We have enhanced the description of our methods, the formulation of the noise model, and the claims regarding multiparametric nanoparticle analysis. These revisions help us provide a more precise and rigorous presentation of Deep Nanometry (DNM) as a potential new standard for nanoparticle analysis.

Major Comment 1

Regarding the unsupervised denoising model. In the formulation of the model, the authors assume that the signal detected is the sum of a noise free signal and signal-independent background noise. We strongly encourage the authors to reassess this statement. It is not clear how the authors can assume that shot noise, an intrinsic source of noise to any optical read out technique, is a signal-independent background noise. Furthermore, when experiments are performed in complex sample mixtures such as serum, there are other background noise sources that should be included, such as those given by the presence of smaller nanoparticles such as proteins, which will lead to variable background scattering noise contributions.

Response (major) 1

We appreciate the reviewer's valuable comments and for pointing out the areas where our description could have been more precise. We apologize for any confusion caused by our original statement. We have revised the relevant sections of the manuscript to improve the clarity of signal composition along with additional analysis.

We will clarify the description of the signal composition and the effect of noise removal models on each. As the reviewer points out, it is overly simplistic to view the measured signal as the sum of a clean signal and signal-independent noise. Instead, the measured signal should be more accurately described as the sum of the signal originating from light scattered by the particles, which is affected by signal-dependent shot noise, and signal-independent background noise. In our denoising process, we can only remove signal-independent background noise using a noise model pre-trained with data from flowing PBS. On the other hand, signal-dependent noise, such as photon shot noise, is not captured by our noise model and thus cannot be removed. In other words, noise sources that are not present in the PBS, e.g. smaller nanoparticles, or proteins, are not modeled and therefore not removed. We have clarified this with the following text, which is included on lines 94-102 of page 3 of the revised main text:

“Our denoising approach assumes that a measured noisy time series, $x = s + n$, $x \in R^T$, is the sum of the measured signal intensity originating from the light scattered by particles (particle signal), $s \in R^T$, and signal-independent background noise, $n \in R^T$, where T is the time series length. The particle signal s is affected by shot noise with a variance that depends linearly on the unknown underlying clean signal. The background noise n has a fixed variance and is independent of the particle signal³⁰, and we therefore describe it as sampled from probability distribution $p(n)$. Our denoising method removes signal-independent noise to recover the particle signal. Even though it is currently not able to remove the remaining shot noise, we experimentally show in the following section that this approach leads to improved sensitivity.”

“[30] de Rond, L. *et al.* Quantification of Light Scattering Detection Efficiency and Background in Flow Cytometry. *Cytometry* **99**, 671 (2021).”

Furthermore, to experimentally investigate the aforementioned signal-composition and effectiveness of our denoising model, we further conducted an additional analysis viewing the variance of the attenuated signal intensity before and after denoising as a function of the corresponding ground-truth signal intensity. To avoid redundancy, we here explain the details of the analysis by including the content we added from line 211 to line 231 on page 6 of the revised main text as it is. The results show that photon-derived shot noise was the main component of residual noise after noise removal.

“Additionally, we analyzed the attenuated signal variance before and after denoising as a function of the ground-truth signal intensity to evaluate how closely DNM approaches the shot noise limit. We consider the shot noise limit to be the state in which there is only noise that depends on the optical signal and whose variance is proportional to the ground-truth signal intensity. In the analysis, we first max normalized the ground-truth time series, then binned these normalized ground-truth signal intensities into 100 intervals. For each bin, we calculated the variance of the attenuated signal at corresponding time points. We plot the results of this analysis in **Figure 2(f) (R4_1 in this response)**.

Before denoising, the variance of the signal from the attenuated arm increases approximately linearly with ground-truth signal intensity, matching the behavior of shot noise. The y-intercept of the line – the variance of the noise when the ground-truth signal intensity is zero – is approximately equal to the variance of the PBS-derived fluidic noise. Since the variance of the sum of two independent random variables is additive³⁶, this indicates that the noise in the attenuated time series is the sum of the PBS-derived fluidic noise and shot noise. To make this clear, we subtracted the variance of the

PBS-derived fluidic noise from that of the attenuated time series and plotted it as the dashed line in **Figure 2(f) (R4_1 in this response)**.

We then repeated this analysis on the variance of the denoised signal. The slope of the variance here is close to the slope before denoising, indicating that the shot noise indeed remains. However, the y-intercept is now close to zero, indicating the removal of the PBS-derived fluidic noise. The curve slightly deviates from the ideal linear shape due to imperfections in the denoising result.”

“[36] Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P. & Meester, L. E. Covariance and correlation. 135–150 (2005) doi:10.1007/1-84628-168-7_10”

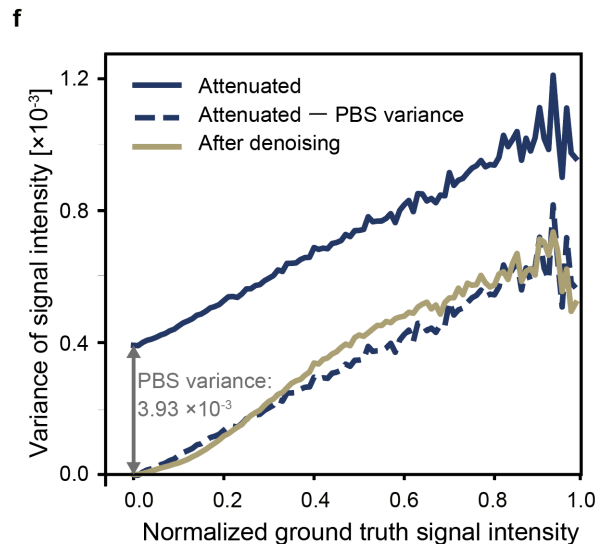


Figure 2(f), Figure R4_1 in this response:

The variance of the attenuated signal before denoising (blue line) and after denoising (olive line) as a function of the corresponding normalized ground truth signal intensity. The dashed blue line is the variance of the noisy attenuated signal minus the variance of signal-independent PBS-derived fluidic noise.

Background noise analysis:

In addition, we compare the noise level in the complex sample, which is serum in our case, with that in PBS. As a result, our histogram analysis shows that the noise distributions in both serum and PBS are quite similar, indicating minimal differences made by background particles in our measurements (**Figure R4_2**). The reason why the background is not affected by the small particles and proteins from the serum could be because we sufficiently dilute the serum with clean, particle-free water.

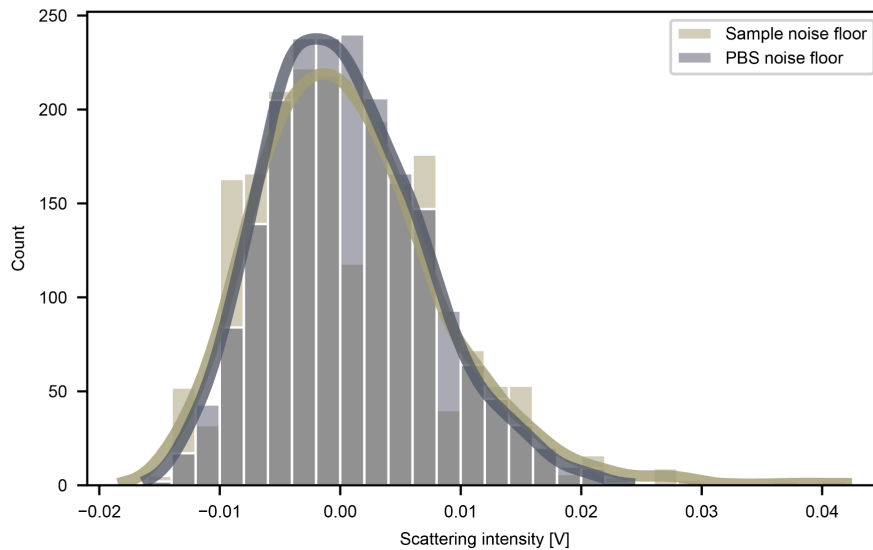


Figure R4_2:

The measured scattering intensity histogram and kernel density plot of PBS noise floor and serum noise floor. The two distributions are similar, indicating minimal differences made by background small particles, such as proteins, in the serum.

Major Comment 2

Regarding the random sampling in the consensus model. The authors describe a consensus model where 100 denoised signals are sampled and the point-wise median as taken as the consensus. Is it possible to quantify or comment on the confidence or uncertainty in choosing such an approach. Namely, what is the confidence that such a particular peak is assigned to noise vs signal. The authors should also briefly comment on the reasoning behind choosing the median to represent whether a portion of the data is assigned to signal or to background noise. For example in figure 3a, the time trace before and after denoising is shown, yet some peaks that look like possible signals are assigned to noise, this begs the question of what is the uncertainty or rather confidence probability of such assignment.

Response (major) 2

We thank the reviewer for their insightful question regarding why we chose the median instead of the mean and the uncertainty and confidence when selecting the median.

One of the reasons why we chose the median as a consensus statistic is its robustness against outliers. By using the median of 100 denoised samples, we aim to capture the most consistent behavior of the signal while minimizing the influence of random fluctuations caused by noise. The median ensures that the result represents the central tendency of the data, and will be stable as long as the number of outliers lies below 50%. In contrast to the median, the mean of samples is more sensitive to outliers. If only a few percent of the denoised samples have a sufficiently high intensity, this may distort the mean and lead to (false) detections even when there is only a negligible predicted probability of a particle being present.

Regarding the confidence in distinguishing signal from noise, the median approach inherently reflects the prediction confidence of the model. For any given time point, if the median of the samples lies above the detection threshold, at least 50% of the predictions suggest the presence of a signal. This provides a probabilistic measure of confidence: the model estimates that it is more likely that there is a particle than not. To quantify uncertainty

further, one could adjust the percentile used for the consensus. For example, choosing the 10th percentile would provide more conservative results, requiring 90% of the samples to be above the detection threshold to indicate a signal, thus increasing confidence at the cost of potentially missing weaker signals. Conversely, using the 90th percentile would lower the confidence threshold, increasing sensitivity but raising the risk of false positives.

Based on the discussion above, we have added the following text to lines 118-134 of page 4 to explain how the median of denoised samples can be interpreted.

“For evaluation, the random denoised samples from a trained signal model must be aggregated into a single consensus solution with which peak detection can be performed. We chose to use the point-wise median of the samples as it can be used to interpret the model’s confidence for this task. At each time point, denoised random samples can be classified as either above or below the peak-detection threshold. If the median at that point in time is above the threshold, at least 50% of the denoised samples must be above. Therefore, DNM has predicted the probability distribution over denoised solutions to have at least 50% of its probability above the detection threshold. In other words, the model believes it is more likely that there is a particle than not.

For more certainty from the model, a lower percentile, for example the 10th, could be used as a consensus. This would demand that the model place 90% of its samples above the detection threshold to reveal a signal peak. Inversely, a higher percentile, for example the 90th, would require that the model place only 10% of its samples above the detection threshold to reveal a signal peak.

In contrast to the median, the mean of samples is more sensitive to outliers. Even if only a few percent of the noise-removed samples have a sufficiently high intensity above the detection threshold, this could distort the mean and lead to false detections.”

Major Comment 3

Regarding the quantitative nature of the denoised time traces. In Figure 3d show the retrieved scattering amplitude signals after denoising and argue that these are quantitatively valid because these follow a size^6 scaling according to Rayleigh scattering theory. However the authors fail to convincingly show whether such denoising skews the retrieved scattering amplitude distribution. For instance figure 2c shows the distributions of the ground truth, attenuated vs denoised of the same nanoparticle sample, yet it is visible that the shape of the distribution of the scattering height is affected between ground truth and denoised.

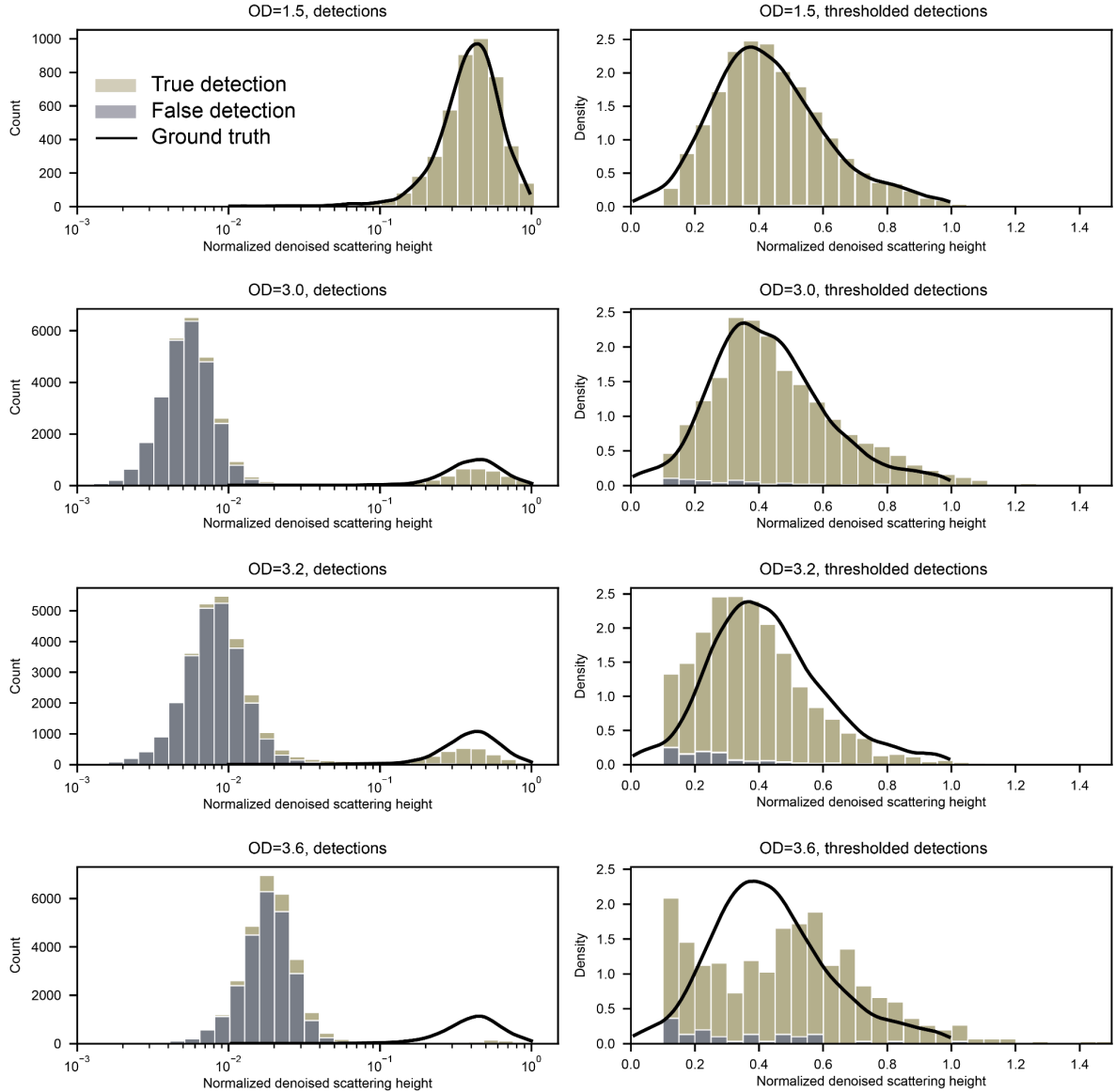
Furthermore the authors should explain the source of the uncertainties in the scattering amplitude scaling shown in Figure 3d; and test whether the denoising algorithm has any affect whatsoever. In the context of imaging, machine learning has been shown to be extremely good at solving segmentation, classification and detection problems, whilst showing mixed results for quantitative extraction of information. The authors have yet to show whether the quantitative aspect of the peak amplitude (scattering and fluorescence) is conserved by this algorithm and whether this affects all types of signals equally.

Response (major) 3

We appreciate the reviewer's insightful comments regarding the quantitative evaluation of the retrieved scattering amplitude distribution.

We performed an additional analysis, which is similar to the one shown in **Figure 2**, to investigate the potential distortion caused by denoising. We split the scattered light using a beam splitter to simultaneously measure the high-SNR time series in one arm and the low-SNR time series in another arm with different attenuation levels.

After detecting peaks in the denoised time series at different attenuation levels and those in the high-SNR time series, we compared the normalized peak height distributions. We assigned true and false detection labels of olive and blue, respectively, to the detections in the denoised data based on corresponding ground truth peaks. **Figure R4_3 left** shows stacked histograms with the x-axis on a logarithmic scale. As a result, we observe that the two distributions match well for attenuation OD 1.5 (median SNR of 107.77 before denoising), indicating that the denoising process preserves the scatter amplitude distribution under this condition. For the attenuation OD of 3.0 and 3.2 (median SNR of 3.05 and 2.97 before denoising, respectively) the distribution exhibits a bimodal pattern. The right mode corresponds to peaks that have been preserved by denoising and contains mostly true detections. The left mode corresponds to peaks that have been flattened by the denoiser and contains mostly false detections. With increasing OD, we observe that more and more true peaks are lost (incorrectly flattened). **Figure R4_3 right** shows a thresholded version of the data, rejecting detections below 0.1, as it may be done in a real applied setting to remove false detections. We show normalized probability density estimates with a linear scale on the x-axis. The result confirms that the general shapes of the distribution, which an experimenter without access to ground truth would measure, are well preserved at least for OD 1.5, 3.0, and 3.2 and less well preserved for the most extreme OD 3.6 (median SNR of 1.73).



Supplementary Figure S5, Figure R4_3 in this response:

Histograms of normalized denoised scattering peak heights at optical densities (OD) of 1.5, 3, 3.2, and 3.6, overlaid with a kernel density estimate (KDE) plot of the ground truth. We consider peaks with a height >0.01 in the ground truth series true detections (olive) and all others false detections (blue). A black line represents the distribution of detections in the ground truth series. **Left panels** show stacked histograms with the x-axis on a logarithmic scale. We observe that the two distributions match well for attenuation OD 1.5 (median SNR of 107.77 before denoising), indicating that the denoising process preserves the scatter amplitude distribution under this condition. For the attenuation OD of 3.0 and 3.2 (median SNR of 3.05 and 2.97 before denoising, respectively) the distribution exhibits a bimodal pattern. The right mode corresponds to peaks that have been preserved by denoising and contains mostly true detections. The left mode corresponds to peaks that have been flattened by the denoiser and contains mostly false detections. With increasing OD, we observe that more and more true peaks are lost (incorrectly flattened). **Right panels** show a thresholded version of the data, rejecting detections below 0.1, as it may be done in a real applied setting to remove false detections. We show normalized probability density estimates with a linear scale on the x-axis. The result confirms that the general shapes of the distribution, which an experimenter without access to ground truth would measure, are well preserved at least for OD 1.5, 3.0, and 3.2 and less well preserved for the most extreme OD 3.6 (median SNR of 1.73).

Based on the analyses above, we have added **Figure R4_3** as **Supplementary Figure S5** and the following sentence to the main text in lines 204-209 of page 6 to clarify this point:

"In addition, by comparing the height distributions of peaks detected in the denoised time series and those in ground truth time series, we found that the general shape of the intensity distribution after denoising is preserved for SNRs ranging from high (107.7) to moderately weak value of 2.97 (see more detailed comparison in **Supplementary Figure S5**)."

Regarding the scattering amplitudes and the associated uncertainties in **Figure 3d**, we thank the reviewer for pointing out the need to discuss the source of uncertainties and the effect of denoising on the recovered signal amplitudes. Based on the following analysis performed on the data shown in **Figure 3d**, we would show that the major source is the intrinsic variation of the bead size^{1,2,3} (27 ± 3.8 nm, 40 ± 6.7 nm, and $60 \text{ nm} \pm 10.8$ nm), and the denoising effect helps us recover the original distribution of particles.

In **Figure R4_4**, we present the experimentally measured scattering amplitudes (height) before (raw) and after denoising for comparison. We set the detection threshold value, below which the 99.9% percentile of measured values in the raw PBS time series and the denoised PBS time series are included, respectively. Here, using Rayleigh scattering theory⁴, the scattering amplitude for small particles is expressed as

$$\begin{aligned} \text{Scattering amplitude } (d) &= A \times \sigma_{\text{scatt}}(d, n_{\text{med}}, n_{\text{particle}}, \lambda) \\ &= A \times \frac{2\pi^5 d^6 n_{\text{med}}^4}{3\lambda^4} \left| \frac{m^2 - 1}{m^2 + 2} \right|^2 = cd^6, \quad (1) \end{aligned}$$

where σ_{scatt} is the scattering cross-section of the particle, $m = n_{\text{particle}}/n_{\text{med}}$, d is the particle size, n_{med} is the refractive index of the medium, n_{particle} is the refractive index of the particle, and λ is the excitation wavelength. The scattering amplitude is proportional to the scattering cross-section, scaled by a factor A . All of these, except d , are parameters of the experimental setup and can be reduced to the constant factor c . In the fitting, using the medians of the experimentally obtained scattering amplitudes for 40 nm and 60 nm particles, respectively, we estimated $c = 3.5 \times 10^{-11}$ in the **equation (1)**. We have described the Rayleigh scattering theory and how we determined the factor c in **Supplementary Note 4**.

Due to variations in bead sizes, there is uncertainty in the expected scattering amplitudes. Using the manufacturer's mean and standard deviations for bead sizes (assuming a normal distribution), we applied **equation (1)** with $c = 3.4 \times 10^{-11}$ to estimate expected amplitude distributions and visualize 50%, 70%, and 90% credible intervals. Finally, we overlaid swarm plots showing the measured peak heights from raw and denoised data (**Figure R4_4**).

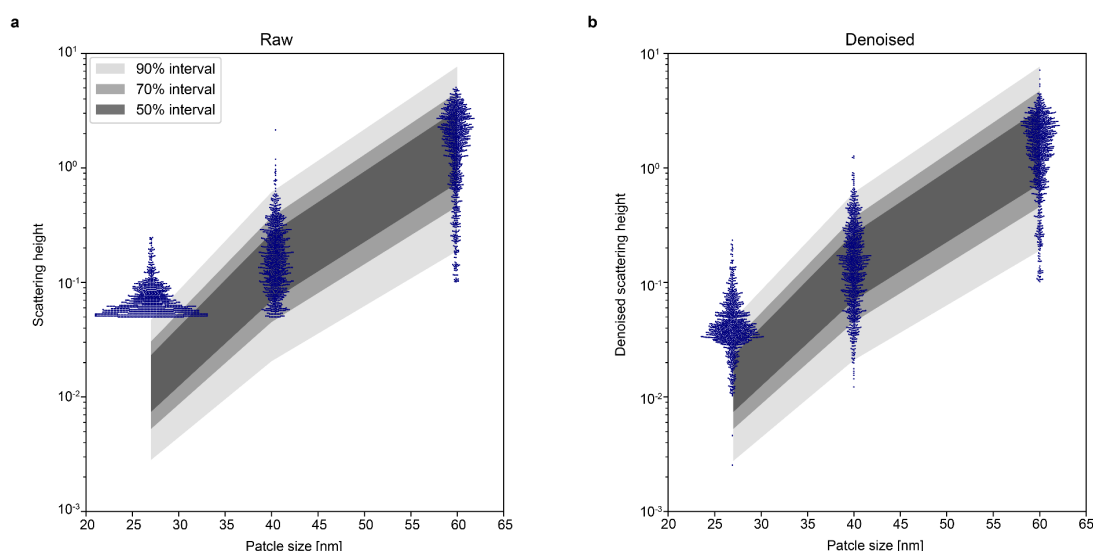


Figure R4_4

The swarm plot of (a) raw and (b) denoised scattering heights for particle sizes 27, 40, and 60 nm. The grey-shaded areas represent the 90% (5th to 95th percentile), 70% (10th to 90th percentile), and 50% (25th to 75th percentile) credible intervals estimated from the particle size distribution provided by the manufacturer. The respective percentiles in the bead size distribution were then mapped to the corresponding percentiles in the theoretical scattering height. In detection, we set the detection threshold values below which the 99.9% percentile of measured values in the PBS time series and the denoised PBS time series are included, respectively.

Noise removal has allowed us to detect small peaks for the difficult-to-detect 27 and 40 nm beads. As a result, the peak height distributions after denoising agree better with the theoretical intervals. This comparison suggests that detection performed on the raw data can lead to systematic errors, which are reduced when using denoised data. Also, the agreement between the theoretical interval and the scatter height in **Figure R4_4** indicates that the main source of scatter uncertainties, which is represented by the large variance, is due to the intrinsic size distribution of the beads.

In addition, our method removes only signal-independent noise and is currently not suitable for fluorescence data where signal-dependent shot noise is dominant. We discuss this aspect in detail in **Responses (major) 4 and 5**.

References

1. <https://www.thermofisher.com/order/catalog/product/jp/ja/F8787>
2. <https://www.thermofisher.com/order/catalog/product/jp/ja/F8795>
3. <https://www.thermofisher.com/order/catalog/product/5006A>
4. Bohren, Craig F., and Donald R. Huffman. *Absorption and scattering of light by small particles* (John Wiley & Sons, 2008).

Major Comment 4

Regarding the training set. The training set for denoising corresponds to a 1s time-series of ultrapure water. This seems like a valid approach for training the model on the noise contributions that are sample-independent. The authors should discuss how this extends to more complex samples where sample-dependent noise cannot be ignored.

Response (major) 4

Thank you for the insightful comment. We agree that tackling sample-dependent noise would be highly impactful. While noise coming from non-targeted smaller particles or proteins can be experimentally minimized simply by dilution as shown in **Response (major) 1**, there is no doubt that it is important to address data that is strongly affected by signal-dependent shot noise. An example of such data is the fluorescence data we discuss in **Response (major) 5**.

As the reviewer understands and as we explained in **Response (major) 1**, our noise model is trained using examples of noise where the underlying signal intensity is zero, i.e. ultrapure water, enabling it to capture signal-independent noise statistics. This in turn enables the VAE to remove signal-independent noise. In a naive approach to removing signal-dependent noise, the noise model would have to capture the noise statistics for a range of underlying clean signals. This would require training it with examples of noise with a known underlying clean signal, i.e. paired data of corresponding noisy and ground-truth time series, which is impossible to obtain in many practical scenarios including our case of nanoparticle detection near the detection limit.

A promising approach to removing signal-dependent noise could lie in a variant¹ of our approach which was recently developed for signal-dependent noise in images. We are currently trying to determine whether it is possible to adapt this method to time series data, such as fluorescence signals in flow cytometry.

Based on the discussion above, we have added the following sentence in the discussion section in lines 494-499 of page 17 of the main text:

“An important milestone for future work will be developing a method capable of additionally removing signal-dependent shot noise. While in this work, the sensitivity was significantly improved by removing only the signal-independent noise as it was dominant in the scattered signal, we also observed that the signal-dependent noise in fluorescence data, for example, was too intense to ignore. To this end, a promising solution could be to extend another version⁵³ of our denoising approach, originally developed for removing signal-dependent noise in images.”

“[53] Salmon, B. & Krull, A. Unsupervised Denoising for Signal-Dependent and Row-Correlated Imaging Noise. arXiv preprint arXiv:2310.07887v2, (2024). Cited as [1] in this Response (major) 4.”

Major Comment 5

Regarding the applicability between scattering and fluorescence signals. The authors should comment if there is any reason why the denoising algorithm would be affected whether the detected signal is from scattering or fluorescence. Specifically the authors state in the SI that the “noise model is autoregressive because the scattering noise is temporally correlated” as shown in figure S2. Given that the fluorescence signal is both a spatially and temporally incoherent light source, would this model still be suitable?

Response (major) 5

Thank you for the question regarding the applicability of our denoising algorithm to scattering and fluorescence signals. We will clarify the requirement of our denoising methods and potential effects of the coherence of light sources and correlation in time series. After that, we discuss the reason why our current model is not applicable to fluorescence time series.

We first apologize for our misleading description of the required assumption in our denoising algorithm. The formulation of an autoregressive noise model does NOT require temporal correlation but could learn to model uncorrelated noise as a special case. In image analysis, it has been shown that spatially uncorrelated noise can be successfully removed using

variants^{1,2} of our approach. For clarity, we will include the following paragraph in the **Supplementary Note 1**:

“The noise model is autoregressive because the scattering noise is temporally correlated, i.e., the value of a noise realization at one time point depends on the values of the noise realizations at the previous time points. This can be confirmed by examining the autocorrelation plot of the noise data (**Supplementary Figure S3**), where we observe correlations between different noise realizations. However, the formulation of an autoregressive noise model does not require temporal correlation but could learn to model uncorrelated noise as a special case. In image analysis, it has been shown that spatially uncorrelated noise can be successfully removed using variants^{1,2} of our approach.”

“[1] Salmon, B. & Krull, A. Unsupervised Denoising for Signal-Dependent and Row-Correlated Imaging Noise. *arXiv preprint [arXiv:2310.07887v2](https://arxiv.org/abs/2310.07887v2)*, (2024). Cited as [1] in this Response (major) 5.
[2] Prakash, M., Delbracio, M., Milanfar, P. & Jug, F. Interpretable Unsupervised Diversity Denoising and Artefact Removal. *ICLR 2022 - 10th International Conference on Learning Representations* (2021). Cited as [2] in this Response (major) 5.”

In addition, we discuss the potential effects of the coherence of light sources and temporal correlation of fluorescence time series while emphasizing again that they will not affect our model since it does not require temporal correlation of noise to work.

Regarding the effects of coherence of light in the correlation of time series, we did not observe them experimentally. Since we detect 1D temporal signals integrating the spatially distributed photons with a single pixel detector, we don't think that the spatial coherence of the light source affects our denoising algorithm. In addition, in the temporal domain, the optical frequency is orders of magnitude higher than the detector bandwidth.

Regarding the temporal correlation, the arrival of individual photons and the resulting shot noise are temporally uncorrelated in theory, independently of the coherence or incoherence of light. In practice, however, there is a smoothing effect due to the limited bandwidth of the detector which turns each photon detection into a peak that extends over multiple samples in time (see **Figure R4_5**).

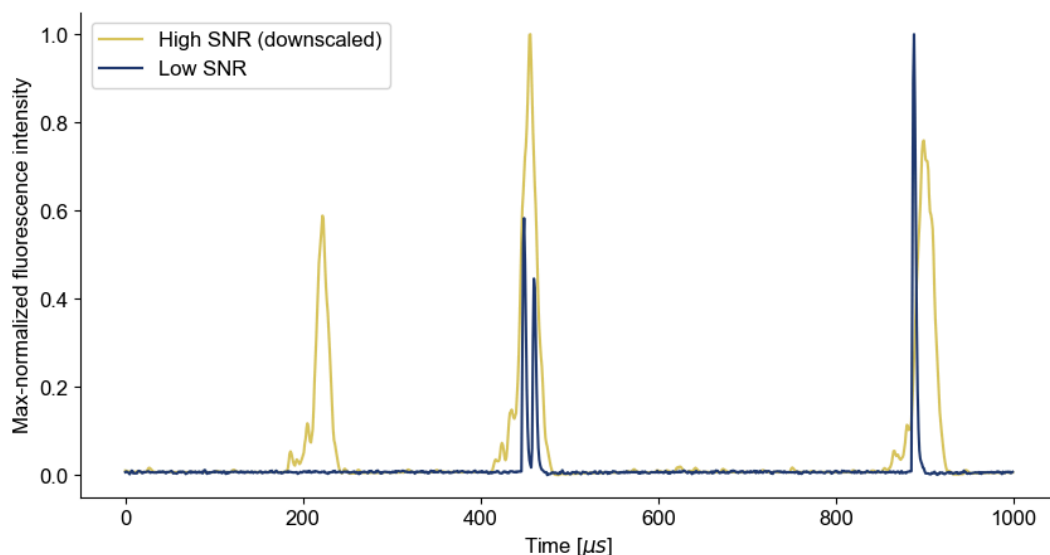


Figure R4_5
Example of corresponding fluorescence time series with high and low signal-to-noise ratio (SNR), captured simultaneously using the two-arm set-up.

Nevertheless, we admit that the model we show in this work is currently not applicable to the fluorescence signals we observed. The main difference between these two types of signals in denoising is the dominant noise sources. Our method assumes that the noise is signal-independent, which is the case for the dominant noise source in the scattering signals, allowing the denoiser to work effectively. In contrast, the fluorescence signals we observed are mainly affected by signal-dependent noise such as photon shot noise.

Figure R4_5 shows the example high and low signal-to-noise ratio (SNR) fluorescence time series captured simultaneously using the two arms split by the beam splitter. While the high SNR signal is the result of many photons, we suspect that the low SNR signal shows individual photons as individual peaks. The high SNR time series reveals three particles, but the first particle is missed entirely in the low SNR time series. This is because the number of photons arriving at the detector follows a Poisson distribution and there is a chance of zero photons arriving at the detector for a low intensity level. Since the mean and variance of Poisson distribution depend on the ground-truth signal, this type of noise can currently not be removed by our method.

To remove this signal-dependent noise, we believe that the unsupervised approach¹ that was developed for signal-dependent noise in images could lead to a solution if an extension to time series data can be discovered (see **Response (major) 4**).

Minor Comment 1

The manuscript could use a quick polish to amend typos found in the figure: concensus vs consensus.

Response (minor) 1

We thank the reviewer for pointing out the typo in the figure. We have corrected "concensus" to "consensus" to ensure accuracy in the manuscript. We appreciate your careful review and attention to detail.

Minor Comment 2

The author should comment on the conditions that lead to the setup running at $1e5$ events/s and state whether this same throughput rate was generally maintained throughout the different measurements shown in the figures in the manuscript. For instance, Figure 4b claims a 4 minute trace. If the setup were running at $1e5$ events/s this would lead to about $6e6$ events/min, yet only about $1e4$ EVs are detected at the highest serum concentration. Would it thus be correct that only 1% of the events detected correspond to EVs. If so, what do the other events correspond to? Perhaps, background signal.

Response (minor) 2

We appreciate the reviewer's comment on our setup operating at $1e5$ events per second. After more rigorous assessment in response to another reviewer, the detection throughput was estimated higher than $2e5$ events per second.

This throughput rate represents the system's maximum capability and, as the reviewer points out, varies depending on sample concentration. In other words, the higher the serum concentration, the greater the number of particles detected per second in proportion. In **Figure 4b**, we measured diluted serum samples and detected only about $1e4$ EVs over a 4-minute trace due to the lower concentration. Additionally, we think that particles other than EVs in the serum were also detected. We have clarified this in the manuscript by adding the following sentence in lines 324-325 on page 11:

“Note that this detection throughput varies depending on the input concentration of particles.”

=====

We believe that we have thoroughly addressed the reviewer's comprehensive concerns. We hope the reviewer finds the updates satisfactory and look forward to further evaluation.

We again appreciate the reviewer's valuable feedback and for helping us improve our work.

Best regards,
Authors

Reviewer #1:

Overall comments

The authors have developed an unsupervised deep learning-based denoising method to realize a nanoparticle analyzer with unprecedented throughput given its sensitivity. The technology, named “Deep Nanometry” (DNM), is able to detect 30-nm polystyrene spheres at a rate of 100,000 events/s. For comparison, a few commercially available state-of-the-art flow cytometers can detect ~70 nm polystyrene beads at a rate of ~30,000 events/s (in practice). Machine learning can resolve dim signals from the noise where humans cannot even spot the signals. Machine learning is probably able to accomplish denoising because noise sources and optical signals originating from particles passing through the laser beam both have characteristic behavior in time. Nevertheless, a 6 dB improvement in the signal-to-noise ratio is impressive and significant when aiming to detect dim nanoparticles. The authors applied DNM to measure the concentration of extracellular vesicles (EVs) in serum from cancer patients and healthy donors. In my opinion, the authors have well-addressed my previous major and minor comments and I particularly appreciate the added noise analysis. I have just a few more minor comments left, please see below.

Response to overall comment

We sincerely appreciate the positive evaluation of the reviewer on our revised manuscript. We are particularly delighted that the reviewer appreciated the added noise analysis, which has further deepened our understanding and will contribute to developing more advanced noise removal approaches in the future. Additionally, we would like to thank the reviewer for the further suggestions on revisions, each of which helped us further clarify our claims and strengthen the overall argument of our paper.

Minor Comment 1

In Figure S10 the authors investigate the detection limit of DNM. For that they use 40-nm polystyrene beads, with a specified median diameter of 40 nm and a standard deviation of 6.7 nm. In contrast to the standard deviation, only the specified median diameter is traceably determined, i.e. with an uncertainty statement. The standard deviation is rather the maximum possible standard deviation according to the manufacturer. NIST-traceable beads from ThermoFisher typically have a smaller standard deviation than specified. It seems that your dataset confirms my statement. However, no action needed here, as it is difficult to obtain the “true” standard deviation.

Below 30 nm, the data does not follow the specified curve, hence the authors define the detection limit as 30 nm for polystyrene spheres. I agree with that approach. Next, however, the authors claim “DNM is able to analyze 30 nm polystyrene beads at over 100,000 events/second, scaling to millions of events.” and “DNM can analyze 30 nm polystyrene beads at a throughput of over 100,000 events/second”. I do not agree with the latter statements, because 30-nm polystyrene beads also have a considerable standard deviation and given that the detection limit is 30 nm for a polystyrene sphere, half the population of 30-nm polystyrene beads would fall below the detection limit. Technically, DNM is therefore able to detect 40 nm polystyrene beads and not 30 nm polystyrene beads. I guess this is a matter of phrasing the text well to avoid an invalid claim.

Response (minor) 1

We appreciate the reviewer's comment pointing out the potentially misleading statement. We are pleased that the reviewer agreed with the approach we have taken to estimate the limit of detection (LoD) and with our conclusion that the LoD is 30 nm.

We acknowledge the reviewer's concern about the phrase "can analyze 30 nm polystyrene beads," which may mislead the readers into thinking that the median diameter of the bead populations is 30 nm. As the reviewer correctly noticed, this is not the interpretation we intended, and we aim to avoid any such misunderstanding.

On the other hand, even if a portion of the 30 nm bead population falls below the detection limit, DNM may detect the remainder of the population and meet the criteria of the IUPAC definition of the limit of detection (LoD): the probability of correctly identifying a true signal should be at least 0.99, and the recovery rate should be at least 50% [1,2,3]. Based on this, we believe it is reasonable to reference the detection of 30-nm beads. However, we agree that the phrasing should better reflect the nuances of this interpretation.

To address the reviewer's concern while considering the discussions above, we have revised the manuscript to provide a more accurate and clear description of our findings. The updated claims are as follows in line 27-28 of page 1:

"DNM detects polystyrene beads with a detection of limit of 30 nm at a throughput of over 100,000 events/second."

in line 64-66 of page 2:

"DNM detects polystyrene beads as small as 30 nm based on theoretical fitting of experimental measurements within the size distribution of 40 nm beads at a throughput of over 100,000 events/second."

in line 491-493 page 17 :

"DNM demonstrated the sensitive detection of polystyrene beads as small as 30 nm based on theoretical fitting of experimental measurements within the size distribution of 40 nm beads"

[1] Pan, Lei, et al. "Determination of X-ray detection limit and applications in perovskite X-ray detectors." *Nature Communications* 12.1 (2021): 5258.

[2] Currie, Lloyd A. "Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995)." *Pure and applied chemistry* 67.10 (1995): 1699-1723.

[3] Long, Gary L., and James D. Winefordner. "Limit of detection. A closer look at the IUPAC definition." *Analytical chemistry* 55.7 (1983): 712A-724A.

Minor Comment 2

The authors used rainbow calibration particles to calibrate MESF values for FITC, Alexa488 and Tide FluorTM245 5WS fluorescence. However, rainbow beads do not have FITC, Alexa488 and Tide FluorTM245 5WS molecules and the emission spectrum of rainbow

beads therefore differs from the fluorescent molecules used by the authors to label EVs. Possibly assigned MESF values of the manufacturer to the rainbow beads are therefore only valid if the authors used the same excitation laser and spectral filters as the manufacturer, which is unlikely the case. Hence the question: how did the authors perform the MESF calibration? Did the authors consider possible differences between the fluorescence optics of the manufacturer and the authors?

Response (minor) 2

We sincerely thank the reviewer for the valuable comment regarding the calibration of Molecules of Equivalent Soluble Fluorophore (MESF) values using rainbow calibration particles. We acknowledge that the emission spectrum of the rainbow beads differs from the fluorescent molecules (FITC, Alexa488, and Tide Fluor™ 5WS) used to label EVs, as the beads do not contain these fluorophores. We also recognize the potential discrepancies between our fluorescence optics and those of the manufacturer. To address this, we have added a detailed explanation of our MESF calibration process and revised the manuscript accordingly.

Specifically, the number of fluorescent molecules per particle was derived from previous studies [1] and the manufacturer's specifications [2]. In the previous revision, potential differences between our optical setup and the manufacturer's conditions were not explicitly considered. Following the reviewer's comment, we carefully reviewed the manufacturer's calibration conditions and compared them with our own, as summarized in Table 1.

For FITC, the excitation wavelength and the spectral filters match between our case and the previous study [1], confirming that the assigned MESF values are valid (Table1). We omitted the comparison for Alexa488, as FITC is more frequently employed for MESF estimation [1] and sufficiently addresses the system's sensitivity for green fluorescence. However, for APC, the spectral filters are different [2], suggesting that our discussion of APC fluorescence could be influenced by unaccounted discrepancies. To address this potential issue and avoid any confusion, we have removed the corresponding discussion of deep red fluorescence in Supplementary Figure 13 and have excluded the figure from the Supplementary Data.

Table 1 The comparison of MESF calibration conditions [1,2]

		Ours	Maia, Joana (2022)	Manufacturer's Specification sheet
Instruments		DNM	Apogee A60	Beckman colter CytoFlex LX
FITC	Excitation	488 nm	488 nm	N.A.
	Laser power	500 mW	200 mW	N.A.
	Emission filter	525/50 nm	525/50 nm	N.A.
APC	Excitation	637 nm	N.A.	638 nm
	Laser power	150 mW	N.A.	50 mW
	Emission filter	690/50 nm	N.A.	660/10 nm

We also have added the following description in the caption of Supplementary Figure 13:

“Here, the number of fluorescent molecules per particle was derived from previous study¹⁶.”

References

[1] Maia, J. *et al.* Employing Flow Cytometry to Extracellular Vesicles Sample Microvolume Analysis and Quality Control. *Front Cell Dev Biol* **8**, 593750 (2020).

[2] <https://www.thermofisher.com/order/catalog/product/jp/ja/F8795>

Minor Comment 3

In the Discussion, the authors ambitiously wrote: “An important milestone for future work will be developing a method capable of additionally removing signal-dependent shot noise”. Shot noise originates from the fundamental quantum nature of light, specifically the discrete and probabilistic arrival of photons. This noise is inherent to electromagnetic radiation and the detection thereof and cannot be removed. Experimental limitations induced by shot noise can in general be mitigated by increasing the number of collected photons. In my opinion, the milestone for future work to “removing signal-dependent shot noise” seems therefore impossible.

Response (minor) 3

We thank the reviewers for insightful feedback on our future research goals. As the reviewer correctly noted, photon shot noise is a fundamental form of noise that scales with the square root of the number of photons, and its complete elimination is inherently challenging. Nonetheless, there are studies [1,2,3] in the fields of image and video analysis that successfully mitigate shot noise by exploiting spatiotemporal and spectral continuity. While the challenge becomes greater for one-dimensional data due to reduced available peripheral information compared to video and images, we believe that with advanced algorithms and careful signal processing, significant suppression of shot noise may still be achievable.

[1]Salmon, B. & Krull, A. Unsupervised Denoising for Signal-Dependent and Row-Correlated Imaging Noise. (2023).

[2] Li, Xinyang, et al. "Real-time denoising enables high-sensitivity fluorescence time-lapse imaging beyond the shot-noise limit." *Nature Biotechnology* 41.2 (2023): 282-292.

[3]Eom, Minho, et al. "Statistically unbiased prediction enables accurate denoising of voltage imaging data." *Nature Methods* 20.10 (2023): 1581-1592.

Reviewer #2:

Overall comments

The authors have provided a thorough response to the questions raised in the initial review. The latest manuscript shows clear improvements and has addressed all my concerns. I recommend the publication of the manuscript.

Response to overall comment

We sincerely appreciate the reviewer's positive feedback and recommendation for publication. The reviewer's time and effort in providing thoughtful insights have significantly improved the depth and clarity of the manuscript.

Reviewer #4:

Overall comments

I'd like to commend the authors for their detailed, thorough and clear responses to all comments and suggestions raised not only by me but by the other reviewers. The revised manuscript has vastly improved and better articulates the structure, advantages and current limitations of the algorithm. It is this reviewer's opinion that the manuscript is ready for publication. I very much look forward to reading their future work, and implementing these algorithms in my own research.

Response to overall comment

We would also like to thank the reviewer for his encouraging comments. If the reviewer has any further interest or questions after publication, please feel free to contact us at any time. Once again, we thank both reviewers for their invaluable feedback and support, which we have no doubt that have greatly improved this work.

The paper presents a new technique termed DNM combining an unsupervised signal denoising algorithm with SP-FCM and demonstrating enhanced sensitivity in detecting and measuring various physical characteristics of protein biomarkers in serum compared to conventional SP-FCM. Despite existing neural network methods that have been proposed for single-photon flow data analysis (classification) and for imaging flow image denoising, there is now still a lack of effective one-dimensional flow data denoising. The author modifies and validates an unsupervised learning (neural network-based) signal denoising method for this field, which constitutes a major contribution. We appreciate predictably potential biomedicine applications of DNM, especially noninvasive detection and assessment of early cancer. Still, I have concerns about algorithms and readability of the paper.

Main:

The author first employs a CNN to learn and model the probability distribution of background noise data collected by imaging particle-free samples like ultrapure water, and then employs a VAE to remove noise with the learned probability distributions from practical imaging data (beads/patient's serum/in control). However, in PMT imaging of practical samples, signal-dependence shot noise (Poisson noise) dominates and must be well considered. However, in the first stage, only signal-independence noise is taken into account. This leads to a problem, which comes as my core concern, that the probability distribution learned in the first stage is not consistent with that in practice. It can result in inadequate denoising performance and weak generalization. Despite the effectiveness of the proposed algorithm in Results, there is a lack of sufficient comparison with other algorithms, especially with supervised deep learning methods. From an algorithmic perspective, the algorithm described by the author may remove part of shot noise that displays similar characteristics to environment/instrument/sensor noise in imaging of particle-free samples. However, the theoretical gap should be fully and analytically discussed.

The algorithm proposed is trained and executed in an unsupervised manner. Owing to incomplete or overconfident modeling of signals and noise, unsupervised denoising algorithms often display a degenerate performance (validation PSNR metrics for example) not as good as those in a supervised manner. For this reason, in most papers involving unsupervised denoising algorithms, results of unsupervised models and supervised models are carefully and fairly compared (on the same dataset, w/wo ground-truth labels). It is recommended to provide such an experimental analysis.

The authors claim that DNM achieves a high throughput of 100,000 events/second (Line 255-258). DNM is based on VAE, and its denoising execution requires multiple samples (Line 105). However, Is the denoising performed in realtime? How long does it take to process events generated within one second? What is the specific configuration (CPU, CPU memories, GPUs, system version, python/C++ version) of the workstation/computing server used in this study? Addressing these issues will make the algorithm part easy to follow.

Line 498-501 and Line 507-514. Loss function configurations of the two neural network models are ambiguous.

The algorithm proposed is central to the paper but I cannot find any available code/software attached. Whether and how the code/software and demon is available?

In the implementation, the noise model is trained with data obtained by recording with particle-free ultrapure water. Therefore, the noise distribution learned by the neural network should be instrument-specific and signal-independent. As a consequence, the denoising VAE seems to be incompetent in dealing with signal-independent noise.

In DNM, the one-dimensional noise is learned by a CNN, which is commonly used to deal with 2-D image data. Considering that scattering noise is temporally correlated, recurrent neural networks (e.g. LSTM) seems to be more suitable to capture the temporal correlation in time series.

More details about the training objective of the noise model should be provided.

Others:

In terms of the “voting” mechanism of the 100 samplings from the denoising VAE, why choose “median” instead of other statistical indicators such as “mean average”? And I wonder how diverse are the different samplings from the VAE?

Line 111 and Figure 1. Given that, in general, the excitation intensity and beam uniformity and stability of a 408 laser are not as good as those of a 488 laser, but with a higher risk of photodamage, I am confused as to why the authors chose a 408 laser to obtain SSC.

Line 498-501 and Line 507-514. It is strongly recommended to provide a neural network structure/architecture diagram (in an extended or supp figure, for example).

Line 530 and Equation 2. The super/sub-scripts of the sum symbol encompass $n+1$ datapoints instead of n .

Line 88-91. Suggest changing ‘address’ to another word. This is unproven in the paper and seems to be hard to prove/validate that with a trained VAE, solutions obtained by reparametrizing are consistent with the solution space of the original ill-posed problem.

Line 139-142. There are too many references, which may reduce the readability of the paper.

Figure 3. Suggest replacing the symbol of PBS with a dashed box.

Supp Note 1, Line 38. What is the distribution of $p_{\theta}(z)$?

Figure S2. I cannot understand why the autocorrelation curve contains a trough (Lag = 4 μ s).

Overall, I recommend Major/Minor revision.