

# On the geometry of regular icosahedral capsids containing disymmetrons



Kai-Siang Ang<sup>a</sup>, Laura P. Schaposnik<sup>b,\*</sup>

<sup>a</sup> The Harker School, San Jose, CA 95128, USA

<sup>b</sup> University of Illinois, Chicago, IL 60607, USA

## ARTICLE INFO

### Article history:

Received 6 September 2016

Received in revised form 11 January 2017

Accepted 11 January 2017

Available online 19 January 2017

### Keywords:

Icosahedral viruses

Icosahedral symmetry

Trisymmetron

Pentasympmetron

Disymmetron

Capsomer

Virus capsid structure

Triangulation number

## ABSTRACT

Large icosahedral virus capsids are composed of symmetrons, organized arrangements of capsomers. There are three types of symmetrons: disymmetrons, trisymmetrons, and pentasympmetrons, which have different shapes and are centered on the icosahedral 2-fold, 3-fold and 5-fold axes of symmetry, respectively. Sinkovits and Baker (2010) gave a classification of all possible ways of building an icosahedral structure solely from trisymmetrons and pentasympmetrons, which requires the triangulation number  $T$  to be odd. In the present paper we incorporate disymmetrons to obtain a geometric classification of large icosahedral viruses formed by regular penta-, tri-, and disymmetrons, giving all mathematically consistent and theoretically possible solutions. For every class of solutions, we further provide formulas for symmetron sizes and parity restrictions on  $h$ ,  $k$ , and  $T$  numbers. We also present several methods in which invariants may be used to classify a given configuration.

© 2017 Elsevier Inc. All rights reserved.

## 1. Introduction

Viruses are known to have two essential components: the genetic material and the protein *capsid*, which surrounds the DNA or RNA. In 1956 Watson and Crick noticed that because the amount of genomic material in a virus could only encode for a few capsid proteins much smaller than the overall capsid, the genetic material needed instead to code for a few smaller proteins, produced in large numbers and arranged symmetrically (Watson and Crick, 1956).

The 60 symmetries of the icosahedron would then imply that there should be 60 proteins forming the capsid, but this is not what has been observed in most cases. In 1962 Caspar and Klug used the idea of quasi-equivalence to explain how more than 60 proteins could come together (Klug and Caspar, 1962). With 60 proteins, each face of the icosahedron has three proteins, one near each vertex. They proposed that the faces could be further triangulated, with each smaller triangle still having three proteins. The triangulation can be seen on a triangulated sphere as in Fig. 1(A).

The original icosahedral vertices are at the center of the red pentagons. The number of smaller triangles per original icosahedron

face is known as the triangulation number  $T$ . The five or six proteins around a single vertex come together to form units called *capsomers*, which can be modelled as small spheres at the vertices produced by the triangulation. The total number of capsomers is related to the triangulation number by the formula (Wrigley, 1969).

$$N_{cap} = 12 + 10(T - 1). \quad (1)$$

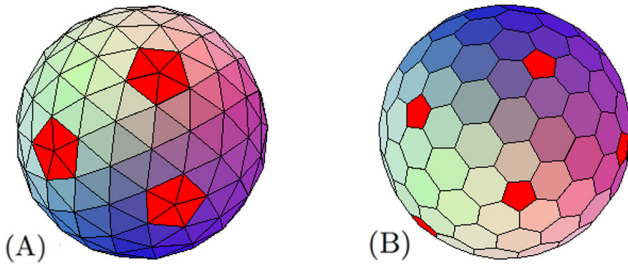
In Fig. 1(B) one can see the dual of the triangulated sphere, in which capsomers are represented by the pentagons or hexagons. Throughout the paper we shall use the triangulated sphere representation, as we can convert the 3D surface into a 2D coordinate system by using any triangulated icosahedral face to tile the 2D plane. Whilst sometimes in the literature readers may find references to crystallography in order to study these lattices, we shall make the present manuscript self-contained and thus mention below the main ideas and mathematical properties, some of which shall be proven in the Appendices for completion. For further reference in the area, the reader should refer to, e.g., Phillips et al. (1956).

The triangulation of the icosahedral face provides then the lines of a lattice of equilateral triangles, and whose axis are denote by  $h$  and  $k$ , as can be seen in Fig. 2(A) below.

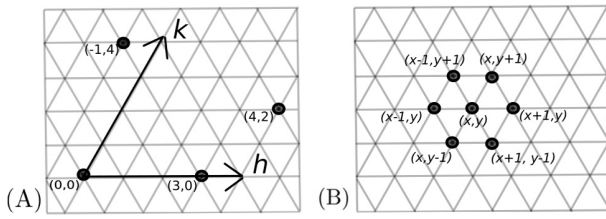
The triangulation number  $T$  can also be expressed in terms of  $h$  and  $k$  through the formula (Klug and Caspar, 1962)

\* Corresponding author.

E-mail address: [schapos@uic.edu](mailto:schapos@uic.edu) (L.P. Schaposnik).



**Fig. 1.** Icosahedral capsid via the dual triangulated sphere, where 5-fold centers in red and  $(h, k) = (1, 3)$ . (A) Triangulated sphere; (B) dual space.



**Fig. 2.** (A) The triangular lattice with  $h$ - and  $k$ -axes; (B) Lattice point  $(x, y)$  and its 6 adjacent points.

$$T = h^2 + hk + k^2, \quad (2)$$

and the  $h, k$  parameters can be seen in terms of Fig. 2 by noting that if one started from one 5-fold axis, proceeded  $h$  units along the triangulated sphere's edges, turned  $\frac{\pi}{3}$  radians counterclockwise, and proceeded  $k$  units, then a neighboring 5-fold vertex would be reached (see Fig. 3).

To reverse the process, one simply takes the triangle formed by three adjacent 6-fold centers of symmetry and map this triangle onto each face of the icosahedron. Since capsomers are located on vertices of the grid, they can be described through lattice points given by coordinates in the  $h$  and  $k$ -axes. In polar coordinates, the  $h$ -axis is oriented along  $\theta = 0$  and the  $k$ -axis is oriented along  $\theta = \frac{\pi}{3}$  (an example is shown in Fig. 2(A) below). Note that pentasymmetrons become hexagons in two dimensions.

In 1969, Wrigley noticed in Wrigley (1969) that large icosahedral virus capsids tend to dissociate into certain organized collec-

tions of capsomers called *symmetrons*. These symmetrons have the shapes of regular pentagons (*pentasymmetrons*), equilateral triangles (*trisymmetrons*), and line segments (*disymmetrons*). Each of these symmetrons is centered on a corresponding axis of symmetry of the icosahedron (e.g., a pentasymmetron must be centered on an icosahedral vertex, through which a 5-fold axis of symmetry passes) (Wrigley, 1969). Additionally, they must be arranged in a way that conforms exactly to the symmetry of the icosahedron. Due to their shapes, symmetrons can be characterized simply by their edge lengths, which we shall denote by  $d, t$ , and  $p$  corresponding to di-, tri-, and pentasymmetrons respectively (these lengths are often denoted by  $e_{DS}, e_{TS}$ , and  $e_{PS}$  in the literature). Since the icosahedral vertices are always occupied by a pentasymmetron capsomer, one necessarily has  $p > 0$ . Moreover, the symmetron shapes also lead to simple formulas for the number of capsomers per symmetron:

$$N_{DS} = d, \quad (3)$$

$$N_{TS} = \frac{t(t+1)}{2}, \quad (4)$$

$$N_{PS} = 1 + \frac{5p(p-1)}{2}. \quad (5)$$

The number of capsomers in each symmetron depends on the edge length, and the symmetries of the icosahedron determine how many of each symmetrons there can be. One should note that icosahedral symmetry does not fully determine the number of symmetrons of each type since there are often multiple solutions that do or do not contain disymmetrons, and this can be seen for instance in Table 3. The above Eqs. (3)–(5) can be combined with Wrigley's formula (1) relating the total number of capsomers  $N_{cap}$  to  $T$  to produce the equation

$$T - 1 = 3p(p-1) + t(t+1) + 3d. \quad (6)$$

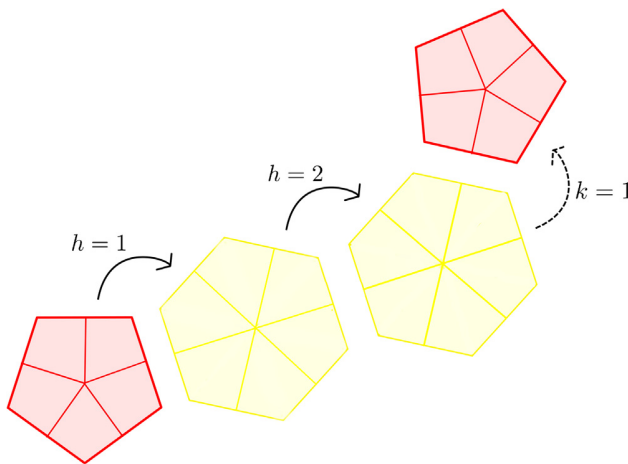
Although not all virus capsids are made of symmetrons, tri- and pentasymmetrons have been found to compose several viruses, examples of which can be found below in Table 1.

In 2010, Sinkovits and Baker gave a complete classification of all ways to arrange penta- and trisymmetrons to form the icosahedral capsid (Sinkovits and Baker, 2010). However, disymmetrons were not considered, and hence as noted by Wrigley, only accounted for viruses with odd  $T$  numbers (Wrigley, 1969). A different approach was taken by Grytczuk (2006), who used number theory to describe a systematic procedure to find all solutions to  $T - 1 = 3p(p-1) + t(t+1) + 3d$ , but he did not account for the physical shape of the symmetrons or the fact that  $T$  may be parametrized as  $h^2 + hk + k^2$ , leading to many solutions not corresponding to physical configurations (Grytczuk, 2006). We dedicate this paper to classify regular symmetron arrangements from a geometric perspective, making full use of all geometric information.

## 2. Towards a classification of icosahedral viruses

In what follows we shall give a description of how icosahedral viruses can be classified by analysing the disymmetrons contained in the capsid, leading to six different classes of viruses. In the absence of disymmetron, i.e. when  $d = 0$ , one recovers the results of Sinkovits and Baker (2010), which we shall refer to as Class 1. In order to classify viruses for which  $d > 0$ , we shall begin with definitions for bordering, edge, and interior capsomers.

**Definition 1.** A lattice point  $(x, y)$  is **adjacent** to another lattice point, if it is connected to it by exactly one edge in the lattice.



**Fig. 3.** Triangulation number  $T = 7$  with  $h = 2$  and  $k = 1$ . The lines inside the pentagons and hexagons are from the triangulated sphere, while the polygons themselves are from the dual representation.

**Table 1**

Known viruses, their  $(h, k)$  characterizations, their  $T$  numbers, their symmetrons sizes  $(d, t, p)$ , and the method (Met.) of analysis (Klose et al., 2016; Sinkovits and Baker, 2010). Abbreviations: NS, negatively stained; IIV, Invertebrate iridescent virus; TIV, *Tipula* iridescent virus; SIV, *Sericesthis* iridescent virus; CIV, *Chilo* iridescent virus; FV3, Frog virus 3; PBCV-1, *Paramecium bursaria chlorella* virus; PpV01, *Phaeocystis pouchetii* virus. The quality of Wrigley's images allowed for several possible interpretations. For frog virus 3 (FV3), results are by Yan et al. in unpublished results, (see Sinkovits and Baker, 2010).

Virus	$(h, k)$	$T$	$(d, t, p)$	Met.	Ref.
SIV (IIV-2)*	(5, 8)	129	(0, 10, 3)	NS	Wrigley (1969)
SIV (IIV-2)*	(4, 10)	156	(9, 10, 3)	NS	Wrigley, 1969
SIV (IIV-2)*	(7, 7)	147	(0, 10, 4)	NS	Wrigley (1969)
TIV (IIV-1)	(7, 7)	147	(0, 10, 4)	NS	Manyakov (1977)
CIV (IIV-6)	(7, 7)	147	(0, 10, 4)	Cryo	Yan et al. (2009)
FV3	(7, 8)	169	(0, 11, 4)	Cryo	(☆)
PBCV-1	(7, 8)	169	(0, 11, 4)	Cryo	Yan et al. (2005)
PpV01	(7, 10)	219	(0, 13, 4)	Cryo	Yan et al. (2000)
Faustovirus	(7, 12)	277	(0, 15, 4)	Cryo	Klose et al. (2016)

In the lattice representation a di-, tri-, and penta- symmetron is a collection of lattice points which form a line, a triangle or pentagon. From the above definition, the point  $(x, y)$  is adjacent to  $(x + 1, y)$ ,  $(x, y + 1)$ ,  $(x - 1, y + 1)$ ,  $(x - 1, y)$ ,  $(x, y - 1)$ , and  $(x + 1, y - 1)$  as seen above in Fig. 2(B).

**Definition 2.** In a symmetron  $S$ , a capsomer is an **interior capsomer** if all adjacent lattice points are occupied by capsomers of  $S$ . Otherwise, it is an **edge capsomer**.

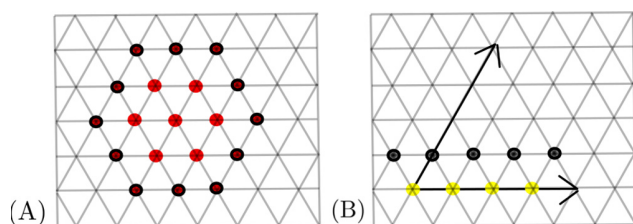
**Definition 3.** Set the grid so that a disymmetron  $D$  lies along the line  $k = 0$ . We say that any  $k = \pm 1$  capsomer adjacent to a capsomer of  $D$  is a **bordering capsomer** with respect to  $D$ .

**Definition 4.** A symmetron is a **bordering symmetron** that **borders** the disymmetron  $D$  if either:

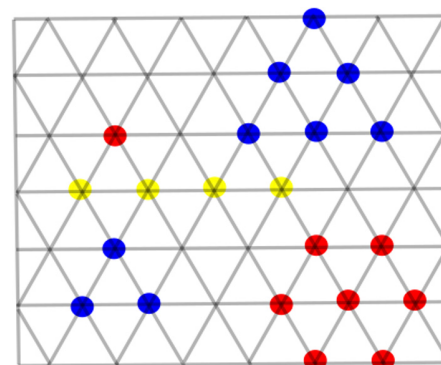
- it has at least two capsomers with the same  $k$  coordinate which are bordering capsomers with respect to  $D$ .
- or, all of its capsomers are bordering capsomers with respect to  $D$ .

The  $k$  coordinate of the bordering capsomers determines which **side** of  $D$  the symmetron is bordering. (Figs. 4 and 5)

As can be seen from the above definitions, bordering always occurs with respect to a disymmetron, so for simplicity we shall write *pentasymmetrons border* when referring to *pentasymmetrons bordering disymmetrons*. Because of the convex shape of penta- and trisymmetrons, a bordering penta- or trisymmetron will have a single string of edge capsomers bordering exactly one side of the disymmetron. The symmetron shapes also dictate that if a symmetron  $S$  has bordering capsomers with  $k = 1$ , then all of the capsomers of  $S$  have  $k \geq 1$ . Moreover, 2-fold symmetry about the disymmetron dictates that another symmetron  $S'$  must be bordering the disymmetron on the opposite side. From the above, on each side of the disymmetron there can be at most one bordering di-, at most one bordering tri-, and at most one bordering pentasymmetron.



**Fig. 4.** (A) An example of a pentasymmetron with edge capsomers (black) distinguished from interior capsomers (red); (B) An example of a disymmetron of length 4 (yellow) and its 5 (black)  $k = 1$  bordering capsomers.



**Fig. 5.** A disymmetron (yellow) and above it, its bordering pentasymmetron (red) and trisymmetron (blue). The penta- and trisymmetron below the disymmetron are not bordering, since neither of the items in the definition apply.

By a *symmetry of overlapping argument* (through which one applies icosahedral symmetry to find an overlap of symmetrons, giving a proof by contradiction), a disymmetron centered on an edge  $e$  can only be adjacent to the four disymmetrons centered on edges belonging to icosahedral faces with edge  $e$ , the two trisymmetrons centered on icosahedral faces with edge  $e$ , and the two pentasymmetrons centered on the endpoints of edge  $e$ . Note that two disymmetrons bordering each other must be parallel, but if they are centered on two edges of a single icosahedral face  $F$ , then this is not possible if  $d > 1$  (3-fold symmetry about the center of  $F$  dictates that they must be at an angle of  $\frac{\pi}{3}$  to each other). Therefore disymmetrons can not border each other unless  $d = 1$ . In A, we prove that bordering symmetrons must exist if disymmetrons exist, i.e., when  $d > 0$ .

In what follows we shall use geometric methods to give a classification of icosahedral viruses based upon what types of symmetrons border the disymmetrons leading to six distinct classes. When only pentasymmetrons border, one has Classes 2 and 5; when only trisymmetrons border, one has Classes 3 and 6; when both tri- and pentasymmetrons border, one has Class 4. As stated before, when  $d = 0$  gives Class 1. We shall now show how these Classes arise and prove that they are the only possible configurations. Unless stated otherwise, we shall set the coordinate grid so that, if the disymmetron exists, it occupies the lattice points from  $(1, 0)$  to  $(d, 0)$ . We shall also be applying 2-, 3-, or 6-fold symmetries, implicitly referring to the formulas given in B.

### 2.1. Class 1: No disymmetrons, $d = 0$

Sinkovits and Baker showed that there are three configurations for  $d = 0$ , and we refer the reader to Sinkovits and Baker (2010) for

the proof that these are all that exist. These authors considered three different configurations because several different formulas arise (we will derive them in Section 3) and the distinction was necessary to conform to their restriction that  $h \leq k$ .

Along this paper, we drop the restriction  $h \leq k$ , and thus treat all three configurations of Sinkovits and Baker (2010) as the same. Note that handedness can be accounted for with a reflection of the plane, or a switch between  $h$  and  $k$  coordinates. What these configurations have in common is that, after a reflection of the plane if needed, one can set a coordinate grid so that:

- There is a pentasymmetron  $P_1$  with center at  $(0, 0)$ .
- $P_1$  has an edge from  $(1 - p, p - 1)$  to  $(0, p - 1)$ .
- There is a trisymmetron edge from  $(1 - p, p)$  to  $(t - p, p)$ .
- There is a pentasymmetron  $P_2$  with an edge from  $(t - p + 1, p)$  to  $(t, p)$ .
- $P_2$  has its center at  $(t - p + 1, 2p - 1)$ .

The above information, which we shall refer to as the Grid Description (GD), combined with the icosahedral symmetries, is sufficient to fill the plane. Examples can be seen below in Fig. 6.

In what follows we shall describe the five classes for which  $d > 0$  by giving its GD and showing why this description provides sufficient information to determine the class of a virus.

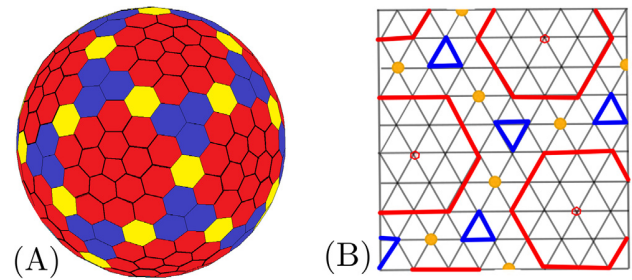
## 2.2. Class 2: Only pentasymmetrons border I

This Class has the following Grid Description:

- There is a pentasymmetron  $P_1$  with center at  $(0, 0)$ .
- $P_1$  has an edge from  $(1 - p, p - 1)$  to  $(0, p - 1)$ .
- There is a disymmetron from  $(1, p - 1)$  to  $(d, p - d)$ .
- There is a disymmetron from  $(1 - p, p)$  to  $(d - p, p)$ .
- There is a trisymmetron from  $(d - p + 1, p)$  to  $(0, p)$ .
- There is a pentasymmetron  $P_2$  with an edge from  $(d + 1, p - d)$  to  $(d + p, p - d)$ .
- $P_2$  has its center at

$$(d + 1, 2p - d - 1) = (p - t + 1, p + t - 1).$$

Examples can be seen in Fig. 7. In order to see that the above GD determines the class, consider a pentasymmetron that has an edge with  $k = c$  for some constant  $c$ , and let all of its capsomers have  $k \geq c$ . Considering the 6-fold symmetry around the pentasymmetron, one can see that there will be overlapping capsomers if all of the  $k = c - 1$  lattice points adjacent to the  $k = c$  edge capsomers are occupied by a single disymmetron. Now recall that we have set the grid so that there exists a disymmetron from



**Fig. 7.** Examples of viruses in Class 2 where only pentasymmetrons border: (A) corresponding to  $d = 1, t = 2$ , and  $p = 3$  and  $(h, k) = (2, 4)$ ; (B) corresponding to  $d = 2, t = 2$ , and  $p = 4$  and  $(h, k) = (2, 4)$ .

$(1, 0)$  to  $(d, 0)$ . Without loss of generality, we may consider the bordering pentasymmetron that has  $k = 1$  capsomers from  $(m, 1)$  to  $(n, 1)$ , with  $m \leq 0 \leq n \leq d - 1$ . Then the pentasymmetron center is at  $(m, 1 + n - m)$ .

Notice that  $(1, 0)$  gets mapped to  $(n + 1, 1 - m)$  under a  $\frac{\pi}{3}$  rotation about the pentasymmetron center. The points from  $(n + 1, 1)$  to  $(n + 1, -m)$  can not be pentasymmetron capsomers by C and can not be disymmetrons by 6-fold symmetry, so they must be trisymmetron capsomers. Each of the points are of the form  $(n + 1, p)$ , with  $1 \leq p \leq -m$ . By a  $-\frac{\pi}{3}$  rotation about the pentasymmetron center, they are mapped to  $(m + p, 0)$ .

If there are zero such  $(n + 1, p)$  points, then  $m = 0$ , and we can apply 2-fold and 6-fold symmetries to partially fill the grid. Note that if  $n < d - 2$ , then  $(n + 2, 1)$  can not be a penta- or disymmetron capsomer, so it must be a trisymmetron capsomer. But this would make both penta- and trisymmetrons border the disymmetron, which lies outside of this case. The case of  $n = d - 1$  fixes  $t = 0$ , and there is enough information to see that this arrangement fits the GD and fills the plane. The case of  $n = d - 2$  also fixes  $t = 0$ , and there is enough information to fill the plane. However, this arrangement does not conform to the GD and we shall see in Section 2.4 that it gives a degenerate form of Class 4.

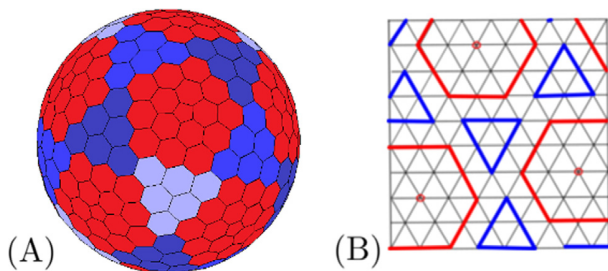
We shall now assume that there is at least one point of the form  $(n + 1, p)$ . These points belong to the same trisymmetron. If there are at least two such points, the trisymmetron orientation becomes fixed. If there is only one such point, the same orientation can occur, but an alternate orientation becomes possible and is covered in Section 2.5, not here. By 6-fold symmetry, there is another trisymmetron with edge along  $(m + p, 0)$ . By 2-fold symmetry about the  $k = 0$  disymmetron, the trisymmetron with edge  $(n + 1, p)$  must also have an edge with  $k = 0$ , fixing the size so that  $d + t = p$ . This is enough information to cover the plane and verify that this conforms to the GD.

## 2.3. Class 3: Only trisymmetrons border I

This Class has the following Grid Description:

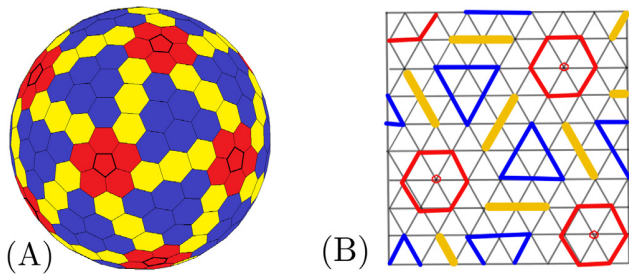
- There is a pentasymmetron  $P_1$  with center at  $(0, 0)$ .
- $P_1$  has an edge from  $(1 - p, p - 1)$  to  $(0, p - 1)$ .
- There is a disymmetron from  $(1, p - 1)$  to  $(d, p - 1)$ .
- There is a trisymmetron from  $(2 - p, p)$  to  $(d - 1, p)$ .
- There is a pentasymmetron  $P_2$  with an edge from  $(d + 1, p - 1)$  to  $(d + p, p - 1)$ .
- $P_2$  has its center at  $(d + 1, 2p - 2) = (t - p + 3, 2p - 2)$ .

Examples of configurations in this class can be seen in Fig. 8. We now show how this configuration arises. As in the previous section, without loss of generality, we consider the trisymmetron having  $k = 1$  capsomers from  $(m, 1)$  to  $(n, 1)$ , with



**Fig. 6.** Examples of Class 1 viruses with no disymmetrons: (A) corresponding to  $d = 0, t = 3$ , and  $p = 3$  and  $(h, k) = (5, 1)$ . Different shades of blue are used to distinguish different trisymmetrons; (B) corresponding to  $d = 0, t = 2$ , and  $p = 2$  and  $(h, k) = (5, 1)$ .





**Fig. 8.** Examples of viruses in Class 3 where only trisymmetrons border: (A) corresponding to  $d = 3, t = 3, p = 2$ , and  $(h, k) = (2, 4)$ ; (B) corresponding to  $d = 3, t = 5, p = 4$ , and  $(h, k) = (2, 4)$ .

$$m \leq 0 \leq n. \quad (7)$$

The third vertex of the trisymmetron is at  $(m, 1 + n - m)$ , which makes the center  $(\frac{2m+n}{3}, \frac{3+n-m}{3})$ .

By **B**, under a  $\frac{2\pi}{3}$  rotation,  $(1, 0)$  goes to  $(m + n, 2 - m)$ . Since these points can not be tri- or disymmetron capsomers by **C** and 6-fold symmetry, the points from  $(n + 1, 1)$  to  $(m + n + 1, 1 - m)$ , which we shall call the  $e$  points, must be pentasymmetron capsomers of the same pentasymmetron by **C**.

Since  $p > 0$ , there will always be at least one  $e$  point. Note that if there are at least two  $e$  points, then they form an edge that defines the orientation of the pentasymmetron. This orientation may occur if there is only one  $e$  point, but there is also an alternate orientation as we discuss in Section 2.6. Rotating the edge between  $(n + 1, 1)$  and  $(m + n + 1, 1 - m)$  by  $\frac{2\pi}{3}$  clockwise, we see that a different pentasymmetron has a  $k = 0$  edge. By 2-fold symmetry around the  $k = 0$  disymmetron, the pentasymmetron containing  $(n + 1, 1)$  must also have a  $k = 0$  edge, fixing the pentasymmetron edge size:  $d + p = t + 2$ . This is enough information to fill the plane and thus determines the class.

#### 2.4. Class 4: Both tri- and pentasymmetrons border

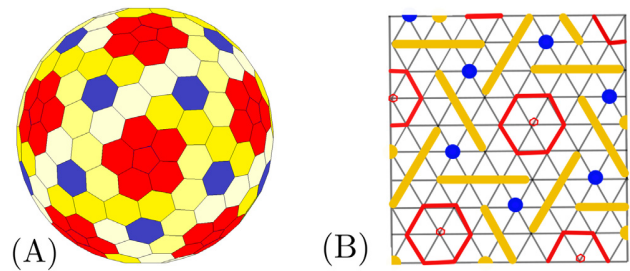
This Class has the following Grid Description:

- There is a pentasymmetron  $P_1$  with center at  $(0, 0)$ .
- $P_1$  has an edge from  $(1 - p, p - 1)$  to  $(0, p - 1)$ .
- There is a disymmetron from  $(1 - p, p)$  to  $(d - p, p)$ .
- There is a trisymmetron edge from  $(1 - p, p + 1)$  to  $(t - p, p + 1)$ .
- There is a pentasymmetron  $P_2$  with an edge from  $(d - 2p + 1, p + 1)$  to  $(d - p, p + 1)$ .
- $P_2$  has its center at  $(d - 2p + 1, 2p) = (t - p + 2, 2p)$ .

Examples can be seen in Fig. 9.

In this case, since both the penta- and trisymmetrons border the disymmetron, they both have no  $k = 0$  capsomers, so  $(0, 0)$  is a disymmetron capsomer. Suppose that a bordering trisymmetron is adjacent to  $(0, 0)$ . It can not also be adjacent to  $(d + 1, 0)$ , since otherwise the pentasymmetron would not be able to border. Applying 3-fold symmetry, we see that there will be a disymmetron capsomer with  $h < 0$  and  $k = 0$ , and this capsomer will be part of a disymmetron parallel to the  $k$ -axis. Thus the disymmetron containing  $(0, 0)$  must be parallel to  $h + k = 0$ . However, it will not be bordering either of the trisymmetrons bordering the  $k = 0$  disymmetron. Hence, the bordering trisymmetron cannot be adjacent to any non-disymmetron  $k = 0$  capsomers, and each edge of the trisymmetron is completely bordering a disymmetron.

Consider now all of the  $k = 1$  bordering capsomers. Since one penta- and one trisymmetron border a given side of the disymmetron, there will be exactly one string of pentasymmetron cap-



**Fig. 9.** Examples of viruses in Class 4 where both tri- and pentasymmetrons border: (A) corresponding to  $d = 4, t = 1, p = 2$ , and  $(h, k) = (1, 4)$ , where different shades of yellow are used to distinguish different disymmetrons; (B) corresponding to  $d = 5, t = 2, p = 2$ , and  $(h, k) = (1, 4)$ .

somers and one string of trisymmetron capsomers. A total of four disymmetrons may be adjacent to but not bordering the  $k = 0$  disymmetron, and two of these disymmetrons may each have at most one adjacent  $k = 1$  capsomer. Because each edge of the trisymmetron completely borders a disymmetron, on either side of the string of the trisymmetron  $k = 1$  bordering capsomers, there will be a disymmetron capsomer. The only other component of the  $k = 1$  bordering capsomers is the string of pentasymmetron bordering capsomers, which is immediately before or after the string of di-tri-disymmetron capsomers.

In view of the analysis above and without loss of generality, we may assume that a trisymmetron edge goes from  $(d - 1, 1)$  to  $(d - t, 1)$ . This means that the third trisymmetron vertex is at  $(d - t, t)$ , so the trisymmetron center is located at  $(\frac{3d-2t-1}{3}, \frac{t+2}{3})$ . 3-fold symmetry means a disymmetron must exist from  $(d - t - 1, 1)$  to  $(d - t - 1, d)$ . By 2-fold symmetry about the  $k = 0$  disymmetron, there must be a trisymmetron with edge from  $(2, -1)$  to  $(1 + t, -1)$  and third vertex at  $(1 + t, -t)$ . By 3-fold symmetry about this new trisymmetron, there is a disymmetron from  $(1 + t, -1 - t)$  to  $(2 + t - d, -2 - t + d)$ .

Consider the  $k = 1$  capsomers adjacent to the  $k = 0$  disymmetron. For a pentasymmetron to border the disymmetron,  $d \geq t + 3$ . Therefore the disymmetron from  $(1 + t, -1 - t)$  to  $(2 + t - d, -2 - t + d)$  passes through  $(-1, 1)$ . Consider the lattice points from  $(0, 1)$  to  $(d - t - 2, 1)$ . These must belong to a single pentasymmetron, since no two symmetrons of the same type may be bordering the same side of a disymmetron. This fixes the pentasymmetron orientation and gives  $d = t + p + 1$ , thus completing the plane.

#### 2.5. Class 5: only pentasymmetrons border II

We shall consider here an exceptional case appearing when only pentasymmetrons border, which lies outside the realm of Class 2. This Class has the following GD:

- There is a pentasymmetron  $P_1$  with center at  $(0, 0)$ .
- $P_1$  has an edge from  $(1 - p, p - 1)$  to  $(0, p - 1)$ .
- There is a disymmetron from  $(0, p)$  to  $(p - 2, 2)$ .
- There is a trisymmetron edge from  $(-1, p)$  to  $(-1, p + 1)$ .
- There is a pentasymmetron  $P_2$  with an edge from  $(p - 2, 3)$  to  $(2p - 3, 3)$ .
- $P_2$  has its center at  $(p - 2, p + 2)$ .

Examples can be seen in Fig. 10.

We now show how this configuration arises. In this case only pentasymmetrons border and there is only one  $(n + 1, p)$  point. In addition to the orientation covered in Section 2.2, there is another orientation in which the trisymmetron has an edge along  $h = 1$ .

Because there is another trisymmetron with a  $k = 0$  vertex capsomer, 2-fold symmetry about the  $k = 0$  disymmetron forces the trisymmetron containing  $(n + 1, 1)$  to have a vertex capsomer with  $k = 0$ . This fixes  $t = 2$ , and the point  $(n + 1, 0)$  can now only be a disymmetron, fixing the disymmetron size so that we have  $p = d + 1$ , and therefore the plane can be canonically filled.

## 2.6. Class 6: only trisymmetrons border II

We shall consider here an exceptional case appearing when only trisymmetrons border, which lies outside the realm of Class 3. This Class has the following GD:

- There is a pentasymmetron  $P_1$  with center at  $(0, 0)$ .
- $P_1$  has an edge from  $(-1, 1)$  to  $(0, 1)$ .
- There is a disymmetron from  $(0, 2)$  to  $(0, d + 1)$ .
- There is a trisymmetron edge from  $(-1, 2)$  to  $(-1, d + 2)$ .
- There is a pentasymmetron  $P_2$  with an edge from  $(0, d + 2)$  to  $(1, d + 2)$ .
- $P_2$  has its center at  $(0, d + 3) = (0, t + 2)$ .

Examples can be seen in Fig. 11.

This is the case in which only trisymmetrons border and there is only one pentasymmetron capsomer among the  $e$  points mentioned in Section 2.3, implying  $m = 0$  in (7). In addition to the orientation covered in that section, there is another orientation in which the pentasymmetron has an edge with  $k = 1$ . The disymmetron containing  $(n, 2)$  has a pentasymmetron vertex capsomer at one end, at  $(n + 1, 1)$ . Therefore, the  $k = 0$  disymmetron must have pentasymmetron vertex capsomers at its ends also. Note that these ends also have  $k = 0$ . If the pentasymmetron has a  $k = 1$  edge, then the only way for it to have a  $k = 0$  vertex capsomer is to have one at  $(n + 1, 0)$ . This gives  $p = 2$  and  $d + 1 = t$ , and the plane can be filled.

## 3. Formulas and invariants

As in Sinkovits and Baker (2010), we now derive formulas for  $d, t$ , and  $p$  based on  $h$  and  $k$ , and we also provide parity restrictions on  $h$  and  $k$ . In our Grid Descriptions, we examined the position of the center of adjacent pentasymmetrons with respect to a pentasymmetron centered at  $(0, 0)$ , which leads to the following table:

Note that the  $(h, k)$  characterization of the icosahedral triangulation requires that  $h, k \geq 0$ , so the coordinates in Table 2 are not exactly  $(h, k)$ , and the signs of these coordinates are relevant. Since  $p > 0$  and (with the exception of Class 1)  $d > 0$ , all  $k$  coordinates are non-negative. Also, for Classes 2, 3, 5, and 6, we can use the extra equations to see that also  $h \geq 0$ .

If the coordinate of the adjacent pentasymmetron center has  $0 \leq \theta \leq \frac{\pi}{3}$ , then the  $h$  steps taken in the  $(h, k)$  characterisation will

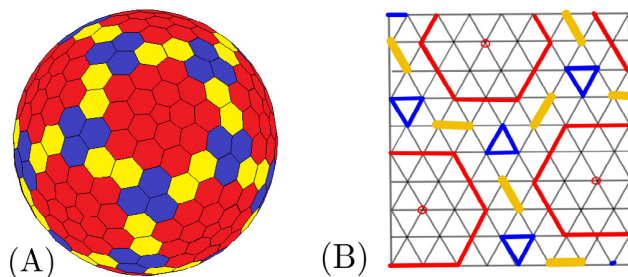


Fig. 10. Examples of viruses in Class 5 giving exceptional case of when only pentasymmetrons border: (A) corresponding to  $d = 2, t = 2, p = 3$ , and  $(h, k) = (5, 1)$ ; (B) corresponding to  $d = 3, t = 2, p = 4$ , and  $(h, k) = (5, 1)$ .

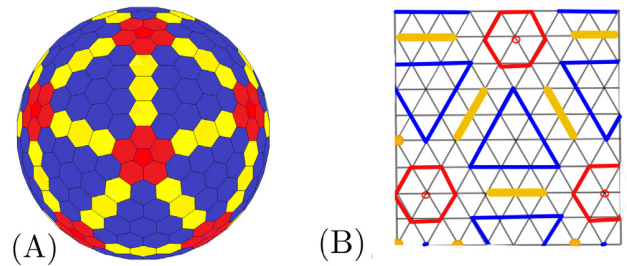


Fig. 11. Examples of viruses in Class 6 giving exceptional case of when only trisymmetrons border: (A) corresponding to  $d = 3, t = 4, p = 2$ , and  $(h, k) = (0, 6)$ ; (B) corresponding to  $d = 3, t = 4, p = 2$ , and  $(h, k) = (0, 6)$ .

Table 2

Coordinates of adjacent pentasymmetron centers with respect to a pentasymmetron at the origin. Note that the  $(h, k)$  characterization of the icosahedral triangulation requires that  $h, k \geq 0$ , so the coordinates above are not exactly  $(h, k)$ .

Class	Coordinates	Extra equations
1	$(t - p + 1, 2p - 1)$	$d = 0$
2	$(p - t + 1, p + t - 1)$	$p = d + t$
3	$(t - p + 3, 2p - 2)$	$t + 2 = p + d$
4	$(t - p + 2, 2p)$	$t + p + 1 = d$
5	$(p - 2, p + 2)$	$t = 2, d + 1 = p$
6	$(0, t + 2)$	$p = 2, d + 1 = t$

be in the direction of  $\theta = 0$  (Case A). If the coordinate has  $\frac{\pi}{3} \leq \theta \leq \frac{2\pi}{3}$ , then the  $h$  steps are in the direction of  $\theta = \frac{\pi}{3}$  (Case B). If the coordinate has  $\frac{2\pi}{3} \leq \theta \leq \pi$ , then the  $h$  steps are in the direction of  $\theta = \frac{2\pi}{3}$  (Case C). Since the  $k$  coordinates of our adjacent pentasymmetron centers are always non-negative, one does not need to consider  $\theta > \pi$ . Finally, when  $h \geq 0$ , one has Case A. If we do not have these restrictions on  $h$ , we may also have Cases B or C.

We can find formulas of the  $(h, k)$  characterisation for the Case A solely based on the coordinates of the adjacent pentasymmetron center. In Case B, the coordinates of the adjacent pentasymmetron center must be rotated clockwise by  $\frac{\pi}{3}$ , i.e.  $(x, y) \rightarrow (x + y, -x)$ . In Case C, the coordinates must be rotated clockwise by  $\frac{2\pi}{3}$ , given by  $(x, y) \rightarrow (y, -x - y)$ . Using these transformations and the coordinates listed in Table 2, we can solve for  $d, t$ , and  $p$  based on  $(h, k)$ . Additionally, we need to reverse  $h$  and  $k$  in each equation to find solutions in a flipped orientation. One should also note that  $d, t$ , and  $p$  must be integral, giving parity restrictions on  $h$  and  $k$ , which in turn give parity restrictions on  $T$  (since  $T = h^2 + hk + k^2$ ), leading to Table 3. For concision, when referring to parity restrictions on  $h, k$ , and  $T$ , we denote equivalence modulo 2 with “ $\equiv$ ”. When specifying equality, we write “ $=$ ”.

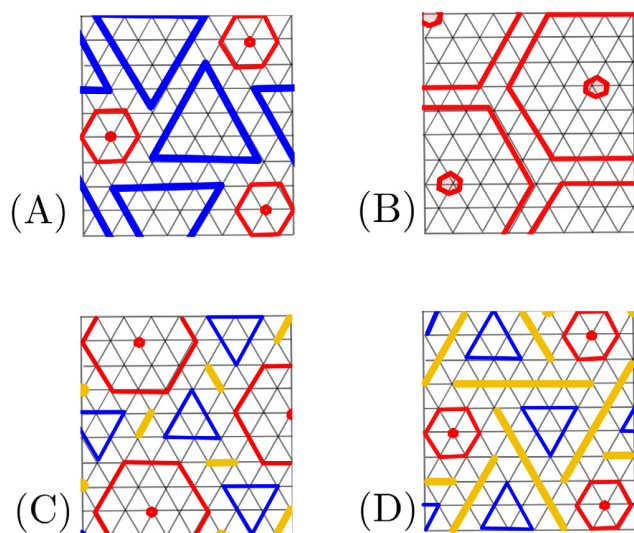
Using the above formulas, one may compute all possible solutions of  $(d, t, p)$  for some particular values of  $(h, k)$ . As an example, for  $(h, k) = (3, 4)$  there are only four possible type of configurations, shown in Fig. 12. Further examples of different solutions for low values of  $(h, k)$  can be seen in Table 4 below.

The following are different methods through which one can distinguish the class of an icosahedral virus:

- Check whether a configuration conforms to each GD.
- Do a visual classification: look for the existence of disymmetrons and, for each pair of types of symmetrons, look at whether these symmetrons' closest edges are parallel or not. This is slightly error-prone.
- Substitute the specific values for  $h$  and  $k$  into Table 3 and see which Class yields the correct values of  $d, t$ , and  $p$ .
- Use the numerical test we describe in D.

**Table 3**Formulas for  $d, t$ , and  $p$  in terms of  $h$  and  $k$ , as well as restrictions on  $h, k$ , and  $T$ .

Class	$d$	$t$	$p$	$h, k$ Parity	$T$ Parity
1	0	$\frac{h-1}{2} + k$	$\frac{h+1}{2}$	$h \equiv 1$	$T \equiv 1$
1	0	$\frac{k-1}{2} + h$	$\frac{k+1}{2}$	$k \equiv 1$	$T \equiv 1$
1	0	$\frac{k-h-1}{2}$	$\frac{h+k+1}{2}$	$h \equiv k$	$T \equiv 1$
1	0	$\frac{h-k-1}{2}$	$\frac{h+k+1}{2}$	$h \equiv k$	$T \equiv 1$
2	$h-1$	$\frac{k-h}{2} + 1$	$\frac{h+k}{2}$	$h \equiv k$	$T \equiv h$
2	$k-1$	$\frac{h-k}{2} + 1$	$\frac{h+k}{2}$	$h \equiv k$	$T \equiv h$
3	$h-1$	$h + \frac{k}{2} - 2$	$\frac{k}{2} + 1$	$k \equiv 0$	$T \equiv h$
3	$k-1$	$\frac{h}{2} + k - 2$	$\frac{h}{2} + 1$	$h \equiv 0$	$T \equiv k$
4	$h+k-1$	$h + \frac{k}{2} - 2$	$\frac{k}{2}$	$k \equiv 0$	$T \equiv h$
4	$h+k-1$	$\frac{h}{2} + k - 2$	$\frac{h}{2}$	$h \equiv 0$	$T \equiv k$
4	$k-1$	$\frac{k-h}{2} - 2$	$\frac{h+k}{2}$	$h \equiv k$	$T \equiv h$
4	$h-1$	$\frac{h-k}{2} - 2$	$\frac{h+k}{2}$	$h \equiv k$	$T \equiv h$
5	$h+1$	2	$h+2$	$k = h+4$	$T \equiv h$
5	$k+1$	2	$k+2$	$h = k+4$	$T \equiv h$
6	$k-3$	$k-2$	2	$h = 0$	$T \equiv k$
6	$h-3$	$h-2$	2	$k = 0$	$T \equiv h$

**Fig. 12.** All possible configurations for  $(h, k) = (3, 4)$ : (A) with  $(d, t, p) = (0, 5, 2)$  in Class 1; (B) with  $(d, t, p) = (0, 0, 4)$  in Class 1; (C) with  $(d, t, p) = (2, 3, 3)$  in Class 3; and (D) with  $(d, t, p) = (6, 3, 2)$  in Class 4.

#### 4. Discussion

The authors in Sinkovits and Baker (2010) asked the question of what combinations of  $d, t$ , and  $p$  lead to valid configurations of symmetrons. In this paper we give an answer to their question, and classify the configurations of admissible icosahedral viruses. In particular, Table coords lists extra equations which must be fulfilled, and it is necessary that  $d > 0$  in non-Class 1 configurations,  $t \geq 0$  and  $p > 0$  in all configurations. For example, we can see that any combination of  $t$  and  $p$  will be a valid Class 1 configuration, which is the same conclusion as the one given in Sinkovits and Baker (2010) but avoids the calculation of limits they had to perform.

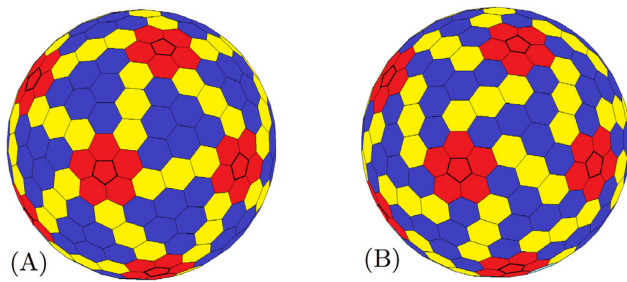
More interestingly, we can ask which solutions could be found when given  $h$  and  $k$  parameters. They could be multiple solutions for  $d, t$ , and  $p$  and are subject to the aforementioned restrictions as well as the parity conditions on  $h$  and  $k$  (or the special conditions on  $h$  and  $k$  for Classes 5 and 6). Thus if  $(h, k) \equiv (1, 1) \pmod{2}$ , we may have solutions from Classes 1, 2, and 4. If  $(h, k) \equiv (0, 0) \pmod{2}$ , we may have solutions from Classes 2, 3, and

**Table 4**For some (not all) values of  $h$  and  $k$  with  $100 \leq T \leq 300$ , all possible solutions of  $d, t$ , and  $p$  and the Classes they belong to.

$(h, k)$	Class: $(d, t, p)$
(0,10)	3: (9,8,1), 4: (9,3,5), 6: (7,8,2)
(0,11)	1: (0,5,6), 3: (10,9,1), 6: (8,9,2)
(0,12)	3: (11,10,1), 4: (11,4,6), 6: (9,10,2)
(1,10)	1: (0,10,1), 1: (0,4,6), 4: (10,4,5)
(1,11)	1: (0,11,1), 1: (0,6,6), 4: (10,3,6)
(1,12)	1: (0,12,1), 1: (0,5,7), 4: (12,5,6)
(2,9)	1: (0,6,5), 1: (0,3,6), 3: (8,8,2), 4: (10,8,1)
(2,10)	2: (1,5,6), 3: (9,9,2), 4: (11,5,5), 4: (9,2,6), 4: (11,9,1)
(2,11)	1: (0,7,6), 1: (0,4,7), 3: (10,10,2), 4: (12,10,1)
(2,12)	2: (1,6,7), 3: (11,11,2), 4: (13,6,6), 4: (11,3,7), 4: (13,11,1)
(3,9)	1: (0,10,2), 1: (0,7,5), 2: (2,4,6), 4: (8,1,6)
(3,10)	1: (0,11,2), 1: (0,3,7), 3: (2,6,6), 4: (12,6,5)
(3,11)	1: (0,12,2), 1: (0,8,6), 2: (2,5,7), 4: (10,2,7)
(3,12)	1: (0,13,2), 1: (0,4,8), 3: (2,7,7), 4: (14,7,6)
(4,8)	2: (3,3,6), 3: (3,6,5), 3: (7,8,3), 4: (11,6,4), 4: (7,0,6), 4: (11,8,2), 5: (5,2,6)
(4,9)	1: (0,8,5), 1: (0,2,7), 3: (8,9,3), 4: (12,9,2)
(4,10)	2: (3,4,7), 3: (3,7,6), 3: (9,10,3), 4: (13,7,5), 4: (9,1,7), 4: (13,10,2)
(4,11)	1: (0,9,6), 1: (0,3,8), 3: (10,11,3), 4: (14,11,2)
(4,12)	2: (3,5,8), 3: (3,8,7), 3: (11,12,3), 4: (15,8,6), 4: (11,2,8), 4: (15,12,2)
(5,7)	1: (0,9,3), 1: (0,8,4), 2: (4,2,6), 2: (6,0,6)
(5,8)	1: (0,10,3), 1: (0,1,7), 3: (4,7,5), 4: (12,7,4)
(5,9)	1: (0,11,3), 1: (0,9,5), 2: (4,3,7), 4: (8,0,7), 5: (6,2,7)
(5,10)	1: (0,12,3), 1: (0,2,8), 3: (4,8,6), 4: (14,8,5)
(5,11)	1: (0,13,3), 1: (0,10,6), 2: (4,4,8), 4: (10,1,8)
(5,12)	1: (0,14,3), 1: (0,3,9), 3: (4,9,7), 4: (16,9,6)
(6,6)	2: (5,1,6), 3: (5,7,4), 4: (11,7,3)
(6,7)	1: (0,9,4), 1: (0,0,7), 3: (6,8,4), 4: (12,8,3)
(6,8)	2: (5,2,7), 2: (7,0,7), 3: (5,8,5), 3: (7,9,4), 4: (13,8,4), 4: (13,9,3)
(6,9)	1: (0,10,5), 1: (0,1,8), 3: (8,10,4), 4: (14,10,3)
(6,10)	2: (5,3,8), 3: (5,9,6), 3: (9,11,4), 4: (15,9,5), 4: (9,0,8), 4: (15,11,3), 5: (7,2,8)
(6,11)	1: (0,11,6), 1: (0,2,9), 3: (10,12,4), 4: (16,12,3)
(6,12)	2: (5,4,9), 3: (5,10,7), 3: (11,13,4), 4: (17,10,6), 4: (11,1,9), 4: (17,13,3)
(7,7)	1: (0,10,4), 2: (6,1,7)
(7,8)	1: (0,11,4), 1: (0,0,8), 3: (6,9,5), 4: (14,9,4)
(7,9)	1: (0,12,4), 1: (0,11,5), 2: (6,2,8), 2: (8,0,8)
(7,10)	1: (0,13,4), 1: (0,1,9), 3: (6,10,6), 4: (16,10,5)
(7,11)	1: (0,14,4), 1: (0,12,6), 2: (6,3,9), 4: (10,0,9), 5: (8,2,9)
(7,12)	1: (0,15,4), 1: (0,2,10), 3: (6,11,7), 4: (18,11,6)

4 (and possibly 5 and 6). If  $(h, k) \equiv (0, 1)$  or  $(1, 0) \pmod{2}$ , we may have solutions from Classes 1, 3, and 4 (and possibly 6). Note that we can also have multiple solutions from the same Class. Furthermore, we can also use our knowledge of the restrictions and Table 3 to find the exact conditions under which we will have solutions in any class.





**Fig. 13.** The structure of the two known viruses with disymmetrons, which have  $(h, k) = (2, 4)$ . (A) Expected example: Class 3 virus with  $d = 3, t = 3, p = 2$ . (B) The real, irregular virus with irregular trisymmetrons (Jaalinioja, 2007).

One of the most simple and useful observations is that, in most cases, there will be up to our solutions, except in the cases of  $(h, k) \equiv (0, 0) \pmod{2}$ , when there will generally be up to six solutions, and of  $h = k \equiv 1 \pmod{2}$ , when there will be two solutions. As an example, Table 4 takes some values of  $h$  and  $k$  and lists all possible solutions of  $d, t$ , and  $p$  and the Classes they belong to.

The formulas and parity conditions clearly show that the classes from Sinkovits and Baker (2010) are not simply degenerate cases of the ones we have found. Indeed, since we consider  $d > 0$ , the disymmetron always creates an extra line of capsomers separating the two edges of pentasymmetrons of adjacent vertices. Moreover, the parity conditions are also consistent with previously known facts such as that even  $T$  numbers are impossible without disymmetrons, as seen in Sinkovits and Baker (2010) and Wrigley (1969).

The solutions presented in this paper suggest the existence of certain configurations of symmetrons, and as in the case of viruses with no disymmetrons Sinkovits and Baker (2010), it would be interesting to consider whether the configurations presented here physically occur in reality. Moreover, since all the cases are mathematically consistent, if those viruses may not exist in Nature, then further biological rules would need to be governing the behaviour of the viruses (in the discussion presented in Sinkovits and Baker (2010), examples of features apparently favoured in Nature are given).

Finally, whilst the viruses studied in this paper have regular structure and symmetrons, one may consider relaxing some of the mathematical constraints imposed, in order to classify non-regular viruses. This would be necessary in order to understand, for instance, the appearance of two known viruses P23-77 and SH1 that have disymmetrons (Happonen, 2012; Jaalinioja, 2007), and which have the same general structure, as depicted in Fig. 13 (A).

The viruses P23-77 and SH1 have structure invariants  $(h, k) = (2, 4)$ . Our theory's five solutions—Class 2 with edge lengths  $(d, t, p)$  given by  $(1, 2, 3)$ ; Class 2 with  $(3, 0, 3)$ ; Class 3 with  $(3, 3, 2)$  (shown in Fig. 13 (B)); Class 4 with  $(5, 2, 2)$ ; or Class 4 with  $(5, 3, 1)$ —are simply theoretically possible arrangements of regular symmetrons, but these viruses do not fit into any of our solutions because of their irregular trisymmetrons, which are not equilateral triangles. However, it is also possible that P23-77 and SH1 are too small to possess the higher-level order captured by symmetrons.

## Acknowledgments

The research in this paper was conducted by the first author under the supervision of the latter, as part of the MIT PRIMES-USA program. Both authors would like to thank Robert Sinkovits for his enlightening comments, Steven Bradlow for inspiring conversations on the geometry of viruses and MIT PRIMES-USA for

the opportunity to conduct this research together. The work of LPS is partially supported by NSF DMS-1509693.

## Appendix A. Proof that bordering symmetrons must exist if disymmetrons exist.

**Proposition 1.** *Bordering symmetrons must exist if disymmetrons exist.*

**Proof.** We set the coordinate grid so that the disymmetron occupies the lattice points on  $k = 0$ . Consider the adjacent  $k = 1$  lattice points. We will prove the statement by contradiction, so assume that no symmetrons border the disymmetron. Recalling which symmetrons may be adjacent to a disymmetron, the lack of bordering means that there can be at most one adjacent  $k = 1$  pentasymmetron capsomer, at most one adjacent  $k = 1$  trisymmetron capsomer, and at most two adjacent  $k = 1$  disymmetron capsomers, for a total of at most four adjacent  $k = 1$  capsomers, implying that  $1 \leq d \leq 3$ .

If  $d = 1$ : if disymmetrons are adjacent, then the 6-fold symmetry relating adjacent disymmetrons forces that  $p = 1$  and  $t = 0$ , in which case bordering occurs. If disymmetrons are not adjacent, then the six points around the disymmetron must be occupied by capsomers of at most two pentasymmetrons and two trisymmetrons. By the Pigeonhole Principle, some adjacent symmetron has at least two adjacent capsomers, in which case bordering occurs.

If  $d = 2$ , there are three adjacent capsomers on each side, and we can do casework based on what type of capsomer is in the middle. The middle capsomer can not be a pentasymmetron capsomer, or else the pentasymmetron would be bordering. If the middle bordering capsomer is a disymmetron capsomer: without loss of generality, one may set one disymmetron at  $(0, 0)$  and  $(1, 0)$  and the other at  $(0, 1)$  and  $(0, 2)$ . Consider the capsomer at  $(1, 1)$ . This is a middle bordering capsomer to the disymmetron at  $(0, 1)$  and  $(0, 2)$ , so it must be a disymmetron capsomer. This fixes  $t = 0$  and  $p = 1$ , in which case bordering occurs. Finally, if the middle bordering capsomer is a trisymmetron capsomer, the disymmetron is at  $(0, 0)$  and  $(1, 0)$ . A trisymmetron has a vertex at  $(0, 1)$  and has a fixed orientation to avoid bordering. It has other vertices at  $(0, t)$  and  $(1 - t, t)$ . Moreover, 3-fold symmetry means a disymmetron exists from  $(0, t + 1)$  to  $(1, t)$ , and 6-fold symmetry relating adjacent disymmetrons fixes a pentasymmetron center at  $(1 + t, 0)$ . But this means that capsomers with  $h = 1$  and  $1 \leq k \leq t - 1$  are unoccupied, and they can not be penta-, tri-, or disymmetrons. This contradiction implies that  $t - 1 < 1$ , meaning  $t = 1$ , which leads to bordering.

If  $d = 3$ : since we assume no bordering occurs, among the adjacent  $k = 1$  capsomers, there must be one adjacent pentasymmetron capsomer (with  $p > 1$ ), one adjacent trisymmetron capsomer (with  $t > 1$ ), and two adjacent disymmetron capsomers. We set the disymmetron from  $(1, 0)$  to  $(3, 0)$ . For  $p > 1$  to be true, the adjacent pentasymmetron capsomer must be at  $(0, 1)$  or  $(3, 1)$ . Without loss of generality, we set it at  $(0, 1)$ , so 2-fold symmetry means that another adjacent pentasymmetron capsomer is at  $(4, -1)$ . If there were an adjacent disymmetron capsomer at  $(1, 1)$ , then it would need to lie from  $(1, 1)$  to  $(1, 3)$  to be properly positioned with respect to 6-fold symmetry about the pentasymmetron containing  $(0, 1)$ . However, by the 6-fold symmetry relating adjacent disymmetrons, this would force  $p = 1$ , which is a contradiction and leads to bordering. Therefore, the adjacent disymmetron capsomers must be at  $(2, 1)$  and  $(3, 1)$ , and the trisymmetron bordering capsomer is at  $(1, 1)$ . Because  $t > 1$ , the trisymmetron also includes  $(0, 2)$  and  $(1, 2)$ . This fixes the orien-



tation of the disymmetron through (2, 1), which must go from (2, 1) to (2, 3). But this disymmetron will border the trisymmetron, which leads to a contradiction. Thus, bordering must occur if  $d > 0$ .  $\square$

## Appendix B. On rotations in the $h, k$ grid.

**Proposition 2.** A point  $(x, y)$  rotated  $\frac{\pi}{3}$  counterclockwise about the origin is mapped to  $(-y, x + y)$ .

**Proof.** A point  $(x, y)$  can be interpreted as the vector  $x\hat{h} + y\hat{k}$ . If we denote the rotation with the function  $R$  acting on some vector  $\vec{v}$ , then notice that  $R$  is a linear function, i.e.  $R(\vec{v}_1 + \vec{v}_2 + \dots + \vec{v}_n) = R(\vec{v}_1) + R(\vec{v}_2) + \dots + R(\vec{v}_n)$  and  $R(a\vec{v}) = aR(\vec{v})$ . This means that

$$R(x\hat{h} + y\hat{k}) = xR(\hat{h}) + yR(\hat{k}).$$

We can see that  $R(\hat{h}) = \hat{k}$  and  $R(\hat{k}) = \hat{k} - \hat{h}$ . Therefore,  $R(x\hat{h} + y\hat{k}) = x\hat{k} + y\hat{k} - y\hat{h}$ , which is  $(-y, x + y)$ .  $\square$

In particular, the above implies the following:

**Proposition 3.** A point  $(x, y)$  rotated  $\frac{2\pi}{3}$  counterclockwise about the origin is mapped to  $(-x - y, x)$ .

By translating the grid so that  $(x_0, y_0)$  becomes the origin, one takes  $(x, y) \rightarrow (x - x_0, y - y_0)$ , leading to the following:

**Proposition 4.** A point  $(x, y)$  rotated  $\frac{\pi}{3}$  counterclockwise about  $(x_0, y_0)$  is mapped to  $(-y + x_0 + y_0, x + y - x_0)$ .

**Proof.** Applying the above results, translation followed by  $\frac{\pi}{3}$  counterclockwise takes

$$(x - x_0, y - y_0) \rightarrow (y_0 - y, x + y - x_0 - y_0).$$

The statement follows by translating the grid back,  $(y_0 - y, x + y - x_0 - y_0) \rightarrow (-y + x_0 + y_0, x + y - x_0)$ .  $\square$

Similarly, since

$$\begin{aligned} (x, y) &\rightarrow (x - x_0, y - y_0) \\ &\rightarrow (x_0 + y_0 - x - y, x - x_0) \\ &\rightarrow (-x - y + 2x_0 + y_0, x - x_0 + y_0). \end{aligned}$$

one has the following:

**Proposition 5.** A point  $(x, y)$  rotated  $\frac{2\pi}{3}$  counterclockwise about the point  $(x_0, y_0)$  is mapped to the point  $(-x - y + 2x_0 + y_0, x - x_0 + y_0)$ .

## Appendix C. Adjacency of capsomers

**Proposition 6.** In the presence of disymmetrons, two capsomers from different pentasymmetrons or two capsomers from different trisymmetrons cannot be adjacent.

**Proof.** In the following proof, the word “symmetron” is used as shorthand for pentasymmetrons and trisymmetrons. By a symmetry of overlapping argument, if the two symmetrons have adjacent capsomers, then they are centered on adjacent vertices/ faces of the icosahedron. Therefore, one symmetron can be mapped to the other by 2-fold symmetry about a 2-fold center. Because equi-

lateral triangles and hexagons are convex and cannot contain the 2-fold center, two rays can be drawn from the 2-fold center such that the region between them, with an angle  $< \pi$ , contains a symmetron. If we reverse the direction of these two rays, they must bound the other symmetron by 2-fold symmetry. Therefore, we can draw a line  $l$  through the 2-fold center that passes through neither of the bounded regions, so it separates the symmetrons. Because of the shape of these two symmetrons, we can always draw  $l$  parallel to one of the lines  $h = 0, k = 0$ , or  $h + k = 0$ .

The 2-fold center can either be a lattice point or halfway between two adjacent lattice points. In the first case, line  $l$ , which is along the triangular grid, separates the symmetrons and does not contain any capsomers of the symmetrons, making adjacency impossible. In the other case, the 2-fold center is between two adjacent lattice points. We orient the coordinate system so that the 2-fold center is  $(\frac{1}{2}, \frac{1}{2})$ . As we saw in the first case, line  $l$  can not be a line along the triangular grid for adjacency to occur, thus without loss of generality, we may assume that line  $l$  has equation  $k = \frac{1}{2}$ .

The disymmetron, which must contain the 2-fold center, must have capsomers at least at (1, 0) and (0, 1). Consider the symmetron with  $k \geq 1$ . For adjacency to occur, this symmetron must have capsomers with  $k = 1$  and  $h < 0$  or  $h > 0$ , and the opposite symmetron must have capsomers with  $k = 0$  and  $h < 0$  or  $h > 0$ , respectively. Moreover, 2-fold symmetry dictates that the  $k \geq 1$  symmetron has capsomers with  $k = 1$  and  $h > 0$  or  $h < 0$ , respectively. So this symmetron has  $k = 1$  capsomers on both sides of  $h = 0$ . But the disymmetron lies at (0, 1). Thus, the convexity of the symmetrons makes this impossible and the statement follows.  $\square$

## Appendix D. A pseudo-code test for Classe

Seen as vectors, the distance squared from (0, 0) to  $(x, y)$  is  $x^2 + xy + y^2$ . We can use the coordinates from Table 2 to find this edge length squared in terms of  $p$  and  $t$ . This motivates the following pseudo-code test, which takes inputs  $d, t, p$ , and  $T$ , and outputs the Class number:

---

```

if (d==0) return 1;
else {
  if (3p2+t2-2t+1==T) return 2;
  if (3p2-6p+t2+4t+7==T && t!=p-1) return 3;
  if (3p2+t2+4t+4==T) return 4;
  if (3p2+4==T && t==2 && p!=2) return 5;
  if (t2+4t+4==T && p==2) return 6;
}

```

---

There are a few overlaps in the Classes, and the numerical test helps find them. The first one is between Classes 2 and 3 when  $t = p - 1$ . These in fact look the same, but according to our bordering definitions, this case ought to belong to Class 2. Similarly, there is an overlap between Classes 5 and 6 when  $p = t = 2$ . Again they appear to be the same, but bordering definitions dictate that this configuration belongs to Class 6. In this way, the above numerical test allows us to identify the unique Class that each configuration belongs to.

## References

- Grytczuk, Aleksander, 2006. Ljunggren's Diophantine problem connected with virus structure. *Ann. Math. Inf.* 33, 69–75.
- Happonen, Lotta, 2012. Life on the Edge-Structural Studies of the Extremophilic Viruses P23-77 and STIV2. University of Helsinki, Helsinki (Ph.D thesis).
- Jaalinoja, Harri, 2007. Electron Cryo-microscopy Studies of Bacteriophage Phi 8 and Archaeal Virus SH1. University of Helsinki, Helsinki (Ph.D thesis).

- Klose, Thomas, Reteno, Dorine G., Benamar, Samia, Hollerbach, Adam, Colson, Philippe, La Scola, Bernard, Rossmann, Michael G., 2016. Structure of faustovirus, a large dsDNA virus. *PNAS* 113 (22), 6206–6211.
- Klug, A., Caspar, D.L.D., 1962. Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* 27, 1–24.
- Manyakov, V.F., 1977. Fine structure of the iridescent virus type I capsid. *J. Gen. Virol.* 36, 73–79.
- Phillips, F.C., 1956. *An Introduction to Crystallography*. Longmans and Green. 324 p.
- Sinkovits, R.S., Baker, T.S., 2010. A tale of two symmetrons: rules for construction of icosahedral capsids from trisymmetrons and pentasymmetrons. *J. Struct. Biol.* 170, 109–116.
- Watson, James, Crick, Francis, 1956. Structure of small viruses. *Nature* 177 (4506), 473–475.
- Wrigley, N.G., 1969. An electron microscope study of the structure of sericesthis iridescent virus. *J. Gen. Virol.* 5 (1), 123–134.
- Yan, X., Olson, N.H., Van Etten, J.L., Bergoin, M., Rossmann, M.G., Baker, T.S., 2000. Structure and assembly of large lipid-containing dsDNA viruses. *Nat. Struct. Biol.* 7, 101–103.
- Yan, X., Chipman, P.R., Castberg, T., Bratbak, G., Baker, T.S., 2005. The marine algal virus PpV01 has an icosahedral capsid with T = 219 quasisymmetry. *J. Virol.* 79, 9236–9243.
- Yan, X., Yu, Z., Zhang, P., Battisti, A.J., Holdaway, H.A., Chipman, P.R., Bajaj, C., Bergoin, M., Rossmann, M.G., Baker, T.S., 2009. The capsid proteins of a large, icosahedral dsDNA virus. *J. Mol. Biol.* 385, 1287–1299.