Title
5-bit classification in crystal classes and genetic code

Author
Giuliano Bettini

Abstract
I show that 5 bits, or better properties, should be enough to classify the 32 crystal classes of 7 crystal systems, and nothing more.
From a rigorous point of view, each of the 32 bit sequences (from 00000 to 11111) must unambiguously identify a class, and each class must unambiguously identify a bit sequence.
In this article, which aims to be very simple, easy, and understandable, I demonstrate the thesis by connecting the various bits to properties already present and known in crystallography.
I use the same approach to show how even in the genetic code 5 properties or entities or bits are used to create complex structures.
I start from the 64 codons table (nucleotide triplets U C A G in RNA, with T substituting U in DNA) that I then examine in a sub-version with 32 codons, still able to codify all 20 amino acids that contribute to the formation of proteins. Considering 32 codons to show symmetries can make sense in an evolutionary process of modifying the genetic code, starting from an ancestral code.
I show in this way that there are analogies between the symmetries of the crystal classes and the genetic code.
The basic idea is that Nature, having identified a “motif” that works, uses it in several fields.

5-bit classification in crystal classes and genetic code

1. 5-bit encoding 32crystal classes

There are many properties of minerals, classified into 32 “classes”. Simply recalling them, they are called and are symmetry planes, axes of symmetry or rotational symmetries, symmetry center, etc. etc.
There are many.
I think instead that 5 properties or entities or bits should be enough to classify the 32 crystal classes, and nothing more.
This is a topic that definitely interests me, and one that I have thought about a lot.
I can't say how long, I think 40 years or more.
I forget it, but every now and then it comes back to mind and I take it up again.
What I certainly remember very well is how it was born: without any particular studies or intuitions.
Simply from the fact that I said to myself:

“32? Thirty-two is 5 bits.”
The fact that 5 bits are enough is not significant in itself. If I have 4 “things”, I can classify them as 2 bits, 00 01 10 11. If I have 32 things, I can classify them as 5 bits, from 00000 00001 00010 and so on, up to 11111.
What is significant is this:
the classification of minerals, ordered and studied in the 32 crystal classes, requires only and exclusively 5 properties.
I say that 5 bits should be enough, each having a very precise meaning, five properties, geometric or physical, or other. I have indicated these bits mnemonically as 342mc, from the most significant bit to the least significant bit.
Strictly speaking, each of the 32 bit sequences (examples: 30000, or 042m0, etc.) must unambiguously identify a class, and each class must unambiguously identify a bit sequence. In other words, there must be a one-to-one correspondence.
Strictly speaking, one could help oneself through the “generators”, [1] Bilbao Server.
I have written a lot about this topic, in particular on Libvrna an article, Libvrna2, “Mother Nature and the 32 crystal classes”. I wrote an article on Mindat. I wrote a book “Crystallography”. There are still unsatisfactory elements.

But this is meant to be a very simple and summary article.
I deliberately use an easy, understandable, immediate approach.
Let's say that empirically I place the 32 crystal classes in an 8x4 table (matrix), 8 rows and 4 columns. Empirically yes, but with a certain reasoning and with a certain objective in mind.
This is the arrangement of Fig1. The symbols are the Hermann Mauguin symbols of the 32 classes, arranged in a table.

1	**1_**	**m**	**2/m**
2	**222**	**mm2**	**mmm**
4	**4_**	**4mm**	**4/m**
422	**432**	**4_2m**	**4/mmm**
3	**3_**	**3m**	**3_2/m**
32	**622**	**6_m2**	**6/mmm**
6	**6_**	**6mm**	**6/m**
23	**m3**	**4_3m**	**m3m**

Fig1 The 32 crystal classes arbitrarily placed in a table

Always empirically, with 3 bits for the 8 rows and 2 bits for the 4 columns I identify the address of the 8x4=32 boxes. Then I add these bits. In the first row at the top, the bits that identify the columns. In the first column, on the left, the bits that identify the rows.

bit		**00** axes	**0c** center	**m0** plane	**mc** m + c
000	No rotational symmetry	**1**	**1_**	**m**	**2/m**
002		**2**	**222**	**mm**	**mmm**
040		**4**	**4_**	**4mm**	**4/m**
042		**422**	**432**	**4_2m**	**4/mmm**
300		**3**	**3_**	**3m**	**3_2/m**
302		**32**	**622**	**6_m2**	**6/mmm**
340		**6**	**6_**	**6mm**	**6/m**
342		**23**	**m3**	**4_3m**	**m3m**

Fig2 The same, identified by 5 bits

Why do I do all this? Let's say I've gone crazy, or that I feel like this, or let's say that I already see the elements for a classification.
The classification that I intuit, or <u>hope</u> , is that the classes can be identified by bits as indicated in Fig2.
To verify this or to demonstrate that it is possible, I have useful "property tables", yellow, blue, pink, purple, which are the following. They are properties of the various classes, already known from crystallography.

To do this, for convenience I use as a basis the same arrangement as Fig1.
On this I note, with the help of color, the properties of the various classes.

1	1_	m	2/m
2	222	mm2	mmm
4	4_	4mm	4/m
422	432	4_2m	4/mmm
3	3_	3m	3_2/m
32	622	6_m2	6/mmm
6	6_	6mm	6/m
23	m3	4_3m	m3m

Fig 3 Axes only (enantiomorphic, chiral groups).

Yellow reminds of the following property: these classes in yellow have only rotational symmetries as crystallographic symmetries.

1	1_	m	2/m
2	222	mm2	mmm
4	4_	4mm	4/m
422	432	4_2m	4/mmm
3	3_	3m	3_2/m
32	622	6_m2	6/mmm
6	6_	6mm	6/m
23	m3	4_3m	m3m

Fig4 The 22 non-pyro classes (10 pyro classes), or, 22 non-polar classes (10 polar classes).

The blue color recalls the following property: these classes are non-polar. A property that I indicate with the bit "c" because, although it is not the classical centrosymmetry, it indicates for these non-polar crystals a sort of balance with respect to the center. I can think of it as a generalized centrosymmetry, obviously including the classical centrosymmetry which is non-polar by its nature.

1	1_	m	2/m
2	222	mm2	mmm
4	4_	4mm	4/m
422	432	4_2m	4/mmm
3	3_	3m	3_2/m
32	622	6_m2	6/mmm
6	6_	6mm	6/m
23	m3	4_3m	m3m

Fig5 Planes, pink. Classes with planes of symmetry.

The pink indicates the following property: these classes have symmetries with respect to planes, one or more symmetry planes.

1	1_	m	2/m
2	222	mm2	mmm
4	4_	4mm	4/m
422	432	4_2m	4/mmm
3	3_	3m	3_2/m
32	622	6_m2	6/mmm
6	6_	6mm	6/m
23	m3	4_3m	m3m

Fig6 Purple. Laue groups, the 11 centrosymmetric classes.

The purple therefore indicates that these 11 classes have the center of symmetry, 11 centrosymmetric classes or Laue Groups. (The 8 in the last column are also maximally symmetric because (Fig5) they also have m).

What can we see from the property tables?
We can see that apparently the classification I was hoping for in Fig2 is already ready, and is foreseen by the properties already known in crystallography.

Let's examine things in more detail.
For example, we can see that as regards the third column of Fig2, I can indicate all the classes in pink, or rather I can indicate them with the bit m, since in fact all of them (Fig5) have the property m of symmetry planes (in pink).
And yet all of them would have the right to be identified with m0 , with the single bit m, if they possessed that and only that property, of having that symmetry m and no other.

Likewise, all the classes in the first column of Fig2 have the yellow property of “only axes”. However, they would be entitled to be identified with 00 if they possessed that and only that property, of having that symmetry and no other.

Likewise, for the second column of Fig2, I can indicate all the classes in blue, since in fact all of them (Fig5) have the property (Fig2) of c-bits, non-polar, in blue.
However, they would have the right to be identified with 0c if they possessed that and only that property c, of having that symmetry c and no other.

Finally, for the fourth column of Fig2, I can indicate all the classes in purple, since they all (Fig6) actually have the property of having m and c at the same time. Therefore, there is no ambiguity of attribution on these, I can actually indicate them with mc and it is then obvious that they also have m (in pink) and also have c (in light blue).

To make a long story short, by making a comparison with the tables, yellow, blue, pink, purple, of the properties, we find that all the classes are correctly placed in the desired 5-bit identification box. But there are some, let's say, anomalies or excesses of attribution.

Since a picture always speaks better than 1000 words, I show it with a summary image. The image clearly highlights the expected positioning of the various classes, and the attribution anomalies to the side.

bit		00 axes		0c center		m0 plane		mc m + c
0	No rotational symmetry	1		1_	1_	m		2/m
002	2	2		222	222	mm2		mmm
040	4	4		4_		4mm		4/m
042	4 2	422	422	432	432	4_2m	4_2m	4/mmm
300	3	3		3_	3	3m		3_2/m
302	3 2	32	32	622	622	6_m2	6_m2	6/mmm
340	3 2	6		6_	6_	6mm		6/m
342	2 3	23	23	m3	m3 m3	4_3m	4_3m	m3m

Fig7 Properties and anomalies or excesses of attribution.

So for example the class mm2 has the property of having, in addition to the axis 2, only the bit m of symmetry planes. Therefore it corresponds to the bit sequence 002m0.
Class 3 has the property of having axis 3 as a property, and that's it, so it corresponds to the bit sequence 30000.
Etc.
While other classes present anomalies.
For example, let's take 6_m2.
I'll go over the process again.
By comparison with the four property tables we see (Fig5) that one table indicates that 6_m2 has the property m.

1	1_	m	2/m
2	222	mm2	mmm
4	4_	4mm	4/m
422	432	4_2m	4/mmm
3	3_	3m	3_2/m
32	622	6_m2	6/mmm
6	6_	6mm	6/m
23	m3	4_3m	m3m

Fig8 The class 6_m2 has the pink property m, planes

Then there is another table (Fig4) which still indicates 6_m2 but with the property c (light blue, non-polar)

1	1_	m	2/m
2	222	mm2	mmm
4	4_	4mm	4/m
422	432	4_2m	4/mmm
3	3_	3m	3_2/m
32	622	6_m2	6/mmm
6	6_	6mm	6/m
23	m3	4_3m	m3m

Fig9 The class 6_m2 has the blue property c, non-polar

This class therefore has bits 302, because it is in row 302 and has bit m, because it is in the pink table, m planes. So it is (would be) 302m0.
But it has the c bit, because it is in the blue, nonpolar table. So there is ambiguity (?).
How can I assign 302m0 to this class?
Empirically it is solved like this: it cannot be 302mc because it is already occupied, and neither is 3020c already occupied.
Strictly speaking, one could conclude that 302mc is correct, and everything else (attribution anomalies) is there but is also a consequence of 302mc.

Likewise , I believe, all the other classes are easily explained.

As a conclusion, it is appropriate to highlight and explicitly write the 5-bit addresses of the various boxes.

bit		00 axes	0c center	m0 plane	mc m + c
000	No rotational symmetry	00000	0000c	000m0	000mc
002	2	00200	0020c	002m0	002mc
040	4	04000	0400c	040m0	040mc
042	4 2	04200	0420c	042m0	042mc
300	3	30000	3000c	300m0	300mc
302	3 2	30200	3020c	302m0	302mc
340	3 2	34000	3400c	340m0	340mc
342	2 3	34200	3420c	342m0	342mc

Fig10 The 5-bit addresses of the various mailboxes

...and what are (should be) the classes to be assigned to the various boxes.

bit		00 axes	0c center	m0 plane	mc m + c
000	No rotational symmetry	1	1_	m	2/m
002	2	2	222	mm	mmm
040	4	4	4_	4mm	4/m
042	4 2	422	432	4_2m	4/mmm
300	3	3	3_	3m	3_2/m
302	3 2	32	622	6_m2	6/mmm
340	3 2	6	6_	6mm	6/m
342	2 3	23	m3	4_3m	m3m

Fig11 The class in each box

2. Appendix. Memo on why bits 342

Here I want to emphasize the 8 “axes only” classes and the resulting 8 rows.
8 means 3 bits.
What is the meaning?
It can be observed that in crystallography the fundamental rotations, among the legal ones, are 2, 3, 4 (180°, 120°, 90°). In fact the rotation 6 is obtained and is the combination 2+3.
It can then be observed that the legal rotations 2, 3, 4 taken individually or in pairs, form no more than 8 combinations which, graphically recalled, are those of Fig 12, first column.
(Why only these? Because that's how crystals are, other strange combinations like axis 4 + axis 3 at 90° are not symmetries that can exist).

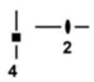
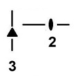
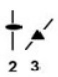
Combinations	Meaning	Classes only axes	3 bit
No rotational symmetry	No axis	monohedron no symmetry 1 00000	000
	Axis 2 alone	sphenoid 2 00200	002
	Axis 4 alone	tetragonal pyramid 4 04000	040
	Axis 4 + Axis 2 at 90°	tetragonal trapezoid 422 04200	042
	Axis 3 alone	trigonal pyramid 3 30000	300
	Axis 3 + Axis 2 at 90°	trigonal trapezoid 32 30200	302
	Axes 3+2 (axis 6)	hexagonal pyramid 6 34000	340
	Axes 3+2 skewed (cube)	tetartoid 23 T 34200	342

Fig12 Eight classes with only “axes” symmetry identified with 3 bits

In fact these combinations with 2, 3, 4 are the legal ones present in crystals, their meaning being recalled in Fig12, second column.
The relevant classes of “axes only” appear in the third column and are, in fact, the classes with axial symmetries only.

What can now be observed is that these 8 classes :

1 being eight, they lend themselves as I have already said to be enumerated with 3 bits or invite to be enumerated with 3 bits, in the 8 combinations 000, 001, 010, 011, 100 etc.;
2 three bits are already naturally present with the axial "qualities" 2, 3, 4;
3 identifying 2, 3, 4 as bits, from 2 least significant bits to 3 most significant bits, the sequences 000, 002, 040, 042, 300, 302 etc. are born (Fig. 12, fourth column), which lend themselves naturally to identifying the previous classes;
4 thus also appears the meaning of the presence of the "strange" bit 4. It is either the rotation 4 itself (in 040 and 042), or axes 2 and 3 respectively parallel (340) or skewed (342, symmetry of the cube).
(I don't know if it is possible to give a more precise crystallographic meaning to bit 4.)

From this, in conclusion, follows the classification of the 32 classes with 5 bits.
There are 3 bits (bits 2 3 and 4) to represent eight rows with the axis identification for each row (000 002 040 042 300 302 340 342)

There are 2 more bits (m and c) to represent the symmetries in each column, 00, 0c, m0, mc (00 no symmetries, 0c c symmetry, m0 m symmetry, or mc both symmetries).
Total: 4x8=32.

3. Appendix. 5-bit encoding, genetic code

This paragraph is written with the intent to show how even in the genetic code five properties or entities or bits are used to create complex structures.
The basic idea is that Nature, having identified a "pattern" that works, uses it in multiple fields.

Some preliminary information can be obtained ex. from [2].

The genetic code is the set of rules by which the information encoded in the nucleotides that make up genes is translated for the synthesis of proteins in cells.
The genetic code is the set of rules used by living cells to translate information encoded within genetic material (DNA or RNA sequences of nucleotide triplets or codons) into proteins. Translation is accomplished by the ribosome, which links amino acids in an order specified by messenger RNA (mRNA), using transfer RNA (tRNA) molecules to carry amino acids and to read the mRNA three nucleotides at a time. The genetic code is highly similar among all organisms and can be expressed in a simple table with 64 entries.
The codons specify which amino acid will be added next during protein biosynthesis. With some exceptions, a three-nucleotide codon in a nucleic acid sequence specifies a single amino acid. The vast majority of genes are encoded with a single scheme (see the RNA codon table). That scheme is often called the canonical or standard genetic code, or simply the genetic code, though variant codes (such as in mitochondria) exist.

Nucleotides. Codons.
So what is a codon? A codon is a sequence of three adjacent nitrogenous bases (and consequently three nucleotides) present in messenger RNA. The discovery that triplets of nucleotides (codons) were the coding units at the base of the genetic code was made by the British scientist F. Crick and the American biologist JD Watson.

Degeneration
Degeneracy refers to the redundancy of the genetic code, that is, two or more codons correspond to the same amino acid. The genetic code is redundant, but, nevertheless, there is no ambiguity in it (see the tables below). For example, both the codons GAA and GAG specify glutamic acid (redundancy), but neither specifies any other amino acid (no ambiguity).
A codon is said to be "quadruple degenerate" if any nucleotide in its third position codes for the same amino acid (for example, UCA, UCC, UCG, and UCU, all corresponding to serine); it is said to be "doubly degenerate" if only two of the four bases in its third position code for the same amino acid (for example, AAA and AAG, corresponding to lysine). In doubly degenerate codons, the equivalent nucleotides in the third position are always either two purines (A/G) or two pyrimidines (C/U).

Having completed this introductory summary, let us consider the 64-codon code as found in the literature. Example, 64 codons table from [3].

Second Letter

1st letter	U		C		A		G		3rd letter
	UUU	Phe	UCU		UAU	Tyr	UGU	Cys	U
U	UUC		UCC	Ser	UAC		UGC		C
	UUA	Leu	UCA		UAA	Stop	UGA	Stop	A
	UUG		UCG		UAG	Stop	UGG	Trp	G
	CUU		CCU		CAU	His	CGU		U
C	CUC	Leu	CCC	Pro	CAC		CGC	Arg	C
	CUA		CCA		CAA	Gln	CGA		A
	CUG		CCG		CAG		CGG		G
	AUU		ACU		AAU	Asn	AGU	Ser	U
A	AUC	Ile	ACC	Thr	AAC		AGC		C
	AUA		ACA		AAA	Lys	AGA	Arg	A
	AUG	Met	ACG		AAG		AGG		G
	GUU		GCU		GAU	Asp	GGU		U
G	GUC	Val	GCC	Ala	GAC		GGC	Gly	C
	GUA		GCA		GAA	Glu	GGA		A
	GUG		GCG		GAG		GGG		G

Fig13 Codons table, from [3]

To continue and to simplify the reasoning on regularities, I consider (Fig14) a 32-codon code. Note that it includes 32 of the 64 codons of Fig13, and is still able (see Figs 13 and 14) to encode all 20 amino acids that contribute to the formation of proteins.
The 32 triplets are made as follows:
in second position, the entire first column always has U, the second column always has C, the third column always has A, the fourth column always has G.

U	C	A	G
UUC	UCC	UAC	UGC
UUG	UCG	UAG	UGG
CUC	CCC	CAC	CGC
CUG	CCG	CAG	CGG
AUC	ACC	AAC	AGC
AUG	ACG	AAG	AGG
GUC	GCC	GAC	GGC
GUG	GCG	GAG	GGG

Fig14 32-codon code

(What is the meaning of a 32-codon coding instead of 64? Hypothesizing an evolutionary process of modification of the genetic code, several authors consider the current construction as coming from simpler cases, so it may make sense to study the regularities and symmetries in a 32-codon subtable).

Still to think about regularities, a good idea could be this:
I rearrange the rows (Fig15).
The first 4 rows end with C.
The second 4 rows, repeated identically, end with G.
16, in green, end with C.
16 in red, repeated, identical, but instead end with G.

For sub-frames, 4 bits are sufficient.
The fifth bit, the most significant bit, marks the transition from C to G in the third position.
In total, 5 bits are used to uniquely identify the 32 triplets, see Fig15, and they are made as follows.
For the columns, the top row shows the bits that identify the columns.
I use the two least significant bits, 00 for the column with U in the second position, 01 for C in the second position, 10 for A in the second position, 11 for G in the second position.
As for row addresses, the first column on the left shows the row bits. I use the three heaviest bits, starting with the most significant bit, and the rows are identified with 000, 001, 010 etc up to 111. The heaviest bit, as already mentioned, marks the transition from C to G in the third position.
As for the second column with U□C, C□C etc., the symbol "□" indicates the empty space filled in succession by U, C, A, G.

Fig 16 is a summary.
In Fig 17 I use convenient colors yellow blue pink purple to distinguish the columns, and to strengthen the analogy with what happens for the 32 crystal classes. Figs 16 and 17 also define the meaning of the various bits.

Row and column bits	Meaning	U 00	C 01	A 10	G 11
000	U□C	UUC	UCC	UAC	UGC
001	C□C	CUC	CCC	CAC	CGC
010	A□C	AUC	ACC	AAC	AGC
011	G□C	GUC	GCC	GAC	GGC
100	U□G	UUG	UCG	UAG	UGG
101	C□G	CUG	CCG	CAG	CGG
110	A□G	AUG	ACG	AAG	AGG
111	G□G	GUG	GCG	GAG	GGG

Fig15 Reordering, and column and row addresses.

Row and column bits			00	01	10	11
	Meaning (meaning)		Second position U	Second position C	Second position A	Second position G
000	C-ending codons	U-starting codons	UUC	UCC	UAC	UGC
001		C-starting codons	CUC	CCC	CAC	CGC
010		A-starting codons	AUC	ACC	AAC	AGC
011		G-starting codons	GUC	GCC	GAC	GGC
100	G-ending codons	U-starting codons	UUG	UCG	UAG	UGG
101		C-starting codons	CUG	CCG	CAG	CGG
110		A-starting codons	AUG	ACG	AAG	AGG
111		G-starting codons	GUG	GCG	GAG	GGG

Fig16 Summary with row and column addresses, and meanings

Bit			00	01	10	11
	Meaning (meaning)		Second position U	Second position C	Second position A	Second position G
000	C-ending codons	U-starting codons	UUC	UCC	UAC	UGC
001		C-starting codons	CUC	CCC	CAC	CGC
010		A-starting codons	AUC	ACC	AAC	AGC
011		G-starting codons	GUC	GCC	GAC	GGC
100	G-ending codons	U-starting codons	UUG	UCG	UAG	UGG
101		C-starting codons	CUG	CCG	CAG	CGG
110		A-starting codons	AUG	ACG	AAG	AGG
111		G-starting codons	GUG	GCG	GAG	GGG

Fig17 Convenient colors yellow blue pink purple to distinguish the columns

4. Bibliography

[1] Bilbao Server. Generators/General Positions of 3D Crystallographic Point Groups
https://www.cryst.ehu.es/cryst/get_point_genpos.html

[2] Genetic code
https://en.wikipedia.org/wiki/Genetic_code

[3] Codons table
https://www.biopills.net/il-codice-genetico-nella-traduzione/