



Surface Science Prospectives

A standard format for reporting atomic positions in measured or calculated surface structures: The CIF file

Laurence D. Marks

Chair, Commission on Electron Crystallography, IUCR
 Department of Materials Science and Engineering, Northwestern University, Evanston, IL 60201, USA

ARTICLE INFO

Article history:

Received 22 October 2009

Accepted 16 February 2010

Available online 3 March 2010

Keywords:

Density functional calculations

Electron–solid diffraction

X-ray scattering

Diffraction

Reflection

ABSTRACT

In his editorial in this issue, the editor-in-chief emphasizes the editorial policy that any paper which involves a crystallographic structure (whether experimentally measured or theoretically calculated) must also include a complete listing of all the atomic positions within the crystal structure, either as supporting information or directly within the paper itself. He also strongly recommends that the complete crystallographic data set be included as supporting information. At the request of the editor-in-chief, I outline here the reasons why this is scientifically desirable. Furthermore, I propose here that the Surface Science community adopt the same standard format for reporting these as is already widely used in bulk crystallography publications, namely the inclusion of a Crystallographic Information Format file (or CIF file) as supporting information. Finally, I describe the details of this specific file format, with illustrative examples.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

We are seeing an increase in the power of techniques to obtain structural information about surfaces, sometimes definitive structure determinations, sometimes more restricted but enough to propose a full structural model. Unfortunately many of the papers that report such data fail to tabulate all the atomic positions within the paper, probably because this can be so excessive.

We have also seen an explosion in the use of Density Functional Theory (DFT) for calculating surface structures. While in most cases representative figures of the structures are shown within the publication, many of these papers also fail to tabulate all of the atomic positions.

A fundamental principle in scientific publication is that sufficient information be made available to the readers so that they can both judge whether the conclusions are valid, and reproduce them if desired. If the raw data for the experiments and/or the atomic positions are not available either within the paper, as supplemental material or deposited in a web repository this fundamental principle is violated. It is proper editorial policy to demand this. I am pleased that the journal *Surface Science* has taken this opportunity to specifically emphasize this with respect to surface structures and related crystallographic data, via the editorial within this issue, and with the invitation for my related contribution herewith. As stated in that editorial, it is required that any paper which reports a crystallographic structure in this journal must also include a complete listing of all its atomic positions, either within the paper of as supporting information.

It is clearly desirable that this be done in a way that is easy to do, and quick to learn and understand. It is also clearly desirable that, if

data is involved in the experimental determination of a structure, that this data also be completely included. The question then arises: what is the best format for presenting such data? Clearly, it is best if the authors all use the same format.

A solution to this problem already exists for bulk structures. Essentially every journal requires deposition of data using the “Crystallographic Information Format” or CIF, established by the International Union of Crystallography in 1990, see <http://www.iucr.org/resources/cif> and [1–4]. The concept is that this file contains sufficient information both to give full details of the atomic structure, as well as in most cases sufficient experimental information (including raw data) so others can check the analysis. It is not that unusual for structures to be reconsidered years after the original publication, and better fits found (e.g. different spacegroups) – for some examples see [5,6]. I propose that the Surface Science community adopt this same file format as a standard. Why reinvent the wheel?

It is telling to consider the following excerpt from the original paper where the CIF format was first described in 1991 [1]:

“There is an increasing need in many branches of science for a uniform but flexible method of archiving and exchanging data in electronic form. Rapid advances in computer technology, coupled with the expansion of local, national and international networks, have fuelled the need for such a facility. The variety and relative inflexibility of existing data exchange formats have inhibited their effective use. This is true even in fields where the basic data requirements are well defined. Problems of data exchange are further exacerbated if the number and nature of data types change rapidly and continuously. Under these conditions specialized and local file formats have proliferated. This diversity was tolerable when electronic data

E-mail address: Laurence.Marks@gmail.com.

transfer was infrequent, or when data processing speeds required file formats finely tuned to specific applications. The developments cited above signal an end to this rationale. A general, flexible, rapidly extensible and universal file format protocol is now essential. It must be machine independent and portable so that accessibility to data items is independent of their point of origin. It must allow new data items to be incorporated without the need to modify existing files.

In addition to archiving data, the use of a universal file would facilitate data exchange between software within a laboratory; between different laboratories; between authors and journals, providing electronic input to the publication process; and between researchers or journals and computerized databases."

The intention of this perspective is twofold: to draw attention to the problem where structures and data are under-reported, and to suggest a solution to it by strongly encouraging deposition of a CIF file with any paper which proposes a surface structure – if no CIF file is deposited, the paper should include a table of all the atomic positions.

2. What is a CIF file

The CIF file format is slightly complicated, but very well documented, see <http://www.iucr.org/resources/cif>, <http://www.it.iucr.org/Ga/contents/> and [1–4]. I will give here only a partial explanation of some of its key features. The structure is based around the slightly earlier concept of a Self-Defining Text Archive and Retrieval (STAR) file [7], and can be broken down into a number of subunits:

text string

string of characters bounded by blanks, single quotes (') double quotes ("), or by semi-colons (;) as the first character of a line

data name

a text string starting with an underline (_) character

data item

a text string not starting with an underline, but preceded by a data name to identify it

data loop

a list of data names, preceded by loop_ and followed by a repeated list of data items

data block

a collection of data names (looped or not) and data items that are preceded by a data code record. A data name must be unique within a data block. A data block is terminated by another data statement or the end of file

data file

a collection of data blocks: the block codes must be unique within a data file

A large number of data names are commonly used (and accepted as part of the CIF dictionary, see <http://www.iucr.org/resources/cif>, <http://www.it.iucr.org/Ga/contents/> and [1–4]) all of which are quite humanly readable. Currently they are primarily targeted towards X-ray diffraction data, but there is room for additional terms (and most existing programs that read CIF files will ignore terms that they don't understand), a point I will return to later. It is useful to give a few examples. The most basic is how one defines the unit cell and a combination of data names and data items. For example, this can be written as:

```
_cell_length_a 3.014950
_cell_length_b 5.222046
_cell_length_c 28.233300
_cell_angle_alpha 90.000000
_cell_angle_beta 90.000000
_cell_angle_gamma 90.000000
```

Note that the numbers are all free-format, which makes them easy to read with a computer program – this is generally true of the CIF format. Atomic positions are reported within a loop structure with additional information such as temperature factors, occupation, valence optionally included, for instance:

```
loop_
_atom_site_label
_atom_site_fract_x
_atom_site_fract_y
_atom_site_fract_z
_atom_site_U_iso_or_equiv
Mg 0.50000000 0.83333333 0.17459595 0.000
Mg 0.00000000 0.00000000 0.73782107 0.000
O 0.00000000 0.66666667 0.95636678 0.000
....
```

Another, relevant example is how one can include experimental diffraction data, again in a standard form with a loop:

```
loop_
_refln_index_h
_refln_index_k
_refln_index_l
_refln_F_meas
_refln_F_sigma
```

1	0	0	169.756706637	14.768994723
2	0	0	60.718573279	4.110587725
2	1	0	52.005509328	19.531694204

....

In crystallography, reciprocal lattice vectors are conventionally always integers, but one could probably use fractional indices as well.

Beyond straight crystallographic information, other terms can be used. For instance, the cif2struct code in the latest version of the Wien2k code ([8] and see <http://www.wien2k.at>) will produce a CIF file with partial information about the DFT method bracketed by a pair of ";" lines:

```
_refine_date '17-10-2009'
_refine_method 'generated from Wien2k code'
_refine_special_details
;
Structure converted from Wien2k struct file, Version 9.1
File Name /home/ldm/Wien/DiLineBig/Ti2gh/Ti2gh.struct
Title 'blebleble'
;
```

Logically one could add other information, for instance the muffin-tin radii for a Wien2k calculation or the maximum plane-wave energy and pseudopotentials used for a code such as VASP.

Various templates exist for different types of bulk diffraction data, see for instance <http://www.journals.iucr.org/services/cif/templates.html>, which can be readily adapted for LEED, SXRD or TED. There are also freely available codes for reading and writing CIF files, see for instance [9], as well as codes to convert between DFT formats and CIF formats (and almost any other), most free. Sometimes there may not be an existing dictionary term, but new ones can be developed any hopefully they will become part of the standard CIF dictionary in the future, see for instance [10]. As an example (not exactly surface related, but comparable) discussions are currently taking place within the Commission on Electron Crystallography of the IUCR (see <http://www.iucr.org/iucr/commissions/ced.html>) on terms for high energy electron diffraction. One suggestion to date is to introduce a data type to define what type of electron diffraction data was collected, for instance

“_electron_diffraction_type” which could have possible values of “selected_area”, “nanodiffraction”, “microdiffraction”, “cbed” with logical extensions for surfaces by adding “leed” and “rheed”.

3. Why deposit a CIF

As stated earlier, depositing information in a CIF (or similar) format satisfies one of the core requirements of science – publication of sufficient data for others to gauge the validity of the results. There are other advantages to using a CIF format. As already mentioned earlier, there are numerous viewers for structures and they all accept this form. There are also many small codes freely available on the web for analyzing CIF files, for instance doing bond-valence sums [11–14] which is a useful method of testing whether a proposed structure is feasible without extensive calculations. Lastly there are large searchable databases to which almost all universities, large companies and national laboratories subscribe, for instance see references [2,15,16]. An example of a full surface structure which is included in one of the standard databases (ICSD) is given in Appendix A.

4. Conclusion

I suggest that it is time for the Surface Science community to adopt the same standard format for the deposition of crystallographic data and atomic positions as is already used widely for bulk crystal structures: inclusion of a CIF file. Why reinvent the wheel? This CIF file format is simple and easy to learn and use. It should be used for both experimental and calculated structures when giving the detailed atomic positions.

Acknowledgements

I would like to thank Peter Blaha, Martin Castell, Ulrike Diebold, Mark Hersam, Sokrates Pantelides, Kenneth Poeppelmeier, Dilano Saldin, George Schatz and Nichola Spaldin for useful comments and for supporting the concept of depositing CIF files for surface structures, as well as the editors of Surface Science.

Appendix A

Given below is the $c(4 \times 2)$ structure of SrTiO_3 (001) [17] as a CIF file, which is entry 97008 in the ICSD database.

```

data_97008-ICSD
_database_code_ICSD 97008
_audit_creation_date 2004-10-01
_chemical_name_systematic 'Titanium oxide - (001) overlayer'
_chemical_formula_structural 'Ti O2'
_chemical_formula_sum 'O2 Ti1'
_exptl_crystal_density_diffn 0.67
_cell_measurement_temperature 293
#Default value included by FIZ Karlsruhe
_cell_measurement_pressure 101.325
#Default value included by FIZ Karlsruhe
_publ_section_title
;
Surface structures of Sr Ti O3 (001): a (Ti O2)-rich reconstruction
with a c(4 X 2) unit cell
;
loop_
_citation_id
_citation_journal_full
_citation_year
_citation_journal_volume
_citation_page_first
_citation_page_last
_citation_journal_id_ASTM
primary 'Journal of the American Chemical Society' 2003 125
10050 10056 JACSAT
loop_
_publ_author_name
'Erdman, N.'
'Warschkow, O.'
'Asta, M.'
'Poeppelmeier, K.R.'
'Ellis, D.E.'
'Marks, L.D.'
_cell_length_a 15.62
_cell_length_b 7.81
_cell_length_c 12.932
_cell_angle_alpha 90(0)
_cell_angle_beta 90(0)
_cell_angle_gamma 90(0)
_cell_volume 1577.6
_cell_formula_units_Z 8
_symmetry_space_group_name_H-M 'P 1'
_symmetry_Int_Tables_number 1
loop_
_symmetry_equiv_pos_site_id
_symmetry_equiv_pos_as_xyz
1 'x, y, z'
loop_
_atom_type_symbol
_atom_type_oxidation_number
Ti4+ 4
O2- -2
loop_
_atom_site_label
_atom_site_type_symbol
_atom_site_symmetry_multiplicity
_atom_site_Wyckoff_symbol
_atom_site_fract_x
_atom_site_fract_y
_atom_site_fract_z
_atom_site_B_iso_or_equiv
_atom_site_occupancy
_atom_site_attached_hydrogens
Ti1 Ti4+ 1 a 0.3761 0. 0.9409 . 1. 0
Ti2 Ti4+ 1 a 0.6239 0. 0.9409 . 1. 0
Ti3 Ti4+ 1 a 0.8761 0.5 0.9409 . 1. 0
Ti4 Ti4+ 1 a 0.1239 0.5 0.9409 . 1. 0
Ti5 Ti4+ 1 a 0.5 0.2461 0.944 . 1. 0
Ti6 Ti4+ 1 a 0.5 0.7539 0.944 . 1. 0
Ti7 Ti4+ 1 a 0. 0.7461 0.944 . 1. 0
Ti8 Ti4+ 1 a 0. 0.2539 0.944 . 1. 0
O1 O2- 1 a 0. 0. 0.9233 . 1. 0
O2 O2- 1 a 0.5 0.5 0.9233 . 1. 0
O3 O2- 1 a 0.5 0. 0.9162 . 1. 0
O4 O2- 1 a 0. 0.5 0.9162 . 1. 0
O5 O2- 1 a 0.2455 0. 0.9146 . 1. 0
O6 O2- 1 a 0.7545 0. 0.9146 . 1. 0
O7 O2- 1 a 0.7455 0.5 0.9146 . 1. 0
O8 O2- 1 a 0.2545 0.5 0.9146 . 1. 0
O9 O2- 1 a 0.3771 0.2077 1. . 1. 0
O10 O2- 1 a 0.6229 0.7923 1. . 1. 0

```

O11 O2- 1 a 0.6229 0.2077 1. . 1. 0
 O12 O2- 1 a 0.3771 0.7923 1. . 1. 0
 O13 O2- 1 a 0.8771 0.7077 1. . 1. 0
 O14 O2- 1 a 0.1229 0.2923 1. . 1. 0
 O15 O2- 1 a 0.1229 0.7077 1. . 1. 0
 O16 O2- 1 a 0.8771 0.2923 1. . 1. 0
 #End of data_97008-ICSD

References

- [1] S.R. Hall, F.H. Allen, I.D. Brown, *Acta Crystallographica Section A: Foundations* 47 (6) (1991) 655.
- [2] S. Hall, *Acta Crystallographica Section A: Foundations* 54 (6 Part 1) (1998) 820.
- [3] I.D. Brown, B. McMahon, *Acta Crystallographica Section B: Structural Science* 58 (3 Part 1) (2002) 317.
- [4] *International Tables for Crystallography Volume G: Definition and Exchange of Crystallographic Data*, IUCR, Chester, 2005 Vol. G.
- [5] A. Spek, *Journal of Applied Crystallography* 36 (1) (2003) 7.
- [6] A. Spek, *Acta Crystallographica Section D: Biological Crystallography* 65 (2) (2009) 148.
- [7] S.R. Hall, *Journal of Chemical Information and Computer Science* 31 (2) (1991) 326.
- [8] K.S.P. Blaha, G.K.H. Madsen, D. Kvasnicka, J. Luitz, WIEN2k, An Augmented Plane Wave + Local Orbitals Program for Calculating Crystal Properties, Universität Wien, Austria, 2001.
- [9] J.D. Westbrook, S.-H. Hsieh, P.M.D. Fitzgerald, *Journal of Applied Crystallography* 30 (1) (1997) 79.
- [10] B.H. Toby, R.B. Von Dreele, A.C. Larson, *Journal of Applied Crystallography* 36 (5) (2003) 1290.
- [11] I.D. Brown, R.D. Shannon, *Acta Crystallographica Section A: Foundations* 29 (3) (1973) 266.
- [12] I.D. Brown, D. Altermatt, *Acta Crystallographica Section B: Structural Science* 41 (AUG 1985) 244.
- [13] N.E. Brese, M. O'Keeffe, *Acta Crystallographica Section B: Structural Science* 47 (1991) 192.
- [14] M. O'Keeffe, N.E. Brese, *Journal of the American Chemical Society* 113 (9) (1991) 3226.
- [15] F. Allen, *Acta Crystallographica Section A: Foundations* 54 (6 Part 1) (1998) 758.
- [16] F. Allen, *Acta Crystallographica Section B: Structural Science* 58 (3 Part 1) (2002) 380.
- [17] N. Erdman, O. Warschkow, M. Asta, K.R. Poeppelmeier, D.E. Ellis, L.D. Marks, *Journal of the American Chemical Society* 125 (33) (2003) 10,050.