A GPU-Accelerated Open-Source Python Package for Calculating Powder Diffraction, Small-Angle-, and Total Scattering with the Debye Scattering Equation

Frederik L. Johansen[1,2]*¶, Andy S. Anker[1]*¶, Ulrik Friis-Jensen[1,2], Erik B. Dam[2], Kirsten M. Ø. Jensen[1]¶, Raghavendra Selvan[2,3]¶

* These authors contributed equally
¶ Correspondence to frjo@di.ku.dk (FLJ), andy@chem.ku.dk (ASA), kirsten@chem.ku.dk (KMØJ) or raghav@di.ku.dk (RS).

1: Department of Chemistry & Nano-Science Center, University of Copenhagen, Denmark
2: Department of Computer Science, University of Copenhagen, Denmark
3: Department of Neuroscience, University of Copenhagen, Denmark

Summary:
The Debye scattering equation, derived in 1915 by Peter Debye, is used to calculate scattering intensities from atomic structures considering the position of each atom in the structure[1,2]:

$$\text{Equation 1} \qquad I(Q) = \sum_{\nu=1}^{N} \sum_{\mu=1}^{N} b_\nu b_\mu \frac{\sin{(Qr_{\nu\mu})}}{Qr_{\nu\mu}}$$

In this equation Q is the momentum transfer of the scattered radiation, N is the number of atoms in the structure, and $r_{\nu\mu}$ is the distance between atoms ν and μ. For X-ray radiation, the atomic scattering factor, b, depends strongly on Q and is usually denoted as $f(Q)$, but for neutrons b is independent of Q and referred to as the scattering length. The Debye scattering equation can be used to compute the scattering pattern of any atomic structure and is commonly used to study both crystalline and non-crystalline materials with a range of scattering techniques like powder diffraction (PD), total scattering (TS) with pair distribution function (PDF) analysis and small-angle scattering (SAS)[1]. Although the Debye scattering equation is extremely versatile, the computation of the double sum, which scales $O(N^2)$, has limited the practical use of the equation.

With the advancement in computer hardware[3], analysis of larger structures is now feasible using the Debye scattering equation. Modern central processing units (CPUs), ranging from tenths to hundreds of cores offer an opportunity to parallelise computations, significantly enhancing compute efficiency. The same goes for graphics processing units (GPUs), which are designed with multiple cores acting as individual accelerated processing units that can work on different tasks simultaneously. In contrast, CPUs usually have fewer cores optimised for more general-purpose computing. This means that a GPU can execute multiple simple instructions in parallel, while a CPU might handle fewer parallel tasks[4]. Therefore, GPUs are better suited for calculations such as the Debye scattering equation, where many computations can be performed simultaneously. Taking advantage of such GPU acceleration could yields computational speeds that surpass those of even the most advanced multi-core CPUs; by orders of magnitude. We introduce a GPU-accelerated open-source Python package, named *DebyeCalculator*, for rapid calculation of the Debye scattering equation from chemical structures represented as xyz-files or CIF-files. The xyz-format is commonly used in materials chemistry for the description of discrete particles and simply consists of a list of atomic identities and their respective Cartesian coordinates (x, y and z). *DebyeCalculator* can take a crystallographic information file (CIF) and a user-defined spherical radius as input to generate an xyz-file from which a scattering pattern is calculated. We further calculate the PDF as described by Egami & Billinge[5]. We show that our software can simulate the PD, TS, SAS and PDF data orders of magnitudes faster than DiffPy-CMI[6]. *DebyeCalculator* is an open-source project that is readily available through GitHub: https://github.com/FrederikLizakJohansen/DebyeCalculator and PyPi (https://pypi.org/project/DebyeCalculator/).

The core functionality of *DebyeCalculator*, represented in the following high-level outline, starts with an initialisation function that sets user-defined parameters or set them to default. They include scattering parameters (such as Q-range, Q-step, PDF r-range, r-step, atomic vibrations, radiation type, and instrumental parameters) and hardware parameters.

During this initialisation phase, the calculation of the atomic form factors (for X-ray) or scattering lengths (for neutrons) is prepared based on the radiation type. Once initialised, *DebyeCalculator* can compute various quantities: the scattering intensity $I(Q)$ through the Debye scattering equation, the Total Scattering Structure Function $S(Q)$, the Reduced Total Scattering Function $F(Q)$, and the Reduced Atomic Pair Distribution Function $G(r)$. In this section, we specifically outline the $G(r)$ calculation using X-ray scattering. This is because the process for calculating $G(r)$ inherently involves the calculations for $I(Q)$, $S(Q)$, and $F(Q)$. When calling the *gr* function, *DebyeCalculator* loads the structure and computes the atomic form factors[7]. Following this, it calculates the scattering intensity $I(Q)$ using the Debye scattering equation and subsequently determines the structure factor $S(Q)$ and $F(Q)$. Necessary modifications, such as dampening and Lorch modifications, are applied before computing the $G(r)$. *DebyeCalculator* outputs the calculated functions to the CPU by defaults to allow for immediate analysis of the results, but user have the flexibility to retain the output on the GPU. It is worth noting that the majority of the compute time is spent on the double sum calculation in the Debye scattering equation. This is where GPU acceleration can enhance performance compared to single core CPUs. For all atom pairs, intermediate products of distances, form factors, and momentum transfers need to be calculated and stored temporarily. Calculating the intermediate products is computationally inexpensive but demands significant memory. This restricts the ability to apply the Debye scattering equation to structures with an increasing number of atoms. The batching schema in *DebyeCalculator* aims to mitigate these memory requirements by breaking down the calculations into smaller chunks that fit into the available GPU memory, thus enabling the calculation of scattering intensities for structures with a large number of atoms. The trade-off is a slight increase in computation time. Users with more substantial GPU memory can accommodate large structures while maintaining high computation speeds.

CLASS *DebyeCalculator*:

  FUNCTION Initialise(parameters...):
    - Set class parameters based on given input or defaults
    - Setup scattering parameters (e.g., Q-values, r-values) and hardware parameters
    - Load atomic formfactor coefficients
    - Setup form factor calculation based on radiation type

  FUNCTION gr(structure_path, keep_on_device=False):
    - Load atomic structure from given structure_path
    - Calculate atomic form factors
    - Calculate scattering intensity $I(Q)$ (Debye scattering equation)
    - Compute structure factor $S(Q)$ based on $I(Q)$
    - Calculate $F(Q)$ based on Q-values and $S(Q)$
    - Apply modifications if necessary (like dampening and Lorch)
    - Calculate pair distribution function $G(r)$ based on $F(Q)$
    - Return $G(r)$ either on GPU or CPU

In order to benchmark our implementation, we compare simulated scattering patterns from *DebyeCalculator* against DiffPy-CMI[6], which is a widely recognised software for scattering pattern computations. Here, our implementation obtains the same scattering patterns as DiffPy-CMI (Supporting Information A), while being faster on CPU for structures up to ~20,000 atoms (Figure 1). Both calculations are run on a x86-64 CPU with 64GB of memory and a batch size of 10,000. Running the calculations on the GPU provides another notable boost in speed (Figure 1). This improvement primarily stems from the distribution of the double sum calculations across a more extensive set of cores than is feasible on the CPU. With smaller atomic structures, an overhead associated with initiating GPU calculations, results in the NVIDIA RTX A3000 Laptop GPU Laptop computations being slower than DiffPy-CMI and our CPU implementation. Once the atomic structure size exceeds ~14 Å in diameter (~300 atoms), we observe a ~5 times speed-up using a NVIDIA RTX A3000 Laptop GPU with 6GB of memory and a batch size of 10,000. The choice of GPU hardware has a substantial influence on this speed advantage. As demonstrated in Figure 1, using a NVIDIA Titan RTX GPU, which offers 24GB of memory, the speed benefits become even more evident. The NVIDIA Titan RTX GPU delivers a performance that is ~10 times faster, seemingly across all structure sizes, underlining the significant role of the hardware.

With the advancements of GPUs like NVIDIA's Grace Hopper Superchip[8], which boasts 576GB of fast-access of memory, there is potential for *DebyeCalculator* to achieve even greater speeds in the future.
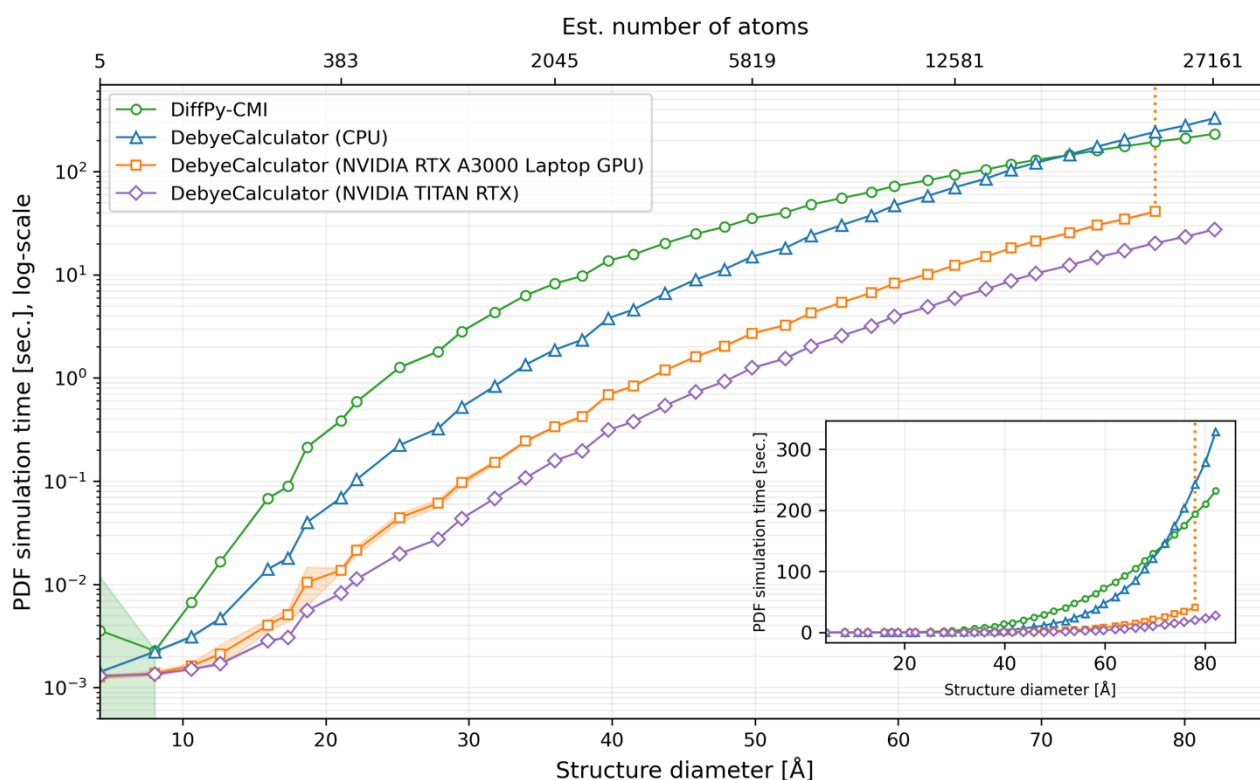
Figure 1: Computation-time comparison of the $G(r)$ calculation using our CPU- and GPU-implementation against DiffPy-CMI[6]. For the CPU-implementation, a batch size of 10,000 was chosen (x86-64 CPU with 6GB). Both the GPU implementations were run with a batch size of 10,000 (NVIDIA RTX A3000 Laptop GPU with 6GB of memory and NVIDIA Titan RTC GPU with 24GB of memory) The mean and standard deviation of the PDF simulation times are calculated from 10 runs. Note that, due to limited memory, the Laptop GPU was unable to handle structures larger than ~24,000 atoms.

Statement of need

Several software packages already exist for simulating the Debye scattering equation, including DiffPy-CMI,[6] debyer[9], Debussy[10,11], TOPAS[12], and BCL::SAXS[13]. Our software distinguishes itself in several ways. Firstly, it is freely available and open-source licensed under the Apache License 2.0. Moreover, it is conveniently implemented as a 'pip' install package and has been integrated with Google Colab[https://github.com/FrederikLizakJohansen/DebyeCalculatorGPU/blob/main/quickstart/QuickStart.ipynb], allowing users to rapidly calculate the Debye scattering equation without the need of local software installations. *DebyeCalculator* can be run through an interactive interface (see Figure 2), where users can calculate $I(Q)$, $S(Q)$, $F(Q)$ and $G(r)$ from structural models on both CPU and GPU.
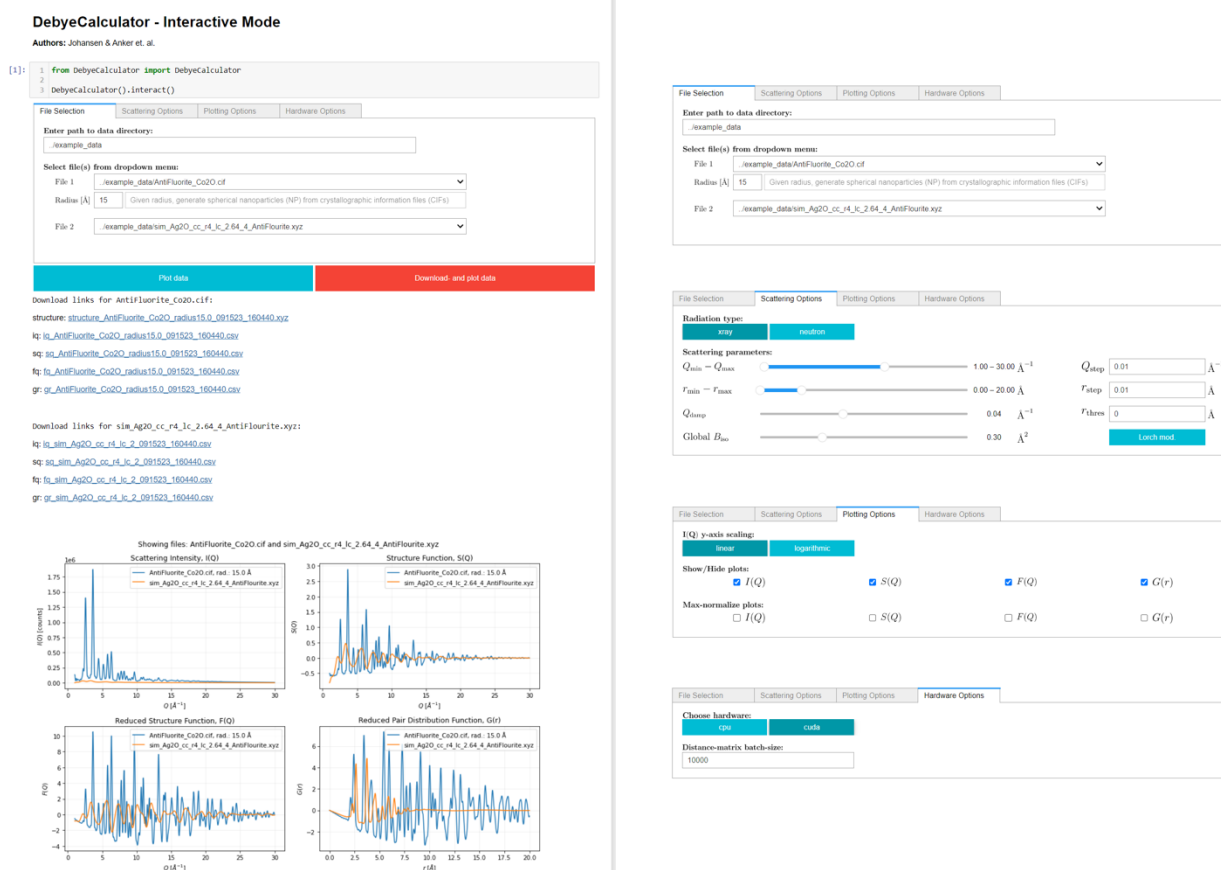
Figure 2: The interact mode of *DebyeCalculator* provides a one-click interface, where user can update parameters and visualise $I(Q)$, $S(Q)$, $F(Q)$ and $G(r)$. Additionally, the $I(Q)$, $S(Q)$, $F(Q)$ and $G(r)$, and xyz-file can be downloaded, including metadata.

Acknowledgements

References

1   Scardi, P., Billinge, S. J. L., Neder, R. & Cervellino, A. Celebrating 100 years of the Debye scattering equation. *Acta Crystallogr. A* **72**, 589-590, doi:10.1107/S2053273316015680 (2016).

2   Debye, P. Zerstreuung von Röntgenstrahlen. *Annalen der Physik* **351**, 809-823, doi:10.1002/andp.19153510606 (1915).

3   Schaller, R. R. Moore's law: past, present and future. *IEEE spectrum* **34**, 52-59 (1997).

4   Garland, M. *et al.* Parallel computing experiences with CUDA. *IEEE micro* **28**, 13-27 (2008).

5   Egami, T. & Billinge, S. J. L. *Underneath the Bragg Peaks, Pergamon* (2012).

6   Juhás, P., Farrow, C. L., Yang, X., Knox, K. R. & Billinge, S. J. L. Complex modeling: a strategy and software program for combining multiple information sources to solve ill posed structure and nanostructure inverse problems. *Acta Crystallogr. A* **71**, 562-568, doi:10.1107/S2053273315014473 (2015).

7   Waasmaier, D. & Kirfel, A. New analytical scattering-factor functions for free atoms and ions. *Acta Crystallogr. A* **51**, 416-431, doi:10.1107/S0108767394013292 (1995).

8   NVIDIA. In NVIDIA. https://resources.nvidia.com/en-us-grace-cpu/grace-hopper-superchip. (2023).

9   Wojdyr. Debye's scattering equation and other analysis of atomistic models. In GitHub. https://github.com/wojdyr/debyer. (2023).

10   Cervellino, A., Giannini, C. & Guagliardi, A. DEBUSSY: a Debye user system for nanocrystalline materials. *J. Appl. Cryst.* **43**, 1543-1547, doi:10.1107/S0021889810041889 (2010).

11   Cervellino, A., Frison, R., Bertolotti, F. & Guagliardi, A. DEBUSSY 2.0: the new release of a Debye user system for nanocrystalline and/or disordered materials. *J. Appl. Cryst.* **48**, 2026-2032, doi:10.1107/S1600576715020488 (2015).
12   Coelho, A. A. TOPAS and TOPAS-Academic: an optimization program integrating computer algebra and crystallographic objects written in C++. *J. Appl. Cryst.* **51**, 210-218, doi:10.1107/s1600576718000183 (2018).
13   Putnam, D. K., Weiner, B. E., Woetzel, N., Lowe Jr., E. W. & Meiler, J. BCL::SAXS: GPU accelerated Debye method for computation of small angle X-ray scattering profiles. *Proteins: Structure, Function, and Bioinformatics* **83**, 1500-1512, doi:10.1002/prot.24838 (2015).
14   Ghedira, M., Vincent, H., Marezio, M. & Launay, J. C. Structural aspects of the metal-insulator transitions in V0.985Al0.015O2. *J. Solid State Chem.* **22**, 423-438, doi:10.1016/0022-4596(77)90020-2 (1977).

Supporting Information A:
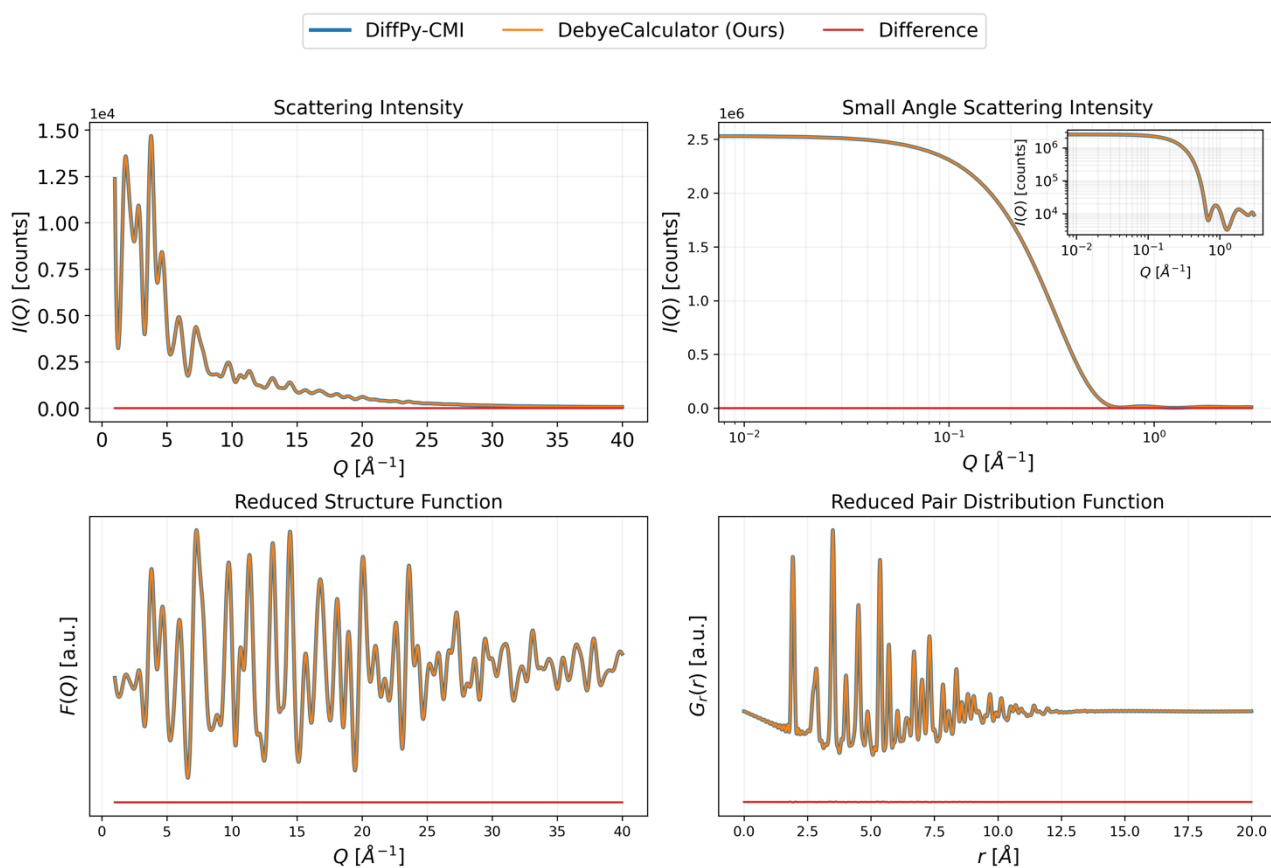
Figure S1: Comparison of the calculated $I(Q)$, SAXS, $F(Q)$ and $G(r)$ of *DebyeCalculator* and DiffPy-CMI[6] on a discrete, spherical cutout with 6 Å in radius from a $V_{0.985}Al_{0.015}O_2$ crystal.[14]