Data Competence for Photonic Nanotechnologies

Jörg Meyer[1], Nigar Asadova[2], Dominik Beutel[3], Uğur Çayoğlu[1], Carsten Rockstuhl[2,3],
Frank Tristram[4]

[1]Steinbuch Centre for Computing, Karlsruhe Institute of Technology;
[2]Institute of Nanotechnology, Karlsruhe Institute of Technology;
[3]Institute of Theoretical Solid State Physics, Karlsruhe Institute of Technology;
[4]3D Matter Made to Order, Karlsruhe Institute of Technology

In nano-optics, the optical properties of scattering objects can be expressed by so-called transition matrices (T-matrices). T-matrices are obtained by solving Maxwell's equations for given configurations under many different illumination conditions. The T-matrix can be extracted from the individual scattering response. Although this might be compute-intensive, T-matrices, as the outcome of these computations, are often not shared nor reused. This is a waste of resources and also does not permit addressing novel scientific questions. In this contribution, we present the strategies of the *data competence for photonic nanotechnologies* (DAPHONA) project to tackle this issue by suggesting a standardized file format for T-matrices and functionalities of a proper web service.

1 Introduction

In nano-optics we deal, among other things, with the optical properties of structures with a spatial extent comparable to or smaller than the optical wavelength. These scatterers have many applications, e.g., in imaging, sensors, or quantum technologies. All optical properties of these scatterers are captured by their transition matrix (T-matrix). This T-matrix describes how an illumination field is converted into a scattering field. T-matrices are the basis for describing complex nano-optical systems that are made out of many individual scatterers. These are systems that consist of coupled particles, a larger number of disordered particles, or structures made out of a periodic arrangement of identical objects. Properties of all these advanced photonic materials can be easily expressed once the T-matrix of the individual constituents is known. Currently, these T-matrices are repeatedly recalculated and not systematically reused. This wastes computational resources and also does not allow to address questions that could be answered on the basis of these data.

Publiziert in: Vincent Heuveline, Nina Bisheh und Philipp Kling (Hg.): E-Science-Tage 2023. Empower Your Research – Preserve Your Data. Heidelberg: heiBOOKS, 2023. DOI: `https://doi.org/10.11588/heibooks.1288.c18092` (CC BY-SA 4.0)

The presented project DAPHONA (Data competence for photonic nanotechnologies) aims to address this shortcoming. We will provide technologies to combine the geometric and material properties of an object and its optical properties, expressed by the T-matrix, into a unified data structure. This data will be used systematically to extract the T-matrix from the data set for a given object. Also, we want to identify objects that have predefined optical properties expressed by a T-matrix. Along these approaches, we will answer many novel questions that can be addressed by the data-driven approach formulated here.

In DAPHONA, we combine subject specific expertise with data expertise. We present our approach to the data model at an early stage of the project to allow coordination with the community and experts from research data management. Our approach is open, based on FAIR principles (Wilkinson et al. 2016), and will provide sustainable benefits to the entire community.

2 Transition Matrix Method

At the center of all considerations are the properties of the individual scattering objects. All, and really all, linear optical properties of such a scatterer can be represented in a highly aggregated form: the transition matrix (abbrev. T-matrix; Waterman 1965, 1971). The T-matrix generally describes how an illumination field with a fixed wavelength and expanded in a suitable basis system (represented with a vector, which contains the amplitudes of the basis functions) is converted into a scattered field, which is also expanded in an appropriate basis system (which, again results in a vector). The relation between illumination and scattered field is a vector-matrix multiplication, in which the illumination vector is merely multiplied by the T-matrix. Starting from the T-matrix as the central element, various ways have been developed in recent years to describe complex photonic systems. In particular, the optical properties of metasurfaces, metamaterials, or photonic crystals can be solved numerically or even analytically. The propagation properties of light in an infinite number of periodically arranged scatterers can be calculated on the basis of a band structure. The optical response of an array of thousands, sometimes even up to a million different objects can be calculated quickly if the T-matrix of each individual object is known.

3 Common File Format

We chose the Hierarchical Data Format (HDF5) as the storage format for T-matrices as it is commonly used by scientific communities. It is an open format, supported by open-source libraries in various programming languages on several platforms, and it is maintained by the HDF Group, a non-profit organization. The data model of HDF5 can represent heterogeneous data objects with various data types and in particular supports n-dimensional datasets. HDF5 is self-describing, i.e., allows to include metadata[1]. An HDF5 file describing a T-matrix needs to include the following minimal information:

---

[1] `https://www.hdfgroup.org`; *Last accessed on September 9th, 2023.*

- T-matrix: i.e., a complex-valued matrix
- Modes: l, m, and the electric or magnetic polarization (l and m are indexes specifying the modes described by vector spherical waves)
- Name, description, and keywords
- ID as a unique identifier
- Frequency of electromagnetic waves
- Name and version of the software that created the HDF5 file

HDF5 does not natively support complex numbers. The real and imaginary parts are stored in separate arrays. The modes need to be associated with the rows and columns of the T-matrix. A full specification of the file format needs to include the applied definitions of the normalization of the modes allowing to define the used vector spherical waves unambiguously. Frequencies may be given as frequency, angular frequency, vacuum wavelength, vacuum wavenumber, or angular vacuum wavenumber that are related to each other. While this set of information is the minimum required more information should be specified in addition:

- Material: description of the isotropic, anisotropic, or bi-isotropic material
- Embedding material: allows to specify the embedding medium in case it differs from the vacuum
- Computation
- Geometry
- Mesh

An isotropic material can be specified by giving the relative permittivity and relative permeability. The metadata on the computation contains information on the used software, version, and parameters used to calculate the T-matrix. The description of the geometry of a scatterer is very important. In Table 1 example parameters are given that describe the geometry of simple 3D shapes. More complicated geometries and material distributions can be described by a mesh. The mesh itself should be specified in an existing standard file format like STL (StereoLithography file) (Roscoe 1988) or gsmh (file format of the tool Gsmh: a three-dimensional finite element mesh generator with built-in pre- and post-processing facilities; Geuzaine and Remacle 2009).

4 Repository for T-Matrix Data

While the definition and establishment of a common file format for T-matrices is an important step for the reusability of T-matrix calculations, in addition, a web platform is needed with the functionality of a repository and data-specific search functionality. Repositories allow providing FAIR data. This includes publishing and sharing of data. New T-matrices may be uploaded and described by metadata. If needed, a persistent

Table 1: Example parameters to describe the geometry of simple three-dimensional shapes and their defining parameters.

Shape	Parameters	Comments
Sphere	radius	
Ellipsoid	radiusx, radiusy, radiusz	
Spheroid	radiusxy, radiusz	
Cylinder	radius, height	Rotational symmetry around z-axis
Cone	radius, height	Rotational symmetry around z-axis
Torus	major_radius, minor_radius	Rotational symmetry around z-axis
Cube	length	
Rectangular cuboid	lengthx, lengthy, lengthz	

identifier like a DOI could be assigned to a dataset. However, for scientists, it would be very useful to be able to search for domain-specific metadata. For example, a scientist might want to query all T-matrices for a given material and geometry. The search engine should not only return exact matches but should allow for fuzzy searches, i.e., also list results for similar query parameters. One objective of the DAPHONA project is to propose a suitable metric defining the similarity of T-matrices for this purpose.

5 Inverse Problem

For a given configuration (materials and geometry) the Maxwell's equations can be solved, and the T-matrix be derived. Inferring the materials and geometry for a given T-matrix is called the inverse problem. The solution of the inverse problem does not need to be unique and in general, is not simple to achieve. One approach to tackle this challenge is to apply machine learning. While supervised machine learning approaches proved to be very successful in many domains, they require large labeled training datasets. A benefit of the standardization and sharing efforts of the DAPHONA project is to compile such a training dataset for future research.

6 Conclusions

Nano optics has many practical applications and aims at the development of new materials with specific (optical) properties. In the project DAPHONA researchers with a background in nano physics and data science collaborate in order to develope a common file format to store and share T-matrices that contain all optical properties of new materials and structures. The calculation of these T-matrices is very computationally intensive. We presented ideas to define a common data format based on HDF5 allowing to share and reuse T-matrices to avoid the recalculation of the same or similar T-matrices. Besides a common file format, a repository with T-matrix-specific search functionalities is required

for the community. Once established a repository with T-matrices can be the basis for further research questions like trying to solve the inverse problem with supervised machine learning.
We present the ideas of DAPHONA at an early project phase in order to be able to coordinate efforts and integrate early feedback from the nano-optics and e-science communities.

Acknowledgements

We acknowledge support by the Federal Ministry of Education and Research (BMBF) within the project DAPHONA (16DKWN039).

References

Geuzaine, Christophe, and Jean-François Remacle. 2009. "Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities". *International Journal for Numerical Methods in Engineering* 79 (11): 1309–1331. DOI: `https://doi.org/10.1002/nme.2579`.

Roscoe, Larry. 1988. "Stereolithography interface specification". *America-3D Systems Inc* 27 (10).

Waterman, Peter C. 1965. "Matrix formulation of electromagnetic scattering". *Proceedings of the IEEE* 53 (8): 805–812. DOI: `https://doi.org/10.1109/proc.1965.4058`.

———. 1971. "Symmetry, Unitarity, and Geometry in Electromagnetic Scattering". *Physical Review D* 3 (4): 825–839. DOI: `https://doi.org/10.1103/PhysRevD.3.825`.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersbergand, Gabrielle Appleton, Myles Axtonand, Arie Baakand, Niklas Blombergand, et al. 2016. "The FAIR Guiding Principles for scientific data management and stewardship". *Scientific data* 3 (1): 1–9. DOI: `https://doi.org/10.1038/sdata.2016.18`.