CTR-structure users' manual

K. Nagai[1] and Y. Wakabayashi[2]
[1] Osaka University
[2] Tohoku University

10th Feb. 2022

Contents

1 What is CTR-structure

This software performs a structure refinement of surfaces within the framework of Bayesian inference based on a CTR scattering dataset. Details of the theoretical background and the performance are reported in some references[1, 2].

Users need to prepare several input files:

1. CTR scattering intensity data file
2. Substrate structure file
3. Surface structure definition file
4. Anomalous scattering factor data.

The file format for each of input file is provided in sec. 3. The third one is the most complicated one as the software allows us to use various constraints.

2 Required environment

The software is compiled by Microsoft Visual Studio 2019 for Windows 10 64bit. The source code is also provided for any other environments.

3 Files needed

The files used without any user modification:

1. `CTR-structure.exe`
2. `f0data.dat`

The files that are modified for user requirements:

1. *int.dat* CTR scattering intensity data file
2. `anomalous.dat` Anomalous scattering factor data.
3. *substrate.dat* Substrate structure file
4. *surf.dat* Surface structure definition file

The filename "`anomalous.dat`" is fixed, while you can name other files written in italic as you like. In this manual, we use the names above for convenience.

File format of each file is explained in the following subsections. Basically, the files are tab delimited text file with CRLF line feed code.

3.1 *int.dat*

Intensity file is basically an input file, but the file format includes a column for calculated intensity. This column is reserved for the output. The software outputs the calculated intensity together with the experimental intensity in the same file format. Users may ignore the calculated intensity column for the analysis.

column number: description

1: h

2: k

3: l

4: Experimentally observed intensity I_{exp}

5: Calculated intensity $I_{\mathrm{calc}} = scale \times |F_{\mathrm{calc}}|^2$. It can be left blank.

6: Statistical error σ_{exp}

3.2 `anomalous.dat`

The anomalous scattering factors f' and f'' are provided in this file.

Each line gives the values for one element. When you do not define the values of f' and f'' for some of the elements, zero is used for them.

column number: description

1: atomic number

2: f'

3: f''

3.3 *substrate.dat*

Substrate structure file gives the structural information of known substrate structure.

The first line gives the information of the bulk unit cell. a, b, c, α, β, and γ are written in the unit of Å and degree.

Each of the following lines provide the structure parameters of each atom in a unit cell.

column number: description

1: atomic index

2: reserved for future extension

3: atomic number

4: x

5: y

6: z

7: B (Å^2)

8: occupancy

x, y and z are written in the fractional coordinate. z axis points outward. The origin is shown

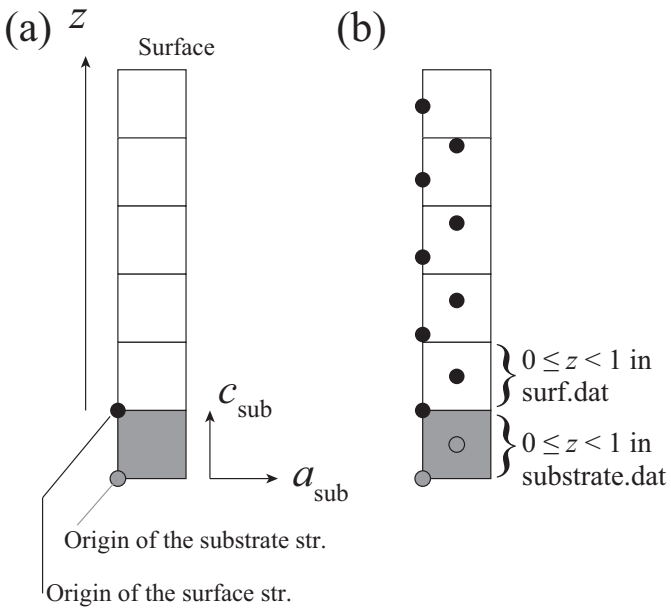

Fig. 1 (a) Coordinations of the substrate structure and surface structure. (b) When you use $0 \leq z < 1$ for the substrate file, the atomic positions are in the gray unit cell, which is $-1 \leq z < 0$ in the film coordination.

in Fig. 1 (a).

It is easy to make 1-u.c.-thick gap or unphysically superposed atoms at the interface between the "substrate.dat" and "surf.dat." It is vitally important to make the interfacial structure coherent.

3.4 *surf.dat*

Surface structure file gives the structural parameters, select which parameter to be modified, and impose some constraints. *This file is usually very long and complicated. We prepared a supporting softwares that converts the substrate.dat+surf.dat into cif file, check the atomic overlap, etc. See the manual for the supporting software for detail.*

The first line defines the parameters independent of each atom, such as scale parameter for overall intensity. This line consists of 8 sets of string and numeric value. The strings are just labels for the numeric values to help users. One can write anything for the string, but do not delete.

The parameters in the first line are listed below.

1. *scale*. *scale*$\times|F_{\rm calc}|^2$ is compared to $I_{\rm exp}$.
2. Switch to select if the vale of *scale* is optimized. 1: modify *scale*, 0: fix *scale*.
3. γ in eqn. (2) below.

4. The switch for the value γ.
5. $\sigma_c$in eqn. (2)
6. The switch for the value σ_c
7. Coefficient for substrate B (intended to be used for low-temperature experiments)
8. The switch for above coefficient

Here, the parameters γ and σ_c are the parameters appear in the following equation, which are used in [2].

$$E(\mathbf{\Theta}) = \frac{1}{N}\left[\sum_{i=1}^{N}\left[\frac{\{I_{\rm exp}(\mathbf{Q}_i) - I_{\rm calc}(\mathbf{Q}_i;\mathbf{\Theta})\}^2}{2\{\sigma(\mathbf{Q}_i)\}^2} - \ln\frac{1}{\sqrt{2\pi}}\frac{1}{\sigma(\mathbf{Q}_i)}\right] + E_{\rm prior}\right], \tag{1}$$

$$\sigma(\mathbf{Q}_i) = \sqrt{\{\gamma I_{\rm calc}(\mathbf{Q}_i;\mathbf{\Theta})\}^2 + \{\sigma_{\rm exp}(\mathbf{Q}_i)\}^2 + \sigma_c^2} \tag{2}$$

The second line is a comment line. The software does not use this line, but assumes to have this line. Do not delete.

The following lines provide the information for each atom.

column number: description

1: Atom index. It must be an integer, and sequence number starting from zero.
2: Atomic number *AtmNum*.
3: N (Number of period. See text.)
4: d (Periodicity in the unit of $\mathbf{c}$. See text.)
5: x
6: y
7: z
8: B (Å^2)
9: Total occupancy of the film, *toccf*. The probability of any element occupies this site.
10: Fractional occupancy of the film, *foccf*. The probability of *AtmNum* appears in all atoms in this site. (detail is described below.)
11: 7-digit switch. Each digit is either 1 or 0. From the left to the right digits, they express if you want to modify $d, x, y, z, B, toccf$ and $foccf$. 1 means *modify it*, and 0 means *fix it*.
12: Atom Index of Mother atom
13: 0-6 (pointing *d-foccf* of Mother atom) } column 12-14: settings for
14: M_d connecting parameters for d (see below)
15: Atom Index of Mother atom
16: 0-6 (pointing *d-foccf* of Mother atom) } column 15-17: settings for

17: M_x connecting parameters for x

$\vdots$

33: minimum value of d

34: maximum value of d

35: minimum value of x

36: maximum value of x

$\vdots$

45: minimum value of $foccf$

46: maximum value of $foccf$

The parameters N and d are used to express periodic structures. The line is repeated N times with the periodicity of d in the unit of the substrate lattice parameter c. If you want to have only one atom, N should be 1 and d can be any value.

There are two occupancy parameters, and a lot of columns for parameter connection. The following subsections are devoted to explain these features.

3.4.1 occupancy parameters

There are two occupancy parameters, *toccf* and *foccf*. The actual occupancy of *AtmNum* on the site is given by $toccf \times foccf$. The total occupancy parameter *toccf* is usually used to describe the surface roughness, and the fractional occupancy parameter *foccf* is usually used to describe the interface atomic mixing.

3.4.2 Connecting parameters

During the structure refinement, we often want to connect several parameters in parallel, antiparallel, etc. The idea of *parameter connection* allows us to write various relations among the structural parameters.

Here, we use the terms *Mother atom* and *Daughter atom* for the parameter connection. The structural parameters for the mother atom, θ_i^{M} (M is the atom index of the mother atom, $i = 0 \cdots 6$ denote $d, x, y, z, B, toccf$ and $foccf$, respectively), are modified by $\Delta\theta_i^{\mathrm{M}}$ in the MC calculation directly. Those for daughter atom, θ_j^{D}, are not modified in the MC calculation directly. Instead, when the θ_i^{M} is modified, θ_j^{D} is also shifted by $M_j \times \Delta\theta_i^{\mathrm{M}}$, where M_j is the multiplier defined in the third column in the one set of connection.

$$\Delta\theta_j^{\mathrm{D}} = M_j \Delta\theta_i^{\mathrm{M}}$$

One component of the structural parameter in a daughter atom requires three columns.

The connection settings are written in the line of daughter atom. Each structure parameter θ_j^{D} can be connected to different mother atom's different parameter.

Example 1:

Atom Ind.		7-digit switch	connection x	connection y	connection z			
20	$\cdots$	0001000						$\cdots$
21	$\cdots$	0000000			20	3	1.0	$\cdots$

The parameter θ_z^{21} is connected to θ_z^{20} with the connecting ratio 1.0. This input results in the same atomic shift in the z direction for atom indexes 20 and 21. Note that atom 21 has the switch '0' for the z parameter to be connected to atom 20.

Example 2:

Atom Ind.		7-digit switch	connection x	connection y	connection z			
21	$\cdots$	0001000						$\cdots$
22	$\cdots$	0000000			21	3	-1.0	$\cdots$

The parameter θ_z^{22} is connected to θ_z^{21} with the connecting ratio -1.0. This input results in the antiparallel atomic shift in the z direction for atom indexes 21 and 22.

The connection works only for one step (granddaughter is undefined). If we write

(Bad example)

Atom Ind.		7-digit switch	connection x	connection y	connection z			
20	$\cdots$	0001000						$\cdots$
21	$\cdots$	0000000			20	3	1.0	$\cdots$
22	$\cdots$	0000000			21	3	-1.0	$\cdots$

the parameter θ_z^{22} is not connected to θ_z^{20}. It is needed to connect with the original mother directly.

Example 3:

Atom Ind.		7-digit switch	connection x	connection y	connection z			
20	$\cdots$	0111000						$\cdots$
21	$\cdots$	0111000						$\cdots$
22	$\cdots$	0000000			20,21	1,2	0.5,-0.5	$\cdots$

In this case, the daughter parameter θ_z^{22} has two mother parameters. The parameter modification $\Delta\theta_z^{22}$ is given by $0.5 \cdot \Delta\theta_x^{20} - 0.5 \cdot \Delta\theta_y^{21}$. This feature may be useful for the modeling of

octahedral rotation in metal oxides. The number of mother atom is practically unlimited.

Note that the number of the mother atoms, parameter index, and the multiplier [in this case, (20, 21), (1,2) and (0.5, −0.5) all consist of two parameters] must be the same. If you wrote different number of parameters, you will see an error message such as "The number of connection conditions of atoms must be the same."

4 Output files

There are a number of output files. Here, we show the output files created by the CTR-structure software.

The output files are

1. **output_name**.log Simple list of the cost function $E(\mathbf{\Theta})$ defined in [1].
2. **output_name**.swap List of the inverse temperature β index for each replica.
3. **output_name**-n.par Record of the parameter history. n denotes the index of β. $n = 0$ shows the parameters for the largest β. The output is recorded every 10 MC steps. The interval is modified by editing the line of "#define RECORD_NUM" in the C++ source file.
4. **output_name**-n.pos The parameter list at the final cycle of the MC calculation. The format is the same with *surf.dat*, and one can use **output_name**-0.pos as the next input file.
5. **output_name**-n.dat The calculated CTR intensity based on the structure given in .pos file. The format is the same with *int.dat*, and the 5th column is replaced with the calculated intensity.

The **output_name** is given by the users as a command line argument. See sec. 7.

The R value we often use is not given by the software. You can easily calculate it from **output_name**-0.dat by using EXCEL.

5 Updating input files

Use **output_name**-0.pos as the next *surf.dat*. The file format of them is identical to each other.

6 Prior probability

Rectangular prior probability[2] is given in the *surf.dat* as the definition of minimum and maximum for each parameter. Other prior probability is applied by editing the C++ source. Modify the function *calcE* for this purpose.

7 Command line calculaiton

Store the files

- *CTR-structure.exe*
- `anomalous.dat`
- `f0data.dat`
- *bulk.dat*
- *film.dat*
- *int.dat*

in a working directory. From the command line, one can start calculation by

> `CTR-structure bulk.dat film.dat int.dat output_name` N $\beta_{\max}$ $\beta_{\min}$

where N denotes the iteration number, $\beta_{\max/\min}$ mean the maximum/minimum values of β, and **output_name** means the directory name for the output files. The data directory is created by the software. The number of replica is defined in the C++ source code, and we usually use 128 replicas. Since **output_name**-n.par, pos, dat files are created as many as the number of replica, the typical number of the output file is 400. (In the default setting, **output_name**-n.par files are saved only for $n = 8m$, where m is integer, to reduce the file size.)

8 Example

Example of the data/setting files together with the windows executable file are provided in the `example` directory. Just start `example.bat` performs a small calculation taking 30 sec. or so.

9 License

This software is distributed free of charge for academic, scientific, educational, and non-commercial uses. Users belonging to commercial enterprises may also use this software at no cost as long as users directly use this package exclusively for their own studies.

Permission to use this software is hereby granted under the following conditions:

1. Results obtained by using CTRstructure may be used in any publications provided that its use is explicitly acknowledged. A suitable reference for this software is
   - K. Nagai, M. Anada, Y. Nakanishi-Ohno, M. Okada, and Y. Wakabayashi, "Robust surface structure analysis with reliable uncertainty estimation using the exchange Monte Carlo method", *J. Appl. Cryst.* **53** 387-392 (2020).
2. You should not redistribute any copy of the files distributed to you unless you have a written permission from the owner.

Part of these terms may be changed without any prior announcement. This software is provided "as is" without any expressed or implied warranty.

example) The crystal truncation rod intensity data were collected at BL13XU, SPring-8, Japan. The data were analyzed by using CTR-structure [1].

We hope you do better science with this software.

Acknowledgements

A part of this software development was supported by a Grant-in-Aid for Scientific Research (Japan Society for the Promotion of Science KAKENHI, grant Nos. JP26105008 and JP26287080). It was also supported by the MEXT Elements Strategy Initiative to Form Core Research Center (Grant No. JPMXP0112101001).

References

[1] M. Anada, Y. Nakanishi-Ohno, M. Okada, T. Kimura, and Y. Wakabayashi. Bayesian inference of metal oxide ultrathin film structure based on crystal truncation rod measurements. *Journal of Applied Crystallography*, 50(6):1611–1616, 2017.

[2] K. Nagai, M. Anada, Y. Nakanishi-Ohno, M. Okada, and Y. Wakabayashi. Robust surface structure analysis with reliable uncertainty estimation using the exchange Monte Carlo method. *J. Appl. Cryst.*, 53:387–392, 2020.