

Hierarchical clustering of the precomputed signals database to solve the parametric inverse light-scattering problem

Artem R. Muliukov^{a,b} and Maxim A. Yurkin^{a,b,*}

^a*Voevodsky Institute of Chemical Kinetics and Combustion SB RAS, Institutskaya str. 3,
630090 Novosibirsk, Russia*

^b*Novosibirsk State University, Pirogova 2, 630090 Novosibirsk, Russia*

**Presenting author (yurkin@gmail.com)*

Light scattering is a powerful proxy for non-invasive characterization of single particles through the solution of the inverse light-scattering (ILS) problem. In most cases one a priori specifies the particle model with several free parameters, which are determined by non-linear regression, i.e. by fitting an experimental signal (typically, a light-scattering pattern – LSP) with the simulated ones. Finding a global minimum of the residual in the space of particle parameters may require thousands of simulations, making direct fit unrealistic.

This computational bottleneck is alleviated by preliminary computing a database of theoretical signals, using a super-computer, for specific class of particles. The ILS problem then boils down to finding the nearest neighbor in this database. A by-product of the direct search is the distances between an experimental signal and all entries in the database. They are further used in a statistical analysis to estimate not only the best-fit parameters, but also their confidence ranges (uncertainties) [1]. This approach has been successfully used to characterize several classes of biological cells in a flow using the LSPs measured with a scanning flow cytometer [1,2]. These cells were described by 4–5 parameters and databases had up to 6×10^5 entries. The processing of each experimental LSP took about 1 second on a desktop, which is still far from real-time, considering measuring speed of 100 particles per second.

The goal of this work is to further decrease the processing time by hierarchically clustering the precomputed database in the form of a binary tree. Knowing the radius of each cluster, i.e. the maximum distance between its element and a center, we can rigorously discard some of the clusters entirely during the nearest-neighbor search. First results show at least ten-fold decrease of number of distance calculations, and hence, of the processing time. Importantly, this reduced number of distances is sufficient for the above-mentioned statistical analysis, approximating the missing distances by that to the center of the corresponding discarded cluster. Overall, the clustering decreases the computational complexity of ILS solution down to almost logarithmic in the database size, which potentially allows much larger databases (hence, better accuracy) with small extra computational costs (apart from one-time investment for the database itself).

References

- [1] Moskalensky, A.E., *et al.*, 2013: Accurate measurement of volume and shape of resting and activated blood platelets from light scattering. *J. Biomed. Opt.* **18**, 017001.
- [2] Gilev, K. V., *et al.*, 2016: Mature red blood cells: from optical model to inverse light-scattering problem. *Biomed. Opt. Express* **7**, 1305–1310.

Preferred mode of presentation: Oral/Poster