eCSE Final Report for ARCHER2-eCSE02-12

Massively parallel and scalable electromagnetic solver for fast analysis of nonlinear optical processes in large clusters of nanoparticles

Ivan Sekulic and Nicolae C. Panoiu

University College London

Abstract

Second-harmonic generation is perhaps the most ubiquitous nonlinear optical process; it consists of the generation of waves with frequency twice as large as that of an excitation optical wave (also called linear or pump wave) that interacts with a nonlinear optical medium. As such, this nonlinear optical process is at the heart of many engineering applications, such as nonlinear plasmonic nanoantennas, nonlinear microscopy and imaging, photonic crystals and cavities. Therefore, accurate and fast characterization of second-harmonic (SH) process is vital for the design phase of an efficient development of many photonic devices. The main aim of this eCSE project was to facilitate numerical and computational investigations of nonlinear SH processes in clusters of nanoparticles with an emphasize on parallel and scalable implementations geared towards execution on distributed memory supercomputers.

1. Project description

Before the start of this eCSE project, OPTIMET was a parallel and scalable C++ code based on the multiple scattering matrix (MSM) method able to numerically analyse electromagnetic wave scattering from large clusters of spherical nanoparticles made of arbitrary optical materials. OPTIMET could be used in a multitude of scenarios of interest in optical technology such as, for example, analysis of metasurfaces used in the design of flat lenses, nonlinear plasmonic nanoantennas, nonlinear microscopy and imaging, and photonic crystals and cavities [1, 2]. In its previous version, OPTIMET was able to only handle linear optical problems, in which the incident field impinging on the cluster of nanospheres, as well as the scattered field resulting from the interaction of the incident field with the cluster, oscillate with the same frequency.

In this eCSE project, we significantly increase OPTIMET's range of functionalities and application domains by extending it along three main directions:

- Incorporated a nonlinear module able to analyse second-harmonic (SH) optical processes in nanoparticle systems. The nonlinear module is seamlessly integrated with the linear code by exploiting the pre-existing software architecture, when possible, as well as already available direct and iterative linear system solvers. It was parallelized using MPI standard in a similar fashion to the linear module, so that the numerical analysis of both optical phenomena, linear and SH, can be performed together on many-node computing architectures.
- In the previous version, OPTIMET was able to handle only spherical particles, which represented a serious limitation because in the real world nanoparticles often have

non-trivial shape. We have introduced the possibility for the analysis of particles of virtually arbitrary shape, both in the linear and nonlinear cases.

- Implemented an adaptive cross approximation (ACA) procedure on top of the linear and nonlinear computational modules. Using this compression strategy, low rank representations of submatrices describing nanoparticle optical coupling were found, both at the fundamental frequency (FF) and SH. In this way, we drastically reduced the memory consumption and sped up the matrix-vector multiplication inherent to iterative solvers.

2. Features developed during the eCSE project

In this section we describe in more detail the work developed during this eCSE project. The summary is given in the bullet points of the previous section.

a) Analysis of second-harmonic optical process in nanoparticle clusters (spherical case) and the adaptive cross approximation acceleration technique

We are interested in the analysis of SH scattering from spherical particles made of centrosymmetric, nonmagnetic and isotropic optical material, with electrical properties described with dispersive complex permittivity function, $\epsilon(\omega)$. The system is excited by a time-harmonic electromagnetic plane wave oscillating at FF, ω. The nonlinear SH optical waves oscillating at angular frequency $\Omega = 2\omega$ originates from two types of nonlinear polarization sources associated to each scatterer. The centrosymmetric crystal lattice of the material forbids the formation of local dipole polarization source inside the nanoparticles; however, due to the symmetry breaking at the vicinity of the surface of the particle, dipole surface nonlinear polarization density is induced. Although this interfacial origin of SH is usually enough to accurately describe second-harmonic generation (SHG) from metallic particles, when a more accurate analysis of purely dielectric targets is required, one must add a nonlocal nonlinear polarization source induced inside the particle, whose origin are the electric quadrupoles and magnetic dipoles induced in the nonlinear medium.

In the context of the MSM method applied to the nonlinear SH case, the electromagnetic fields in the volume external (V_e) (usually vacuum) and internal (V_i) to the n^{th} particle in the cluster of N particles in total, are expanded in series of vector spherical harmonics (VSHs) $\left(\boldsymbol{M}_v^{(1)}, \boldsymbol{N}_v^{(1)}\right)$ and $\left(\boldsymbol{M}_v^{(3)}, \boldsymbol{N}_v^{(3)}\right)$ as [3]:

$$\boldsymbol{E}^e(\boldsymbol{r}) = \sum_{n=1}^{N} \sum_{v\geq 1} \left[b_v^n \boldsymbol{M}_v^{(3)}(k_e \boldsymbol{r}_n) + a_v^n \boldsymbol{N}_v^{(3)}(k_e \boldsymbol{r}_n) \right], \ \boldsymbol{r} \in V_e, \tag{1}$$

$$\boldsymbol{E}^i(\boldsymbol{r}) = \sum_{v\geq 1} \left[c_v^n \boldsymbol{M}_v^{(1)}(k_n \boldsymbol{r}_n) + d_v^n \boldsymbol{N}_v^{(1)}(k_n \boldsymbol{r}_n) \right], \ \boldsymbol{r} \in V_i, \tag{2}$$

with k_e and k_n representing wave numbers associated to the exterior and the interior regions of the n^{th} sphere calculated at the SH frequency, respectively. The corresponding spherical wave expansions of the SH magnetic fields are derived directly from Maxwell-Faraday law using the curl relation. A key step in the calculation of the unknown expansion coefficients (b_v^n, a_v^n) and (c_v^n, d_v^n) associated to the n^{th} particle is the expansion of the nonlinear SH boundary conditions [3, 4], related to the n^{th} particle, in a series of VSHs. In this way, we can derive a T-matrix [5] relating the external field expansion coefficients to the coefficients used in the expansion of the boundary conditions at the SH. For a single particle case, this can be formally expressed as:

$$[\boldsymbol{f}_n] = \boldsymbol{T}_n[\boldsymbol{g}_n], \tag{3}$$

where the vectors $\boldsymbol{f}_n$ and $\boldsymbol{g}_n$ contain the VSH expansion coefficients of the SH external field and the nonlinear boundary conditions, respectively. For the case of a spherical particle, $\boldsymbol{T}_n$ is a diagonal matrix with entries that can be computed analytically. The extension of Eq. (3) to multiparticle systems is achieved by employing translation-addition matrices [6], which are used for the characterization of multiple scattering effects among the particles. The total SH electric field exciting the n^{th} sphere is equal to the sum of the fields radiated by all the other spheres, and is given by:

$$\boldsymbol{E}^{n,ex}(\boldsymbol{r}) = \sum_{\substack{k=1 \\ k\neq n}}^{N} \left[\boldsymbol{M}_v^{(3)}(k_e\boldsymbol{r}_k), \boldsymbol{N}_v^{(3)}(k_e\boldsymbol{r}_k)\right]\boldsymbol{f}_k\,,\ n = 1 \dots N \tag{4}$$

This field can be expanded in the local coordinates associated to the n^{th} sphere employing the translation-addition matrices $\boldsymbol{\alpha}_{n,k}$, evaluated at the SH frequency as:

$$\boldsymbol{E}^{n,ex}(\boldsymbol{r}_n) = \sum_{\substack{k=1 \\ k\neq n}}^{N} \left[\boldsymbol{M}_v^{(1)}(k_e\boldsymbol{r}_n), \boldsymbol{N}_v^{(1)}(k_e\boldsymbol{r}_n)\right]\boldsymbol{\alpha}_{n,k}\boldsymbol{f}_k\,,\ n = 1 \dots N \tag{5}$$

The total SH field incident onto the n^{th} sphere, $\boldsymbol{E}^{n,ex}$, is expanded into a series of VSHs, with the corresponding expansion coefficients forming the vector $\boldsymbol{e}_n$. Then Eq. (5) can be written solely in terms of SH field expansion coefficients as follows:

$$\boldsymbol{e}_n = \sum_{\substack{k=1 \\ k\neq n}}^{N} \boldsymbol{\alpha}_{n,k}\boldsymbol{f}_k\,,\ n = 1 \dots N \tag{6}$$

Multiplying Eq. (6) with the *T*-matrices of scatterers ($\boldsymbol{T}_1, \dots, \boldsymbol{T}_N$), we are left with the system comprising the external field expansion coefficients only, expressed in matrix form as:

$$\begin{bmatrix} \mathbf{I} & -\boldsymbol{T}_1\boldsymbol{\alpha}_{(1,2)} & -\boldsymbol{T}_1\boldsymbol{\alpha}_{(1,3)} & \dots & -\boldsymbol{T}_1\boldsymbol{\alpha}_{(1,N)} \\ -\boldsymbol{T}_2\boldsymbol{\alpha}_{(2,1)} & \mathbf{I} & -\boldsymbol{T}_2\boldsymbol{\alpha}_{(2,3)} & \dots & -\boldsymbol{T}_2\boldsymbol{\alpha}_{(2,N)} \\ -\boldsymbol{T}_3\boldsymbol{\alpha}_{(3,1)} & -\boldsymbol{T}_3\boldsymbol{\alpha}_{(3,2)} & \mathbf{I} & \cdots & -\boldsymbol{T}_3\boldsymbol{\alpha}_{(3,N)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\boldsymbol{T}_N\boldsymbol{\alpha}_{(N,1)} & -\boldsymbol{T}_N\boldsymbol{\alpha}_{(N,2)} & -\boldsymbol{T}_N\boldsymbol{\alpha}_{(N,3)} & \cdots & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{f}_1 \\ \boldsymbol{f}_2 \\ \boldsymbol{f}_3 \\ \vdots \\ \boldsymbol{f}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{T}_1\boldsymbol{g}_1 \\ \boldsymbol{T}_2\boldsymbol{g}_2 \\ \boldsymbol{T}_3\boldsymbol{g}_3 \\ \vdots \\ \boldsymbol{T}_N\boldsymbol{g}_N \end{bmatrix}, \tag{7}$$

or in a more compact form:

$$[\boldsymbol{S}^{\Omega}][\boldsymbol{F}^{\Omega}] = [\boldsymbol{G}^{\boldsymbol{\Omega}}]. \tag{8}$$

The assembly of the multiparticle system scattering matrix $\boldsymbol{S}^{\Omega}$, together with the excitation vector $\boldsymbol{G}^{\boldsymbol{\Omega}}$, resulting from the discretization of the nonlinear boundary conditions, and the solution of the linear system given in Eq. (8) represent the main computational tasks performed by the nonlinear module developed in this project. These tasks are initially programmed as a serial C++ code, and afterwards parallelized using MPI for optimal execution on distributed memory HPC platforms. Physical quantities computed at the postprocessing stage, such as field profiles (*Fig. 1*) and scattering and absorption spectra (*Fig. 2*) are easily computed from the solution of linear system (8).

Because of the complexity of the MSM method and page limitations imposed by this ***Report***, we had no choice but to leave many details regarding the computational algorithm out. Those interested in the nonlinear counterpart of the MSM method can find out more details about it in Ref. [7], which we also attached to this ***Report*** as ***Annex I***.

To speed up the execution of the code, we have included the possibility for the user to compress the system scattering matrices $\boldsymbol{S}^{\omega}$ and $\boldsymbol{S}^{\Omega}$, arising from the MSM analysis at the FF and SH, respectively, using the ACA algorithm [8]. Some of the off-diagonal submatrices in $\boldsymbol{S}^{\omega}$ and $\boldsymbol{S}^{\Omega}$ describing the electromagnetic coupling between separated particles are rank

deficient and thus can be approximated as the product of two matrices with lower rank. This *low-rank approximation* results in the compression of the scattering matrices $\boldsymbol{S}^{\omega}$ and $\boldsymbol{S}^{\Omega}$ hence saving memory and accelerating the matrix-vector product inherent to GMRES iterative solvers. The adaptive cross approximation procedure is implemented in a modular manner on top of the linear as well as nonlinear OPTIMET code.

The ACA procedure becomes particularly useful when dealing with clusters of a large number of particle (>200 scatterers) and large number of VSHs per particle (>10). To corroborate this claim, we considered a cubic cluster of silicon nanospheres embedded in vacuum and located in first octant of the Cartesian coordinate system, aligned with the *x* -, *y* - and *z*-axes. The cubic array contains 7x7x7 = 343 spheres with same radii $r_i = 50\ \mathrm{nm}, \mathrm{i} = 1, \ldots, 343$ which are equally spaced with the centre-to-centre distance $d = 190\ \mathrm{nm}$. Each sphere is anlysed with 160 harmonics in the FF and SH regimes, which amounts to scattering matrices of dimension equal to $dim(\boldsymbol{S}^{\omega}) = dim(\boldsymbol{S}^{\Omega}) = 54{,}880$. This is a computational problem of moderate size with complex system matrices. In terms of computer memory, each matrix element occupies 16 bytes (real and imaginary parts of the type double), which brings us to the total memory allocation of 48.19 GB per scattering matrix. This is a substantial memory consumption which we aim to reduce by using the ACA algorithm.

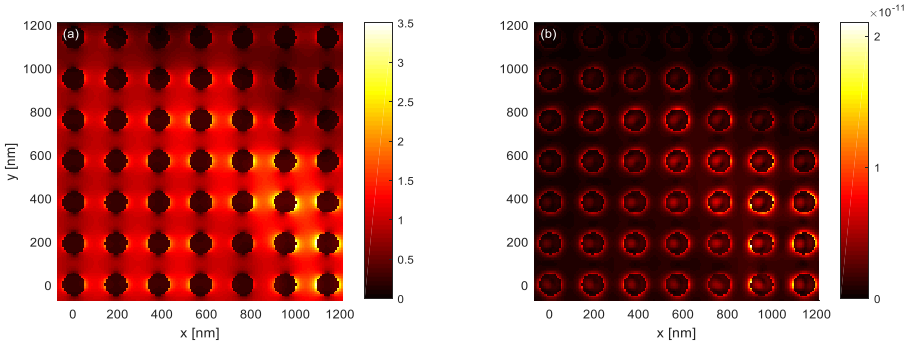

Fig. 1. *Electric field amplitude at the FF (left) and SH (right), calculated for a cubic cluster (343 silicon nanospheres) at the incident wavelength λ_{inc} = 800 nm. The fields are determined in the xy-plane at z=570 nm.*

In *Fig. 2* we plot the scattering and absorption cross sections at FF (top panels) and SH (bottom panels) computed with MSM and ACA compression (dashed lines with asterisk) and compare them to the results obtained without ACA compression (full lines). Very good agreement is observed both at the FF and SH. The scattering matrices compressed with ACA occupy only about 16% of the memory used by the uncompressed ones, thus making this problem amenable even for some high-end workstations.

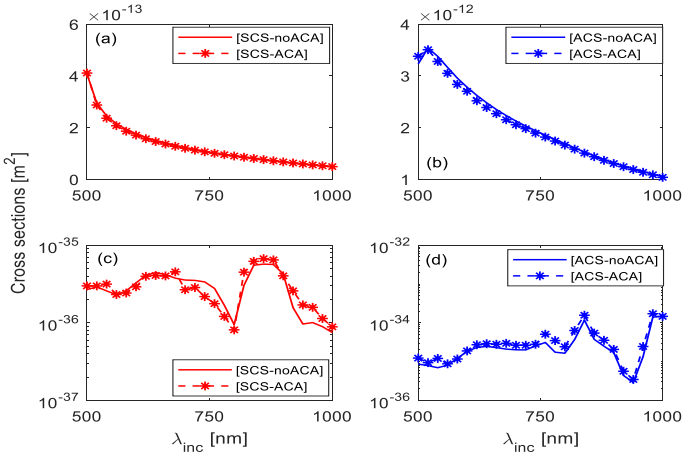

Fig. 2. *Scattering and absorption cross sections at FF (upper panels) and SH (bottom panels) for a cubic lattice of nanospheres*

The relative reduction in the memory requirements becomes more significant with the increase of the number of particles in the system under analysis as more matrix blocks, describing the coupling between distant particles, become rank deficient. This favourable property of ACA brought us to a level at which we can analyze clusters containing an order of magnitude more particles as compared to the situation before this eCSE project (now we can handle thousands of particles instead of hundreds).

b) The extension of the FF and SH analysis to nonspherical particles

The extension of the MSM numerical method to the case of nonspherical particles is by no means trivial. The main difference between the MSM-analysis of spherical and nonspherical (arbitrarily shaped) particles resides in the construction of the *T*-matrix characterising each scatterer. In the case of a sphere, this matrix is diagonal with the elements given by the well-known Mie coefficients. On the other hand, for the case of a particle with arbitrary shape, the *T*-matrix is generally full, with entries equal to integrals evaluated over the surface of the particle involving products of VSHs. The theory behind the MSM analysis of nonspherical particles at FF is well known [9], while the analogous theory valid at the SH is new, to the best of our knowledge, and has been developed as part of this eCSE project. This work is too elaborate to fit this section, so that the interested reader is referred to ***Annex I***, where the theory and computational results are described in detail. A detailed description of VSHs used in the linear and nonlinear MSM analysis can also be found in ***Annex I***.

Here we present just one example of how OPTIMET performs, namely we compare the linear and nonlinear scattering cross-section spectra obtained with MSM method, described in ***Annex I,*** with the cross-section spectra calculated by a commercial solver based on FEM. We analyze two identical gold prolate spheroids with equatorial and polar radii equal to $a = 50\,\mathrm{nm}$ and $c = 70\,\mathrm{nm}$, centred at $(0,\ 0,\ 0)$ and $(0,\ 0,\ 200)\,\mathrm{nm}$. This configuration is interesting as it may find its application to the design of plasmonic nanoantennas or sensing devices since strong near-fields can be induced in the gap between the spheroids, allowing the formation of so-called *hot spots*. The scattering cross-sections are presented in *Fig 3*.

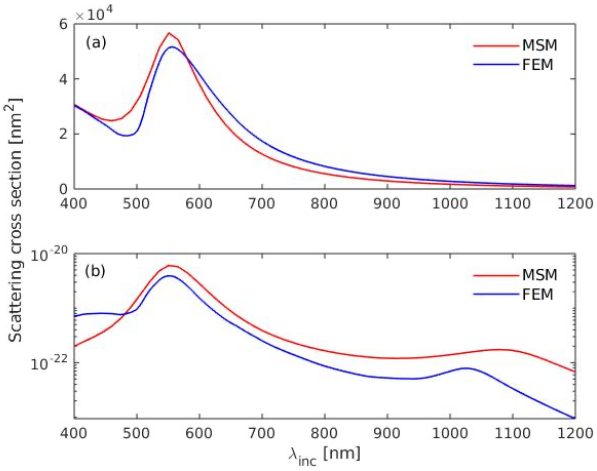

Fig. 3. *Scattering cross sections at FF (upper panel) and SH (bottom panel) for two prolate spheroids made of gold.*

3. Scalings and performance of the code on ARCHER2

In this section, we report on the performance of our parallel implementations on ARCHER2 computing cluster of the OPTIMET code.

The assembly of the scattering matrix $\boldsymbol{S}^{\Omega}$ at SH is parallelized in the same fashion as the assembly of $\boldsymbol{S}^{\omega}$ corresponding to the FF case, so that the interface to the parallel direct and

iterative solvers, already efficiently implemented and tested for the FF case, can be reused. Namely, each MPI process holds all the information about the geometry and materials parameters of the particles in the cluster, and computes one or several columns of the scattering matrix, representing interactions of each particle with all the others. The number of the scattering matrix columns each process is allocated depends on the number of particles in the cluster and number of allocated MPI processes. Load-balanced distribution is ensured across all of the processes.

Different from the linear module, whereby the computation of the excitation vector is very cheap, the evaluation of the nonlinear source vector $\boldsymbol{G}^{\Omega}$ is computationally intensive as it requires the calculation of special Clebsch-Gordan coefficients over multiply-nested sums. These calculations are performed in parallel as well in a load-balanced manner over vector spherical harmonics.

For our performance benchmark we chose an assembly of nanospheres arranged in a cubic lattice. We assumed that the spheres are made of silicon and are located in the first octant of the Cartesian coordinate system, aligned with the *x*-, *y*- and *z*-axes. The lattice contained 8×8×8 = 512 identical nanospheres with radii $r_n = 300\ \mathrm{nm}, n = 1, \ldots, 512$, which were equally spaced with center-to-center distance, $d = 700\ \mathrm{nm}$, thus forming a cubic array with the side length equal to $4.9\ \mu\mathrm{m}$. The cluster is illuminated with a time-harmonic plane wave characterized by wavelength, $\lambda_{\mathrm{inc}} = 500\ \mathrm{nm}$, propagating along a direction defined by the angles $\theta_{inc} = \pi/4$ and $\varphi_{inc} = \pi/2$ and polarized along the θ direction. Each sphere is discretized with 240 spherical harmonics resulting in a total of 122,880 basis functions per FF and SH computations. This discretization level translates to full-system scattering matrices $\boldsymbol{S}^{\omega}$ and $\boldsymbol{S}^{\Omega}$ occupying about 241.6 GB of memory each.

In the following figures we plot the performance of the system scattering matrix assembly and direct as well as iterative GMRES solutions with and without ACA compression, for both the FF and SH cases. We calculated the speedup as compared to the computation time performed on 8 MPI processes. This number of MPI processes was the minimum number we could use on our benchmark calculations before memory issues or computation time restrictions made the calculations impractical. Therefore, the speedup factor η, introduced in the *Success Metrics* of *Objective 3*, must be rescaled considering this issue i.e., considering the calculations on 8 MPI processes as "serial".

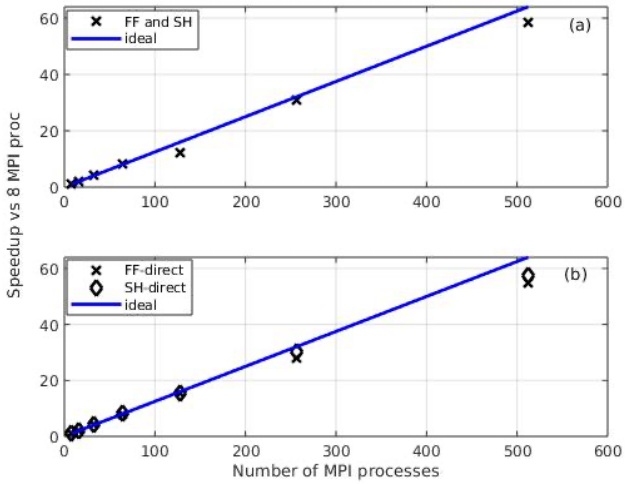

Fig. 4. *Speedup of the system scattering matrix assembly at FF and SH (top panel) as the number of processes is increased from 8 to 512, with respect to the computations performed on 8 MPI processes. The same speedup computations for the direct ScaLAPACK linear system solution (bottom panel).*

In *Fig. 4(a)* we show the speedup of the computation regarding the system matrix assembly (same for the FF and SH) as we increase the number of MPI processes from 8 to 512 and with respect to the computations performed on 8 processes. It can be seen that the speedup is very close to the ideal case, especially for a smaller number of MPI processes. Similarly, in *Fig. 4(b)* we show the speedup of the direct ScaLAPACK solution of the linear systems at the FF and SH when the number of processes is increased from 8 to 512 with respect to the computations obtained on 8 processes. The linear system solution speedups for FF and SH cases are very similar, since the direct solver is immune to the matrix condition numbers. This plot shows that a very close to ideal speedup factor was achieved in this case, too. Accordingly, the speedup factor η takes the value between 1 and 0.9 as the number of MPI processes is increased, both in the scattering matrix assembly phase and when the solution of the linear systems is computed. The decrease in this factor is explained by the increase of the communication time between the processes as their number increases.

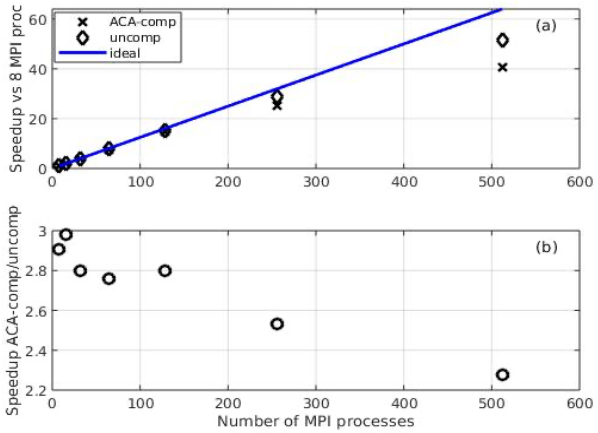

Fig. 5. *Speedup of the matrix-vector product involving the compressed ACA system for the FF or SH frequencies, and the uncompressed system, as the number of processes is increased from 8 to 512, with respect to the computations performed on 8 MPI processes (top panel). Speedup of the matrix-vector operations in GMRES of compressed vs uncompressed systems when the number of processes increases from 8 to 512 (bottom panel).*

In *Fig. 5(a)* we present the speedup of the matrix-vector operation inherent to the GMRES iterative solver when the scattering matrix of the system is uncompressed and compressed with the ACA procedure. For this purpose, we first had to update the GMRES solver to be able to operate on such compressed systems (we call GMRES-ACA this version of the GMRES). By a simple inspection of the plots in *Fig. 5(a)* we see that our parallel compressed matrix-vector product scales similarly as the uncompressed one, thus validating our parallel compressed iterative scheme. Moreover, in *Fig. 5(b)* we test the speed of the ACA-compressed system matrix-vector operation. Due to the low rank approximation of certain submatrices, we gained on the computational performance as compared to that in the case of the uncompressed system matrix-vector product for all number of MPI processes considered. This favourable performance of the ACA procedure is particularly significant when the condition number of the system matrix is not good, and there are no suitable preconditioners, which is often the case in the SH scattering analysis of clusters containing a large number of particles. In particular, the data in *Fig. 5(b)* shows that a speedup factor ranging from 2.2 to 3 can be achieved when the ACA compression technique has been employed.

This work was funded under the embedded CSE programme of the ARCHER2 UK National Supercomputing Service (http://www.archer2.ac.uk).

References

[1] R. W. Boyd, *Nonlinear Optics* (Academic Press, 1992).
[2] N. C. Panoiu, W. E. I. Sha, D. Y. Lei, and G. C. Li, "Nonlinear optics in plasmonic nanostructures", *J. Opt*. **20**, 083001 (2018).
[3] A. Capretti, C. Forestiere, L. Dal Negro, and G. Miano, "Full-wave analytical solution of second-harmonic generation in metal nanospheres", *Plasmonics* **9**, 151 (2014).
[4] J. I. Dadap, J. Shan, K. B. Eisenthal, T. F. Heinz, "Second-harmonic Rayleigh scattering from a sphere of centrosymmetric material", *Phys Rev Lett* **83**, 4045 (1999).
[5] M. I. Mishchenko, G. Videen, V. A. Babenko, N. G. Khlebtsov, and T. Wriedt, "Comprehensive T-matrix reference database: A 2004-06 update", *J. Quant. Spectrosc. Radiat. Transfer* **106**, 304 (2007).
[6] O. R. Cruzan, "Translational addition theorems for spherical vector wave functions," *Quart. Appl. Math*. **20**, 33 (1962).
[7] I. Sekulic, J. W. You, and N. C. Panoiu, "*T-matrix method for computation of second-harmonic generation upon optical wave scattering from clusters of arbitrary particles*", https://arxiv.org/abs/2208.13977.
[8] M. Bebendorf, "Approximation of boundary element matrices", *Numer. Math*. **86**, 565 (2000).
[9] M. I. Mishchenko, L. D. Travis, and D. W. Mackowski, "T-matrix computations of light scattering by nonspherical particles: a review", *J. Quant. Spectrosc. Radiat. Transfer* **55**, 535 (1996).