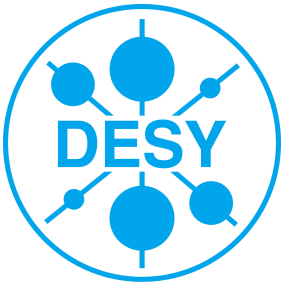

Data Evaluation Software for X-ray Standing Waves Investigation of Element-Specific Density Profiles in Biomembranes

Egor Marin*[1] and Dmitri Novikov†[1]

[1]Moscow Institute of Physics and Technology, Dolgoprudniy, Russia
[2]DESY, FS-PEX

September 7, 2016

Abstract

Various methods of atom position characterization in in bulk and surface regions of crystals, using x-ray standing wave (XSW), have emerged during development of x-ray experimental techniques in 20-th century. The XSW approach can be also applied to non-periodic objects, deposited on a crystal surface, with the standing wave generated inside a bulk used to probe the atomic arrangements above the surface. Alternatively, the standing waves can be produced by Bragg diffraction on long period artificial multilayer structures (ML). Various attempts were made to apply the ML-generated XSW for biomembranes characterization. Recently, in was demonstrated that the method can be directly applied to biomembranes in native surrounding. Here we present a Python software, developed for fitting of XSWF data from such objects. The new software was successfully tested on both model and experimental datasets and has proved to be fast and user-friendly. Noticeably, the program allows the user to introduce his own models, to put constraints in an existing model and simultaneously fit multiple data sets. This allows one to involve *a priori* information, which is often necessary in biological objects investigation.

*marin@phystech.edu
†dmitri.novikov@desy.de

Contents

1 Introduction **3**

2 Methods **3**
2.1 X-ray standing wave fluorescence 3
2.2 Light atoms distribution 5
2.3 Fitting procedure 6

3 Results and discussion **7**
3.1 Fitting workflow 7
3.2 Model datasets 7
3.2.1 Single-gaussian convergence 8
3.2.2 Two-gaussian distributions 8
3.3 Real data fitting 9

4 Conclusion and future plans **10**

1 Introduction

The discovery of X rays by Wilhelm Conrad Roentgen in 1895 was followed by extensive development of experimental techniques of matter characterization, based on high permeability and wavelengths range of x-rays. In 1960-s, concept of x-ray standing waves as a interference pattern between incident and Bragg reflected x-ray waves from a crystal (that was later expanded to several ordered structures) was proposed by P. W. Batterman [1]. Since then, several techniques, requiring high monochromaticity of x-rays (usually, from synchrotron sources) were developed and successfully applied to surface and interface characterization with high spatial resolution and chemical selectivity.

As an important biological objects, lipid layers and bilayers (usually referred to as biomembranes) are currently an object of interest of both fundamental science and industry. First, biomembranes separate inner part of the cell from outer space and serve as a substrate for membrane proteins. They are the targets for more then a half of modern drugs, and biomembranes properties play a key role in both *in vitro* and *in silica* fundamental studies of these proteins. Second, chemical properties of lipid molecules allow them to spontaneously form differentstructures such as bilayers, cubic phases, micelles, etc. in a solution, allowing one to use them for various research and drug-delivery purposes.

X-ray standing wave, with a period matching to the typical thickness of an membrane, can be generated within a Bragg peak from an artificial multilayer structure. The fluorescence yield from specific membrane atoms, which is measured as a function of angle close to Bragg peak, can be used to locate the position of atoms above the surface.

In 1990-s, first attempt to investigate single-layer Zn-labeled biomembranes using x-ray standing wave fluorescence (XSWF) was made [2]. However, plausible quantitative description of native light atoms (such as P, S) within membranes had become available only recently [3, 4]. Here we present a tool for evaluation of XSWF yield data from membrane objects. The software was tested on both model and experimental data. For tested objects, description employing simple one- or two-gaussian models of light atoms distribution was applied.

2 Methods

2.1 X-ray standing wave fluorescence

X-ray standing wave is a pattern of interference between incident and reflected x-ray waves from a multilayer structure. Since the high contrast of XSW requires the intensities of both waves to be comparable, they can be only observed within Bragg peaks, which appear only at very small angles (typically $\approx 1^\circ$) due to the high energy of x-rays.

Standing wave shows periodicity that equals to that in a multilayer structure, divided by the order of the Bragg peak (e.g. if we go from first to second Bragg reflection, one

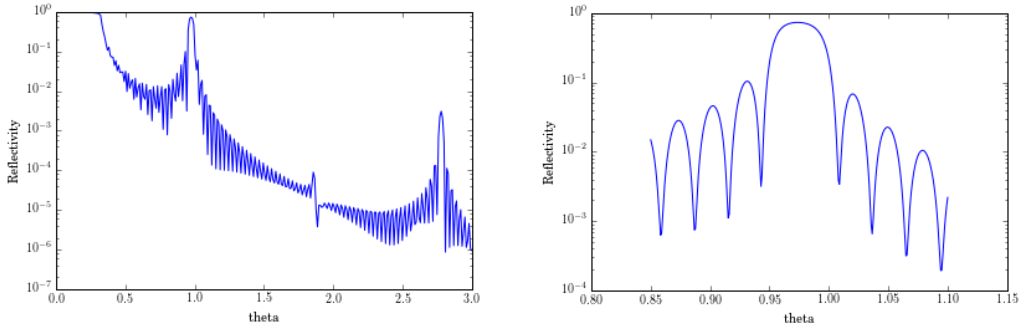

Figure 1: Typical reflectivity (logarithmic scale) from a 48 Åperiod multilayer, $\lambda = 1.5$ Å (left), and magnified first Bragg peak (right).

additional wave node will appear, doubling the number of nodes and reducing the period by half).

Phase of the standing wave can be precisely tuned by the incident angle variation in a small range within a Bragg peak (see Figure 2). For more detailed description of a standing wave, one can see e.g. ref. [5].

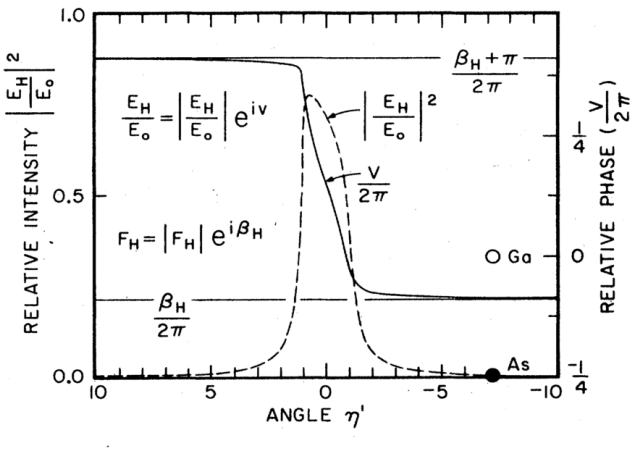

Figure 2: Angular variation of the reflectivity (solid) and reflected beam phase (dotted) for a multilayer structure. Courtesy from ref. [6]

For x-ray standing wave calculation we use the "X-ray server" [7], that delivers reflectivity and standing wave intensity above the surface of a ML structure with given parameters, such as chemical composition, thickness and various imperfection parameters.

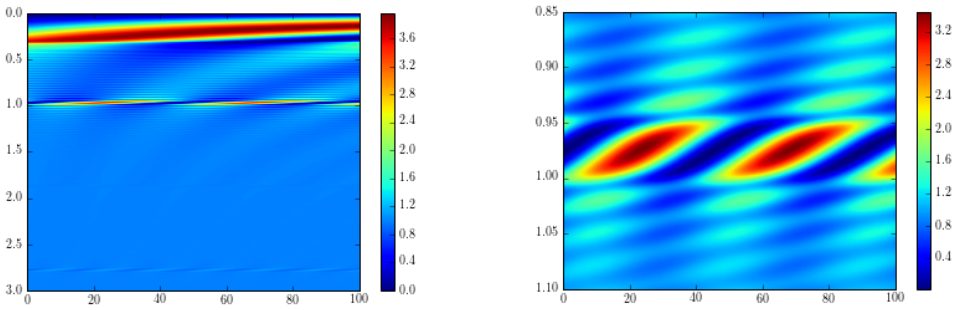

Figure 3: Standing wave profile above the surface (from 0 to 100 Å), calculated by "X-ray server". Vertical and horizontal axes are the gracing angle θ and distance above the surface z, respectively. The pattern on the left corresponds to the reflectivity curve as shown at the Figure 1 (left). Three standing waves appearing around $\theta \approx 0.1°$, $\theta \approx 2°$ $\theta \approx 3°$ correspond to total reflection and first and third Bragg peaks, respectively. Magnified wave at a first Bragg peak (shown on the right). One might clearly see that the period of the wave is ≈ 48 Å, same as the one of ML structure

2.2 Light atoms distribution

Fluorescence from light atoms, present in a molecule, can be in principle detected from almost any chemical component (such as O, N, C, P). In reality, long wavelength components from atoms as N, C and O get strongly absorbed by the water layer and atmospheric gases on the way to fluorescence detector.

However, P atoms look especially promising for detection, since they are natural labels of the hydrophilic head in phospholipids (1 atom per molecule) allowing one to precisely locate it via XSWF.

In the software we implemented two possible fitting regimes, assuming that lipids on the surface have native configuration (such as mono- or bilayer). For monolayer, distribution of P atoms can be described as a gaussian distribution with a mean value around a natural lipid length and standard deviation σ describes the level of disorder within the layer.

We also use an additional linear parameter k, taking into account the geometrical conditions of experimental setup, as described [8].

$$\rho\left(z,\theta\right) = c \cdot \exp\left(-\frac{\left(z-z_0\right)^2}{2\sigma^2}\right) + k\theta \tag{1}$$

where k, c, z_0 and σ are the fitting parameters.

The two-gaussian model allows to control the ratio between total amount of fluorescing atoms in two peaks, thus, enabling the processing of non-symmetrical distributions. Taking into account that the area under a gaussian is proportional to σc, we use the following model:

$$\rho(z,\theta) = c_1 \cdot \exp\left(-\frac{(z-z_{01})^2}{2\sigma_1^2}\right) + c_2 \cdot \exp\left(-\frac{(z-z_{02})^2}{2\sigma_2^2}\right) + k\theta \tag{2}$$

where

$$c_2 = \frac{c_1\sigma_1}{\sigma_2 N} \tag{3}$$

with fitting parameters $z_{01,2}$, $\sigma_{1,2}$, k and N (the ratio).

Implemented restraints allow one to set limits for each parameter (minimum and maximum, also including infinite borders), or to fix any set of parameters to allow more precise refinement. This is also helpful during the search of initial parameters.

2.3 Fitting procedure

From the fitting of reflectivity measurements (such as Figure 1), one can extract the parameters of the ML structure and calculate SW from this structure, using the "X-server", as a tabulated function of $A(z,\theta)$ – intensity of the wave at each point and each angle.

Experimental data is usually available as dependence $I(\theta)$, with the fluorescence yield at certain wavelengths (corresponding to the certain chemical components of the sample) at a mesh around chosen Bragg peak. Since we have $\rho(z,\theta)$ from Equation 1 or Equation 2 and the intensity $A(z,\theta)$ of the SW, one can obtain the model curve as:

$$I_{model}(\theta) = \sum_{z_{min}}^{z_{max}} \rho(z,\theta) \cdot A(z,\theta) \tag{4}$$

As a target function for minimization, we employ widely used among physicists chi-square parameter, that is, in our case:

$$\chi^2 = \frac{1}{L-p} \sum_{\theta_{min}}^{\theta_{max}} \left(\frac{I_{model} - I_{obs}}{\Delta_{obs}}\right) \tag{5}$$

where $\Delta_{obs} = \Delta_{obs}(\theta)$ is the experimental error at each angle. L and p are the number of experimental points and number of parameters, respectively.

As an independent goodness-of-fit parameters, we also use crystallographic r-factor-like value as a measure of consistency between data and model:

$$R = \sum_{\theta_{min}}^{\theta_{max}} \frac{|I_{obs} - I_{model}|}{|I_{obs}|} \tag{6}$$

Fitting routine uses LMFIT package [9], available for both PYTHON 2 and PYTHON 3 programming languages. Notably, out implementation allows one to run the script with both versions of PYTHON.

3 Results and discussion

3.1 Fitting workflow

On a reasonably big lattice ($81 \times 200 = 16200$ points) fitting of single gaussian model is not time consuming (from almost any reasonable parameters set it converges to the minimum, if there exists one, in less then 20 seconds on 1 core of IntelCore i5-4200M CPU with 2.50GHz). However, fitting of two-gaussian model is much more sophisticated and requires reasonable initial conditions and restraints for most of parameters. We usually run single-gaussian mode first, to estimate the initial amplitude, and then try to locate optimal values for gaussian centers and half-widths. It was noticed that use of reasonable restraints for both σ and z_0 parameters could be very helpful.

Specifically, one could set the "window" for z_0 parameter with $\approx \Lambda$ (ML period) width, e.g. MIN=30, MAX=80, and control the distance between to the surface by putting constraints on σ_1 (e.g. $\sigma_1 = 10$, MIN=0, MAX=10, to keep 3σ-width of the distribution above the surface). Control of the distance between two different distributions is also possible, but one should always remember that the distance larger then the period of a SW is useless due to summation along z-coordinate and Λ-periodicity of the SW.

It should be also noted that the software allows one to observe the fitting parameters in real time, significantly reducing the amount of wasted time in case of fitting is stuck in a local minimum.

Since the software allows only command-line mode without any graphical interface, one may find the semi-graphical mode with one input file, inspired by [10], rather convenient.

3.2 Model datasets

To estimate the functionality of the algorithm, we implemented 3 separate datasets and attempted to fit them in different regimes (see Table 1). Artificial zero-mean normal-distributed noise of 4% was added to each dataset.

Dataset	$z_{0,\ model}$	$z_{0,\ fit}$	σ_{model},	σ_{fit}	χ^2	R
1-peak-1-fit	50	95	5	5	54	0.026
2Λ-2-fit	30, 120	–, –	5, 5	–, –	–	–
2Λ-1-fit	30, 120	120	5, 5	4.9	10	0.019
3Λ/2-2-fit	30, 96	29.8, 95.9	5, 5	4.9, 4.9	1.1	0.021
3Λ/2-1-fit	30, 96	43	5, 5	16	5.3	0.023

Table 1: Summary for model datasets: simple single-gaussian distribution, fitted with single-gaussian curve, and two bi-modal distributions, shifted by 2Λ and 3Λ/2 with attempts to fit them with single-gaussian and two-gaussian distributions both.

3.2.1 Single-gaussian convergence

Dataset "1-peak-1-fit", simple single-gaussian distribution, has shown great consistency with the model distribution, as one can clearly find from the table.

Moreover, we recorded the model parameters during fitting in order to visually estimate the characteristics of found minimum. One can clearly see steps-like improvement of χ^2 at the (Figure 4).

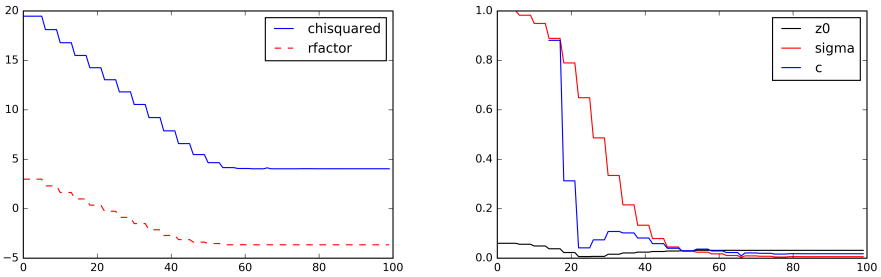

Figure 4: Convergence of 1-peak-1-fit dataset fitting.

The width of each step equals to the number of parameters in the model (4, in our case). In contrast, at the convergence stage one can rarely notice any changes in both χ^2 and R-factor. Also parameters of the model change slightly after approximately half of steps. One can use this estimation for manual interruption of algorithm during the search of plausible initial parameters. It should be noticed that the algorithm shows similar behavior during two-gaussian fitting, which rarely converges without pre-alignment of initial conditions, thus allowing one a powerful tool for the initials search.

3.2.2 Two-gaussian distributions

We have created two bi-modal datasets with 2Λ and $3\Lambda/2$ distances between maxima. The distance in the former distribution was deliberately set a multiple of the SW period in order to illustrate the inefficiency of bi-modal distribution in such a case.

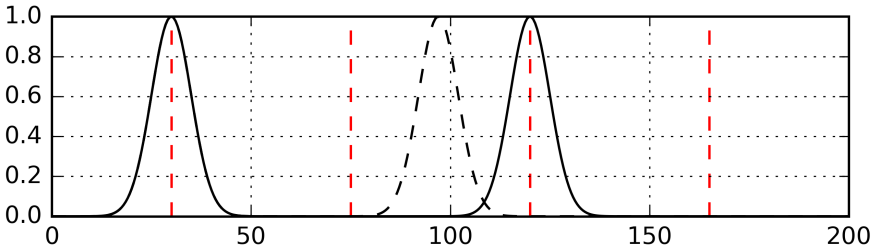

Figure 5: Model atom distribution above the surface. 2Λ dataset shown in solid, $3\Lambda/2$ is dashed. Red dashed lines show the nodes of the SW at the Bragg angle

As one can clearly see from Table 1, 2Λ model can be fitted perfectly with single-gaussian distribution. Moreover, it did not fit any of two-gaussian distributions with reasonably different z_0 or similar σ (during fitting procedure ratio N between areas under gaussians was kept constant, so one could expect $\sigma_1 = \sigma_2$ after fitting, which was not observed). It also should be noted that the amplitude of single-gaussian and two-gaussian fits (now shown), relate as 2:1, proving the fact that single-gaussian curve equals to two similar curves with half amplitude.

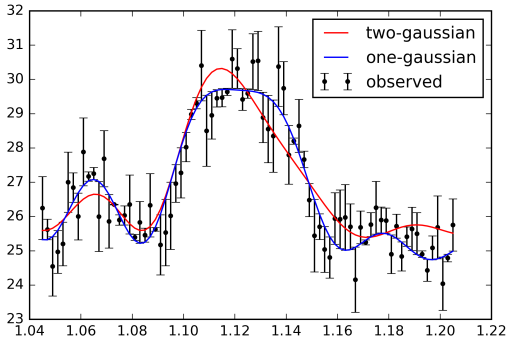

Figure 6: 3Λ/2 1- and 2-fit modes imposed. They could be hardly distinguished by the naked eye, however, numerical indicators clearly show the difference

However, the fitting procedure can clearly distinguish two different gaussian peaks ($\chi^2 = 1.1$ vs. χ^2=5.3) with 3Λ/2 distance, even it is seen nor by the naked eye at (Figure 6) neither by R-factor (2.4% vs. 2.7%).

3.3 Real data fitting

For real-world testing of the program, we took two experimental datasets [11] from a very recent experiment at the ESRF ID10 beamline. Fitted curves show good consistency with the data, both at the qualitative (see Figure 7) and the quantitative level (see Table 2). The obtained results correspond extremely well to the values expected for the objects under investigation.

Dataset	$z_{0,fit}$	σ_{fit}	χ^2	R
SGC_highhum_K	36	8.0	0.026	0.046
PEG_highhum_P, 2-fit	23, 96	8.7, 9.4	0.028	0.043
PEG_highhum_P, 1-fit	60	13.0	0.030	0.046

Table 2: Two experimental datasets from yet unpublished work. Knowledge of the nature of studied membranes allows us to know that these are mono- and bilayers in SGC and PEG cases, respectively. Better consistency of the data in case of PEG-2-fit mode is clearly seen.

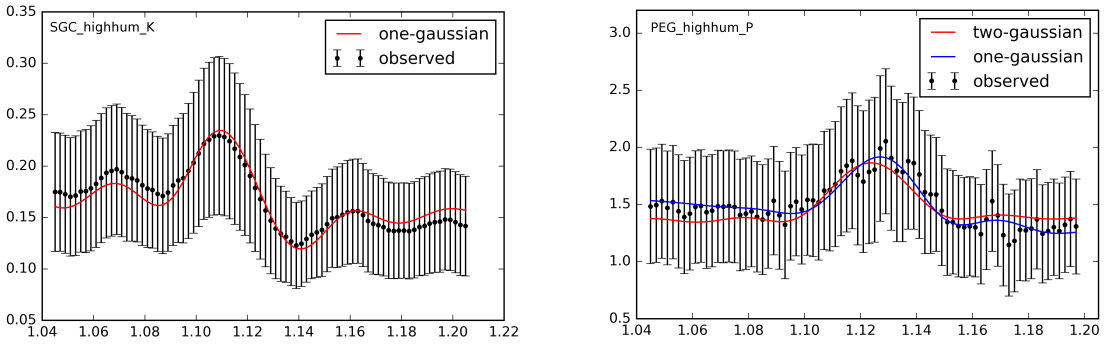

Figure 7: Fluorescence yield curves for both experimental datasets. For values, see Table 2

4 Conclusion and future plans

In conclusion, I have implemented a PYTHON-based software for x-ray standing wave data evaluation. The software was tested on both model and experimental data and has shown its high efficiency and reliability. As a future development of the software, I would suggest making the software public domain and, if demanded, supply it with a graphical user interface.

References

[1] Boris W. Batterman and Henderson Cole. Dynamical Diffraction of X Rays by Perfect Crystals. *Reviews of Modern Physics*, 36(3):681–717, jul 1964.

[2] Jin Wang, Michael J. Bedzyk, Thomas L. Penner, and Martin Caffrey. Structural studies of membranes and surface layers up to 1,000 Å thick using X-ray standing waves. *Nature*, 354(6352):377–380, 1991.

[3] Emanuel Schneck and Bruno Demé. Structural characterization of soft interfaces by standing-wave fluorescence with X-rays and neutrons. *Current Opinion in Colloid & Interface Science*, 20(4):244–252, 2015.

[4] Emanuel Schneck, Ernesto Scoppola, Jakub Drnec, Cristian Mocuta, Roberto Felici, Dmitri Novikov, Giovanna Fragneto, and Jean Daillant. Atom-scale depth localization of biologically important chemical elements in molecular layers. *Proceedings of the National Academy of Sciences*, 113(34):9521–9526, aug 2016.

[5] S. K. Ghose and B. N. Dev. X-ray standing wave and reflectometric characterization of multilayer structures. *Physical Review B*, 63(24):245409, jun 2001.

[6] M. J. Bedzyk and G. Materlik. Two-beam dynamical diffraction solution of the phase problem: A determination with x-ray standing-wave fields. *Physical Review B*, 32(10):6456–6463, 1985.

[7] Sergey a. Stepanov. X-ray server: an online resource for simulations of x-ray diffraction and scattering. *Proceedings of SPIE*, 5536:16–26, 2004.

[8] D. K G De Boer. Glancing-incidence x-ray fluorescence of layered materials. *Physical Review B*, 44(2):498–511, 1991.

[9] Matthew Newville, Antonino Ingargiola, Till Stensitzki, and Daniel B. Allen. LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python. *Zenodo*, 2014.

[10] Wolfgang Kabsch. XDS. *Acta crystallographica. Section D, Biological crystallography*, 66(Pt 2):125–32, feb 2010.

[11] E. Schneck, I. Rodriguez-Loureiro, L. Bertinetti, E. Marin, D. V. Novikov, O. Konovalov, and G. Gochev. In preparation. *Journal of Physics D*.