Further Evaluation and Calibration of a Mie Scattering-based Apparatus for Portable Optical Haematological Analysis

Ourania J. Banti[†1], Maro Michailidou[†2], and Anna Dziuba[3]

Department of Industrial Engineering & Management, International Hellenic University, Greece[1],
Department of Electrical & Computer Engineering, Aristotle University of Thessaloniki, Greece[2],
Faculty of Computing and Telecommunications, Poznań University of Technology, Poland[3]

***Abstract*—This paper presents the second stage of data analysis for the "Sappho PDA"; a non-invasive, Mie scattering-based apparatus for portable haematological analysis via optics. In this paper, the dataset presented in [1], is further examined for the generation of statistical indicators and for the validation of previous results.**

***Index Terms*—Mie scattering, biomedical sensors, haematological analysis, statistical indicators, embedded systems, optics**

I. Mie Scattering

MIE scattering is the scattering of light as it travels through a medium with a uniform refractive index, like a homogeneous sphere with a diameter larger or equal to the wavelength of light used for the measurement. This scattering of the electromagnetic plane wave is described by the forced oscillations of the electric and magnetic fields according to Maxwell's equations. When analysing biological cells, such as red and white blood cells and platelets, Mie scattering becomes crucial, as these cells' sizes are on par with the light's wavelength. In biomedical applications, Mie's theory is crucial, as it accounts for the complex angle-dependent intensity of the scattered light, with forward scattering typically dominating. The Mie solution, consisting of an infinite series of spherical multipolar partial waves is especially applicable to these cellular components, which are similar in size to λ, and this approach remains valid up to sizes approaching the limits of geometric optics for large particles.[1]

Mie Scattering does not account for the varying morphologies of the cells under observation and considers all measured particles to be spheres; for instance, in our experiments, red blood cells, which are characterised by their biconcave disk shape from a lack of a nucleus, are nonetheless approximated as spheres instead of discoids. Still, even with that approximation considered, Mie Scattering has been the preferred standardised technique for measuring particles smaller than 50 μm since the 1990s, as delineated in the International Standard ISO:13320:2020. This factor establishes its suitability for the current device's operational methodology.[1]

II. The "Sappho PDA" Device

The portable and low-cost device presented is utilising Mie Scattering for detecting particles ranging from 1 μm to 15 μm in diameter. The device is consists of a 650 nm laser, a photodiode array TCD1103GFG produced by Toshiba with a resolution of 1500×1 pixels, the Beaglebone Black rev. C, a custom printed case, and a custom circuit board that facilitates the communication between the photodiode array and the microcomputer. A render of the aforementioned device is shown in figure 1. By measuring the occurring Mie

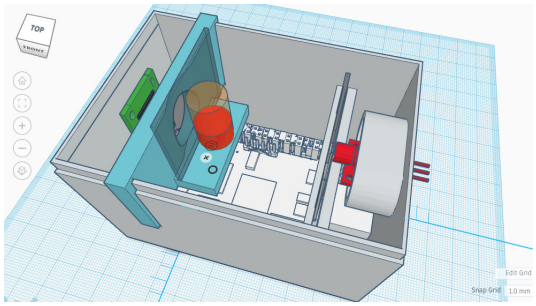

Figure 1: A render of a three-dimensional CAD file of the Sappho-PDA device, with the top cover removed

scattering from the measured cells with the photodiode array, the sensor's output can be examined with statistical analysis in order to approximate the quantity and size distribution of RBCs, WBCs and platelets present.[1][2]

The usage of calibrated solutions of a fixed diameter at 2, 4.8 and 9.6 μm was crucial in confirming the appropriate function of the device and its ability to measure the cells. The size and diameter of WBCs, RBCs and platelets is known from previous research to be about 12-16 μm, 7-8 μm, and 1.5-3 μm respectively, allowing the estimation of the amount of particles present in the solution along with their size.[1] The right half of the measured signal is illustrated in figure 2 for three randomly selected pairs, using particles of a standard diameter and distribution at a specific concentration, comparing unpolarised and polarised samples. The light intensity has been normalised – per sample – to a scale between 0 and 1, using min-max normalisation. Figure 3 illustrates a similar measurement, but for samples taken using aqueous solutions of plant-based beverages, rather than particles of a specific, standardised diameter.

By integrating these documented measurements with the experimental results obtained, it is possible to estimate the concentration and size distribution of the particles within the solution. During past analyses, the authors confirmed, with considerable certainty, that the signals contained in their datasets largely match the expected Mie scattering profiles

predicted by Mie scattering theory, with each experiment exhibiting satisfactory repeatability and accuracy, as previously verified by K. Karakostas.[3][4] Furthermore, the authors have also observed that even using low-cost linear polarisers may reduce optical signal interference and/or improve the measured Mie scattering pattern.[1][2]

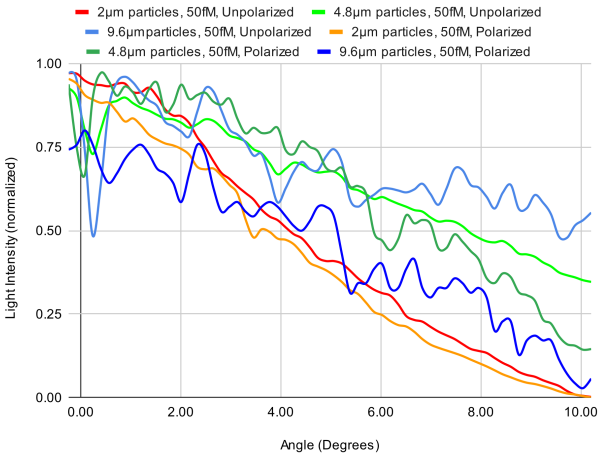

Figure 2: Three random pairs of unpolarised and polarised samples taken for standardised 2µm, 4.8µm, and 9.6µm particles at a 50fM concentration

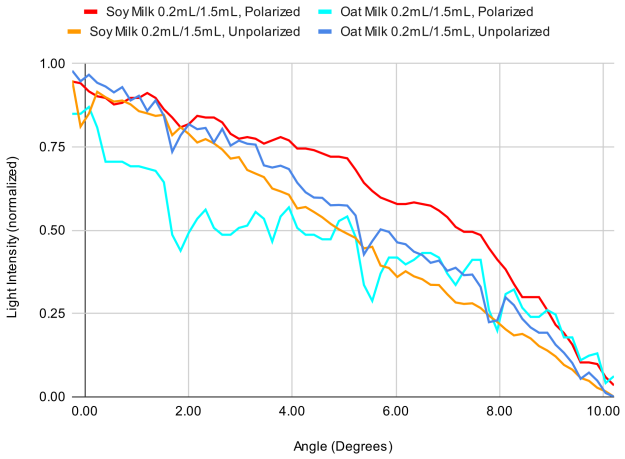

Figure 3: Two random pairs of polarised and unpolarised samples taken for a soy-milk aqueous solution and an oat-milk aqueous solution at a specific % v/v concentration

III. Further Analysis

During the aforementioned stage of analysis, the authors had primarily utilised elements of basic statistics and metrology.[1] Therefore, in this paper, the authors would like to present the second stage of their analysis, in which they've used more advanced algorithms in order to ascertain the accuracy of the previous stage of data analysis. Some of the following algorithms also involve data produced from simulations of Mie scattering theory. MiePlot was used for the generation of these signals.[1]

A. Coherence Test

The coherence test uses both simulated data and experimental data. The result is a single number between 0 and 1, where 0 indicates a total mismatch between the signals, and 1 indicates perfectly-matching signals. In this test, we used nine simulated signals — three for each particle size, for three specific detector distances.[2][5] The results of this test are indicated in table I.

Table I: Lowest and highest coherence ranges for different simulated particle sizes

Sim. Part.	Filter	Low Range	High Range
2µm	Unfiltered	0.05-0.20	0.35-0.60
2µm	2% Mov. Avg.	0.05-0.20	0.40-0.65
4.8µm	Unfiltered	0.10-0.25	0.40-0.60
4.8µm	2% Mov. Avg.	0.10-0.25	0.40-0.60
9.6µm	Unfiltered	0.10-0.25	0.40-0.70
9.6µm	2% Mov. Avg.	0.10-0.20	0.45-0.75

The coherence distribution for 2 µm and for 9.6 µm particles, at different simulated distances from the detector, demonstrates a consistent linear relationship. In contrast, coherence values for particles with a size of 4.8µm do not exhibit a similarly clear linear trend between different distances. The application of a moving average filter, with a size approximately 2% of the length of the original signal, does not mitigate this lack of linearity in the 4.8µm coherence values. Furthermore, the authors observed that due to the simplicity of the theoretical signal, some control signals may test positive as 2 µm signals.

B. K-means Clustering

The authors attempted to use two variations of *K*-means; the standard "naive" algorithm, and the *k*-means++ (initialised) method. For *k*-means++, the algorithm was first initialised with one simulated signal for each particle size, at a specific distance from the detector, for 3 clusters in total. This was later expanded to 9 clusters; one for each simulated scattering pattern mentioned in the header of this section. The results were modest: a marginal improvement over "random guessing" with the naive *K*-means, and approximately a 7% improvement with *k*-means++. Further exploration with alternative *K*-means implementations or other clustering algorithms is required for more significant improvements. Additionally, the utility of *K*-means could be reconsidered in different use cases, using arrays of results from other tests (i.e. "indicators") as inputs.[2][6]

C. Time-Domain Correlation Tests

The R^2 test was used as "goodness-of-fit" tests in the time domain, in pair with coherence being used as a goodness-of-fit test in the frequency domain. Results can be interpreted similarly to the coherence's results, but the values on the negative axis can be arbitrarily large, depending on the overall lack of determination between the sample and the simulation.[7] The results of this test are indicated in tableII.

In the assessment of R^2 tests at 4.8µm and 9.6µm, a clear linear relationship is observed with the respective coherence tests, indicating a consistent pattern within these wavelength categories. In contrast, the 2µm tests do not exhibit this linear relationship, potentially in part due to control signals testing positive as 2µm signals in the coherence test but not in the R^2 test. The "Pearson's r" test[5] reinforced these findings, providing similar results with variations in numerical values.

Table II: Lowest and highest R^2 distribution ranges for different simulated particle sizes

Particle	Filter	Max. Rng.	Min. Rng.
2μm	Unfiltered	0.75-0.86	(-4.54)-(-3.40)
2μm	2% Mov. Avg.	0.76-0.86	(-4.52)-(-3.41)
4.8μm	Unfiltered	0.67-0.83	(-4.59)-(-3.04)
4.8μm	2% Mov. Avg.	0.67-0.84	(-4.60)-(-3.05)
9.6μm	Unfiltered	0.65-0.82	(-6.92)-(-4.98)
9.6μm	2% Mov. Avg.	0.66-0.83	(-6.92)-(-4.98)

Overall, signals that test positive as a specific particle in coherence tests will often test positive in the R^2 test and the Pearson test.

D. Dynamic Time Warping

The Dynamic Time Warping test was used as an alternative time-domain test, both in order to verify the goodness-of-fit of the two signals, like above, and also as a way to cross-check the outputs of the previous chapter's time-domain tests. In this section, the result represents a measurement of distance between the two signals.[8] Table III presents the results of this test, paired with their corresponding coherence values.

Table III: Coherence and dynamic time warping data for different simulated particle sizes

Part.	Filt.	Min. Rng.	Coherence	Max. Rng.	Coherence
2μm	Unfil.	0.77-2.28	0.35-0.6	5.51-6.52	0.05-0.20
2μm	2% MA	0.75-2.24	0.40-0.60	5.52-6.52	0.05-0.20
4.8μm	Unfil.	1.58-2.64	0.40-0.60	6.33-7.49	0.10-0.30
4.8μm	2% MA	1.54-2.61	0.40-0.60	6.35-7.52	0.10-0.25
9.6μm	Unfil.	1.32-2.56	0.45-0.70	7.30-8.89	0.10-0.25
9.6μm	2% MA	1.26-2.49	0.45-0.75	7.32-8.92	0.10-0.20

Overall, the results exhibited a linear relationship with the results of the coherence test and the R^2 test, especially with the larger particle sizes. Furthermore, similarly to the previous test, control signals with high coherence did not test positive as 2μm waveforms, potentially giving the authors multiple ways to avoid such false-positive outputs.

E. Kolmogorov-Smirnov Test

Ultimately, the authors used the Kolmogorov-Smirnov test's p-value as a "confidence check" for the results of the aforementioned tests. Any p-value larger than or equal to 0.05 in this test indicates that there is a sufficient match between the simulated and experimental signals. In the results, 29.78% of the dataset had individually passed the test with a p-value $\leqslant 0.05$.[5] This is relatively in line with the expected results, under the assumption that not all samples are of equal quality, as approximately 38% of the dataset contains samples of individual calibrated solutions of fixed-diameter particles; 20% of the dataset contains various control samples; 15% of the dataset contains non-standardised aqueous solutions (such as solutions of soy milk, oat milk, salts, etc.); 12% contains in-vivo samples; 10% contains mixed solutions of fixed-diameter particles, and the remaining dataset contains data from minor experiments.

IV. Conclusions and Future Plans

Even though most of the above tests yielded promising results, there is still a requirement for further study. In specific, the authors are planning to experiment with the *wavelet transform* for pattern extraction; with *fuzzy logic* systems, and *supervised* and (other) *unsupervised* machine learning algorithms for data classification from indicators, and with further statistical analysis. In addition, the authors are also planning to implement some of the aforementioned statistical tests into the *K*-means algorithm itself; for example, implementing dynamic time warping rather than Euclidean distance during the training phase. However, a newer, better dataset will be critical for the development and evaluation of several of the above experiments. Last but not least, the authors are exploring enhancements to the physical device, particularly focusing on improvements in computational power, optical properties, and further compliance to European and global standards for biomedical devices, as previously outlined in [1].

References

[1] M. Michailidou, O. J. Banti, *Scattering measurements with the Beaglebone microcomputer for portable biomedical sensors*, International Hellenic University, 2022

[2] O. Banti, M. Michailidou, E. Gkagkanis, K. Karakostas and M. E. Kiziroglou, *Fabrication and Development of an Optical Biomedical Sensor*, International Workshop on Microsystems, 2021

[3] Karakostas K., *"Portable system development for Mie scattering analysis, to determine the size of blood cells in in-vivo and in-vitro studies"*, Aristotle University of Thessaloniki, 2019

[4] Konstantinos Karakostas, Stratos Gkagkanis, Korina Katsaliaki, Peter Köllensperger, Alkiviadis Hatzopoulos, and Michail E. Kiziroglou, *"Portable optical blood scattering sensor"*, Microelectronic Engineering 217 (2019) 111129, Elsevier

[5] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, CJ Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E.A. Quintero, Charles R Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors, *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python.*, Nature Methods, 17(3), 261-272., 2020

[6] Simon Haykin, *"Neural Networks and Learning Machines"*, McMaster University, Hamilton, Ontario, Canada

[7] Pedregosa et al., *Scikit-learn: Machine Learning in Python*, JMLR 12, pp. 2825-2830, 2011.

[8] Stan Salvador and Philip Chan, *FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space*, Florida Institute of Technology, 2007