Icosahedral Virus Transitions: Computational Techniques
WMRUGS Talk: Paper Format

Xavier Silva

December 12, 2023

Contents

1 Quick Summary

Parallel computation allows for efficient identification of the symmetries that are preserved during maturation of icosahedral viruses.

2 Group Theory

Group theory is a subset of abstract algebra. In particular, group theory studies the mathematical object known as a group, defined in definition 1.

Definition 1 (Group)**.** A group G is a set along with a binary operation $*$ that satisfy the following properties:

1. Closure: If $a, b \in G$, then $a * b \in G$.
2. Associativity: If $a, b, c \in G$, then $(a * b) * c = a * (b * c)$.
3. Identity: There exists $e \in G$ such that $a * e = e * a = a$ for all $a \in G$.
4. Inverses: For all $a \in G$ there exists $b \in G$ such that $a * b = b * a = e$. Thus $b = a^{-1}$.

Beyond the rigorous definition, groups can be thought of as symmetries of an object (one can think about this through group actions). The idea of groups representing symmetry is crucial to understanding the structure of icosahedral viruses.

2.1 Icosahedral Group

The set of symmetries on an icosahedral virus is the icosahedral group $\mathcal{I}$.

Definition 2 (Icosahedral Group)**.** The icosahedral group, denoted by $\mathcal{I}$, is the 60-element group given by

$$\mathcal{I} := \langle a, b | a^2 = b^3 = (ab)^5 = 1 \rangle$$

These viruses have 2-fold, 3-fold, and 5-fold symmetries, which relates to the structure of the icosahedral group. We wish to investigate whether icosahedral symmetry is preserved during virus maturation.

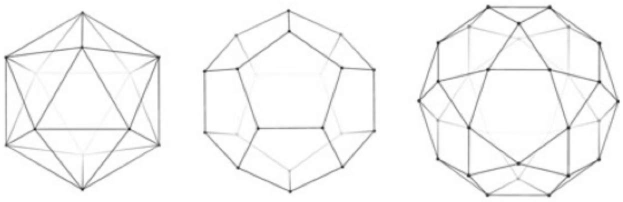


Figure 1: Images of the icosahedron, dodecahedron, and icosidodecahedron.

2.2 Realization of Icosahedral Group

We realize icosahedral symmetry as a 6×6 matrix group. The operation matrix groups use is matrix multiplication.

Definition 3 (Generators of $\mathcal{I}$)**.** The generators of $\mathcal{I}$ are:

$$a = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad b = \begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

2.3 Maximal Subgroups of $\mathcal{I}$

The icosahedral group has maximal subgroups $A_4, D_{10},$ and D_6.

- A_4 has multiple 2-fold and 3-fold axes.
- D_6 has a singular 2-fold and 3-fold axis.
- D_{10} has a 5-fold and a 2-fold axis.

These are subsets of $\mathcal{I}$ and more specificially, maximal subgroups of $\mathcal{I}$. Similar to how the icosahedral group represents the shape of an icosahedron, these maximal subgroups represent their own shapes. These shapes are shown in figure 2. We use these subgroups because if we are unable to preserve all of icosahedral symmetry, then we wish to at least preserve the symmetry of one of these maximal subgroups.

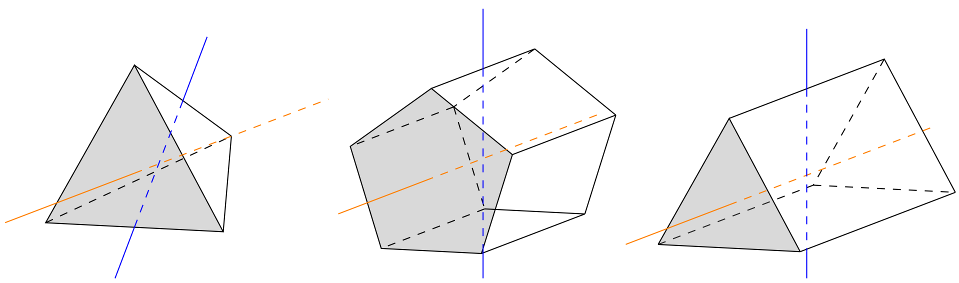


Figure 2: A visual representation of the maximal subgroups. From left to right we have A_4, D_{10}, and D_6. The two fold axes are labeled in blue while the 3 or 5 fold axis is labeled in orange.

3 Point Arrays and Lattices

Icosahedral viruses are characterized by point arrays, which are a set of points in 3 dimensions.

Definition 4 (Icosahedral Point Array)**.** Let $\mathbf{t}, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be 6 dimensional vectors. Then the icosahedral point array generated by these vectors is

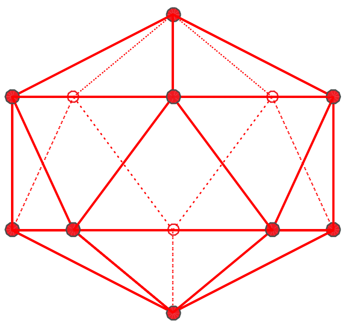
$$P = \mathcal{I}\mathbf{v}_1 \cup \mathcal{I}\mathbf{v}_2 \cup \cdots \cup \mathcal{I}\mathbf{v}_n \cup (\mathcal{I}\mathbf{v}_1 + \mathcal{I}\mathbf{t}) \cup (\mathcal{I}\mathbf{v}_2 + \mathcal{I}\mathbf{t}) \cup \cdots \cup (\mathcal{I}\mathbf{v}_n + \mathcal{I}\mathbf{t}).$$

Notice that while icosahedral viruses are characterized by 3 dimensional point arrays, *icosahedral point arrays* are 6 dimensional. This is because we want icosahedral point arrays to fit inside icosahedral lattices, and no such lattices exist in dimensions lower than 6.[1] However, we can project icosahedral point arrays from 6 dimensions into 3 dimensions using the projection matrix given in [2].

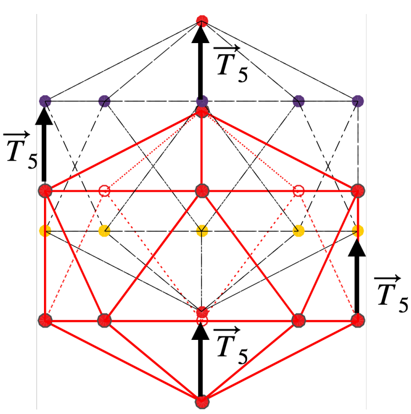
There are 55 standard point arrays from which we build all others. These are called the *one-base* point arrays and take the form $\mathcal{I}\mathbf{v} \cup (\mathcal{I}\mathbf{v} + \mathcal{I}\mathbf{t})$. This form tells us that only one translation vector and one "base" vector is needed to create the point array. [1]

3.1 Constructing Point Arrays

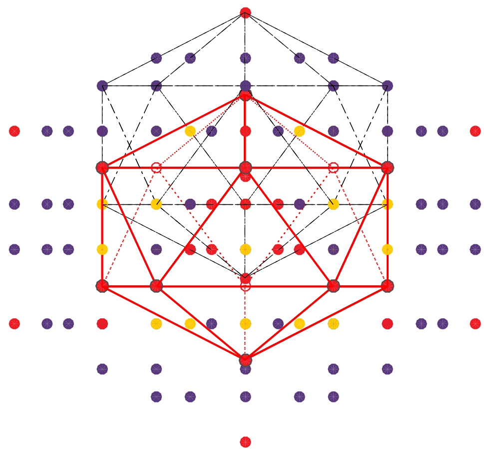
Constructing a one-base point array involves us finding all vectors in the set $\mathcal{I}\mathbf{v} \cup (\mathcal{I}\mathbf{v} + \mathcal{I}\mathbf{t})$. This is shown visually in figure 3.



(a) Start with icosahedron, created by applying icosahedral symmtry on one vector. Mathematically this is $\mathcal{I}\mathbf{v}$.



(b) Shift icosahedron by a translation vector. Mathematically this is $\mathcal{I}\mathbf{v} \cup (\mathcal{I}\mathbf{v} + \mathbf{t})$.



(c) Reapply icosahedral symmetry, getting the final expression $\mathcal{I}\mathbf{v} \cup (\mathcal{I}\mathbf{v} + \mathcal{I}\mathbf{t})$.

Figure 3: The process of creating a one-base point array.

4 Viral Maturation

Icosahedral viruses mature over time, and maturation is how they become infectous. For this paper we will focus on the Cowpea Chlorotic Mottle Virus (CCMV). This virus is characterized by one of the standard point arrays. How CCMV matures is shown by figure 4 on the following page. Notice the points in orange, these points collectively are the point arrays that characterize the CCMV virus. We wish to see if there exists a linear transformation that maps the native point array to the mature point array while preserving icosahedral symmetry (or one of its maximal subgroups).

5 Mathematical View of the Problem

To find a linear transformation that maps from native to mature point arrays while preserving symmetry, we solve matrix equations of the form

$$TB_0 = B_1.$$

In this equation, B_0 and B_1 represent the native and mature point arrays respectively, and T is our desired linear transformation.[2] Since B_0 and B_1 are representatives of the point arrays, there exist many different equations of the form $TB_0 = B_1$ that represent the same problem. Therefore, we only need to find one equation of this form that is solvable.

[1] Because lattices of dimension 5 and lower cannot preserve 5-fold symmetry.

[2] Recall that we realize icosahedral symmetry as a matrix group and icosahedral point arrays are sets of vectors, so it makes sense that we are solving a matrix equation.

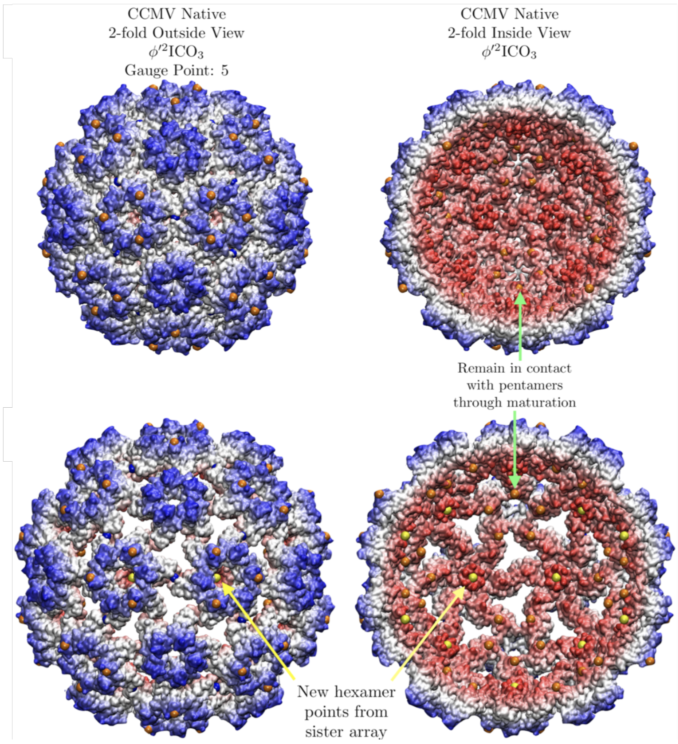


Figure 4: The above diagram shows CCMV in its native state (first row) and mature state (second row). Notice the points in orange are the points that characterize CCMV.

In our example of CCMV, these equations will take the form

$$T \cdot \begin{bmatrix} | & | \\ \mathbf{t}_0 & \mathbf{v}_0 \\ | & | \end{bmatrix} = \begin{bmatrix} | & | \\ \mathbf{t}_1 & \mathbf{v}_1 \\ | & | \end{bmatrix}$$

and we wish to find a choice of these four vectors such that the equation is solvable.

5.1 Transitions Preserving Symmetry

A transition T that preserves all of icosahedral symmetry must have the form:

$$T = \begin{bmatrix} z & x & -x & -x & x & x \\ x & z & x & -x & -x & x \\ -x & x & z & x & -x & x \\ -x & -x & x & z & x & x \\ x & -x & -x & x & z & x \\ x & x & x & x & x & z \end{bmatrix} \tag{1}$$

Our linear transformation T preserves icosahedral symmetry because it is in the centralizer of $\mathcal{I}$ (or one its of maximal subgroups).

Definition 5 (Centralizer)**.** The centralizer of a group G, denoted by $Z(G)$ is the set of elements that commute with all elements of G. That is,

$$Z(G) = \{z \mid gz = zg \ \forall g \in G\}.$$

Any matrix of the form given in equation 1 is in $Z(\mathcal{I})$, the centralizer of $\mathcal{I}$. Similar general matrix forms exist for matrices that are in $Z(A_4)$, $Z(D_{10})$, and $Z(D_6)$. [2]

5.2 CCMV D_6 Transition Example

The following is an example of an equation for the CCMV virus the preserves D_6 symmetry:

$$\begin{bmatrix} u & w & -w & x & s & s \\ -t & y & v & -v & z & -t \\ t & v & y & v & t & -z \\ z & -v & v & y & -t & -t \\ s & x & -w & w & u & s \\ s & w & -x & w & s & u \end{bmatrix} \cdot \begin{bmatrix} -1/2 & -3/2 \\ -1/2 & 1/2 \\ 1/2 & -1/2 \\ -1/2 & -1/2 \\ -1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ -1 & 0 \\ 0 & -1 \\ 0 & -1 \\ 0 & -1 \end{bmatrix} \tag{2}$$

In this equation, the 6×6 matrix on the left is the general form for linear transformations that preserve D_6 symmetry. We have also chosen example vectors that generate the CCMV native and mature point arrays.

This equation is only one of many that represents a CCMV D_6 transition. In order to compute how many of these equations exist, we simply need to mutiply the sizes of the icosahedral orbits of each vector individually.

$$\left| \mathcal{I} \begin{bmatrix} -1/2 \\ -1/2 \\ 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix} \right| = 20 \quad \left| \mathcal{I} \begin{bmatrix} -3/2 \\ 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \\ 1/2 \end{bmatrix} \right| = 12$$

$$\left| \mathcal{I} \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right| = 12 \quad \left| \mathcal{I} \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \\ -1 \\ -1 \end{bmatrix} \right| = 20$$

Above we have the sizes of the icosahedral orbits for each vector used in equation 2 on the previous page. Therefore there are $20 \cdot 12 \cdot 12 \cdot 20 = 57600$ equations of the form $TB_0 = B_1$ that could result in a D_6 symmetry preserving transition for CCMV.

6 Computational Techniques

In order to find a solvable equation (or say such an equation doesn't exist) of the form $TB_0 = B_1$, we must search through up to hundreds of thousands of these equations as the previous example illustrates. This is infesible to do by hand, so we employ various programs to find solvable equations.

6.1 Brute Force in C++

As a first approach, we try to brute force every possible transitions matrix. In order to do this, we take a set of values and create every possible transition matrix from those values. If we take the set $E = \{0, \pm\frac{1}{2}, \pm 1\}$ as the set of values, then for our example equation (see equation 2), then there will be $|E|^8 = 390625$ possible transition matrices to check. Because there are thousands of equations we need to check that look like our example, this approach quickly becomes infeasible. However, it was useful at the beginning of this research in order to get a grasp on the problem and get some results.

6.2 Equation Solving in Python

The final approach to finding transitions is to use a linear matrix equation solver within the Python `sympy` package. This package allows us to give it an equation such as equation 2 and it will tell us if a solution exists or not. We can simply loop through all possible equations and give it to `sympy` and it will tell us if any equations can be solved. If we find such an equation with a solution, then there exists a symmetry preserving transition. If we cannot find any equation, then there does not exist any symmetry preserving transitions.

6.3 Parallel Computing on Jigwé

Since checking of each equation is an independent task, this problem is said to be *embarassingly parallel.* Therefore we can speed up computation by employing parallel computing. Kalamazoo College has a supercomputer known as Jigwé (pronouced 'Cheekua' and is Potawatomi for Thunderbird) in which we can perform parallel computing.

7 Conclusion

By running my programs, I have produced the following results:

- Reproduce D_6 transitions for CCMV found in [2].
- I have figured out what symmetry preserving transitions exist between all of the 55 standard point arrays. Larger point arrays must be built from these 55, so this data also helps in determining what transitions are and are not possible between bigger point arrays.
- There cannot exist any transition that preserves all of icosahedral symmetry between different point arrays, since no such transition exists between any of the 55 standard point arrays.

8 Future Directions

As a final thought, a few future directions one could take with this research:

- I have been solving these equations in 6 dimensions, but since we live in 3 dimensions, we could project these point arrays into 3 dimensions, using the projection matrix given in [2]. Since multiple 6 dimensional transitions exist, it may be possible that they project onto the same transition in 3 dimensions.
- I could attempt to find all possible transitions between larger point arrays because even if a transition exists between all of its components, a transition may not exist for the larger point array.

9 References

[1] Thomas Keef and Reidun Twarock. Affine extensions of the icosahedral group with applications to the three-dimensional organisation of simple viruses. *Journal of mathematical biology*, 59:287–313, 2009.

[2] Giuliana Indelicato, Paolo Cermelli, David G. Salthouse, Simone Racca, Giovanni Zanzotto, and Reidun Twarock. A crystallographic approach to structural transitions in icosahedral viruses. *J. Math. Biol.*, 64(5):745–773, 2012.