

Decomposition of Virus Normal Modes into Spherical Harmonics

An exploration of Symmetry Adapted Functions (SAFs)

Evan Voyles

Advisor: Dr. David Wilson

A thesis presented for the degree of
Bachelors of Arts



Department of Physics
Kalamazoo College
13 February 2020

Contents

1	Introduction	2
1.1	Viruses	2
1.2	Icosahedral Symmetry	2
1.3	T-Number	3
1.4	Normal Modes	5
1.5	SAFs	6
2	Background	7
2.1	Normal Mode Analysis	7
2.1.1	Eigenmodes	12
2.1.2	Wave Equation	13
2.1.3	Calculations	13
2.2	Spherical Harmonics	14
2.2.1	Laplacian	15
3	Methods	16
3.1	Overview	16
3.2	SAFs	16
3.3	Initialization	22
3.3.1	Initialization example	22
3.4	SAFs Calculations and Coordinate Systems	25
3.4.1	SAF Coordinates	25
3.4.2	Virus SAF Generation	27
3.5	Decomposition	29
3.6	Visualizing	29
4	Results	32
4.1	Virus Analysis	37
4.1.1	Turnip Yellow Mosaic Virus (1auy, T = 3)	37
4.1.2	Cowpea Chlorotic Mottle Virus (1cwp, T = 3)	40
4.1.3	Pariacoto Virus (1f8v, T = 3)	42
4.1.4	Nudaurelia Capensis ω Virus (1ohf, T = 4)	44
4.1.5	Bacteriophage Q β Capsid (1qbe, T = 3)	46
4.1.6	Simian Virus 40 (1sva, T = 7d)	48
4.1.7	Bursal Disease Virus VP2 (2df7, T = 1)	50
4.1.8	Bacteriophage MS2 (2ms2, T = 3)	52
4.1.9	Human Papilloma Virus Type 16 (3j6r, T = 7d)	54
5	Conclusion	55

1 Introduction

1.1 Viruses

Viruses are the most abundant parasites on Earth, infecting a wide variety of living organisms ranging from simple plants (e.g. tobacco mosaic virus) to larger hosts, such as humans (e.g. flu virus, HIV). Viruses are difficult to classify as being biologically alive or “not alive,” as they contain the fundamental elements of living organisms, DNA or RNA, but they do not have the ability to independently reproduce. That is, viruses rely on infecting a host and hijacking its cell’s machinery in order to produce proteins from its DNA (or RNA). Viruses must either contain DNA or RNA – but not both – and can be made up of either single or double stranded nucleic acids. The simplest viruses are composed of genetic material enclosed in a protective protein outer shell that is called the viral capsid. The shape and behavior of each virus can be classified into three different categories, spherical (or polyhedral), helical, and bacteriophages(also referred to as phages). Spherical viruses are viruses whose protein capsids exhibit icosahedral symmetry (ico – 20, hedral – faces), helical viruses have rod-shaped capsids, and bacteriophages – as the name suggests – are viruses that have hijacked a bacterium.

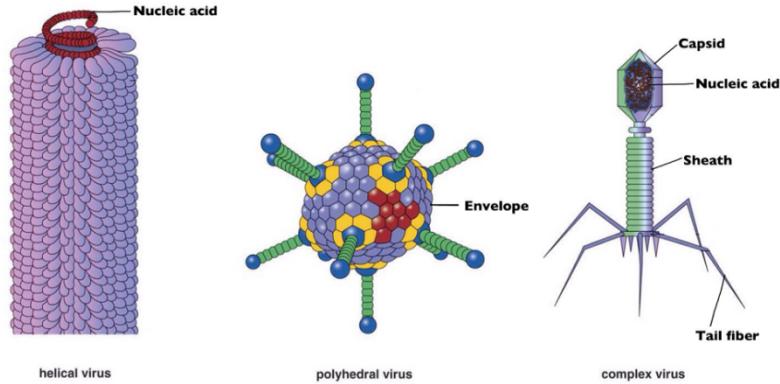


Figure 1: The different shapes of viruses[6]. Viruses can be helical in structure, they can have polyhedral symmetry (often called spherical viruses), or they can also be found in the form of a bacteriophage.

This paper will focus on spherical viruses whose icosahedral capsid structure has been extensively classified by a triangulation number (T) developed by Casper and Klug.

1.2 Icosahedral Symmetry

Viruses are so tiny that there is not enough space within the capsid to produce a large number of unique proteins to form their outer shell. So, viruses use symmetry as a way to combat spatial limitations. Instead of being made

up of different proteins, a virus's DNA or RNA only provide the instructions to form a few distinct proteins, which are symmetrically organized to form a spherical capsid. Icosahedral symmetry is nature's solution to form a symmetric protective shell.

A regular icosahedron is a shape that has 60 rotational symmetries, and can be reduced to a simple, repeating asymmetric unit. Icosahedral symmetry demands the presence of 5-Fold, 3-Fold, and 2-Fold symmetry axis, which describe how many times a shape can be rotated while preserving orientation.

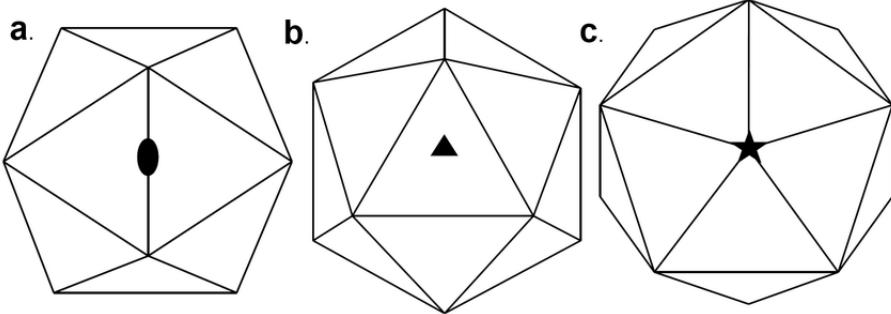


Figure 2: Symmetry axes of a regular icosahedron. (a) is a 2-fold axis, (b) is a 3-fold axis, (c) is a 5-fold axis[7]. For a two fold axis, there are two rotations that maintain the shapes orientation. If we rotate a 2 fold axis 180 degrees, then its orientation is preserved. A three fold axis needs to be rotated 120 degrees in order for it to look the same, and a 5-fold axis has 5 rotations of 72 degrees.

1.3 T-Number

The triangulation number, or T-number, provides information about the arrangement and quantity of proteins that form the viral capsid. Starting with the number of proteins, we take the T-number and multiply it by 60. Therefore, a $T = 1$ virus has 60 proteins on its capsid, a $T = 3$ virus has 180 proteins, a $T = 4$ virus has 240 proteins, and so on. Notice how I skipped a ' $T = 2$ ' virus when stating the number of proteins in the viral capsid. This is because a $T = 2$ is actually impossible, given that the formula to determine T number is:

$$T = h^2 + hk + k^2 \quad (1)$$

Where h and k are the number of hexamers that must be traversed in each direction to connect two pentamers.

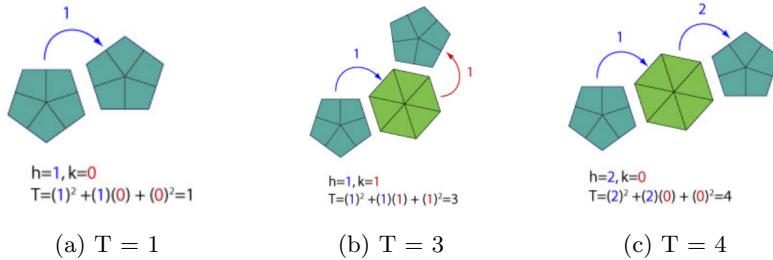


Figure 3: A demonstration of T-number[5], where a step in each direction is represented by h and k . For example, in Figure 3.a, it only takes one step in the same direction to reach another pentamer. Therefore, $h = 1$, and $k = 0$, so the virus would be a $T = 1$ virus. In Figure 3.b, it takes 1 step in each direction, so both h and k are equal to 1. $T = 1^2 + (1)(1) + 1^2 = 3$. In the final case, we have an example where it requires two steps in the same direction in order to link two pentamers.

Let's take a look at some spherical viruses so we can see where the pentamers and hexamers line up.

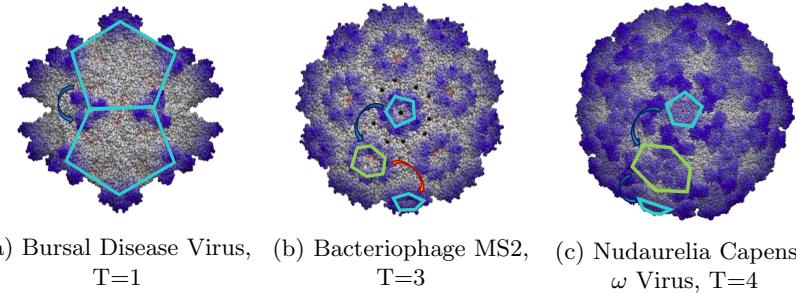


Figure 4: A demonstration of T-number, where a step in each direction is represented by h and k . Figure 4 provides a real life analog to the T-number orientations shown in Figure 3.

As we increase the T-number, the more “sphere-like” the viral capsid becomes. It would be interesting to see if the T-number affects how a virus oscillates, and if there are any specific patterns of vibrations that are dependent on T-number. We are also interested in the T-number of a virus because the orientation of proteins changes how the symmetry axes line up with the intersections of pentamers and hexamers.

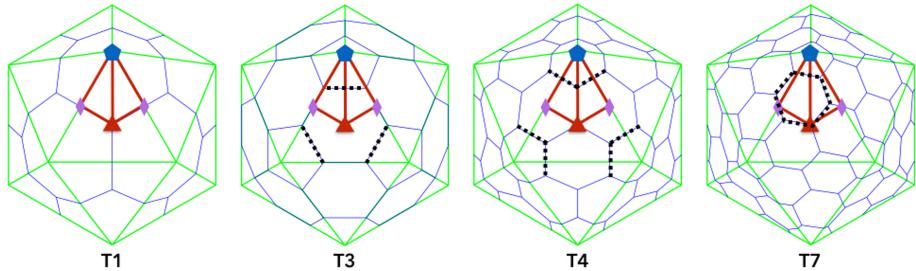


Figure 5: T-Number affects where structural pentamers and hexamers line up with the symmetry axes.[4]. For a T1 virus, we see that there is an intersection of the vertices of 3 pentamers at the 3-fold axis. We see that the 2-fold axis lines on the border of two pentamers, and the 5-fold axis lies right in the center of each pentamer. A T3 virus is oriented slightly differently, with the 2-axis coinciding with the border of two hexamers, the 3-fold axis being found in the middle of each hexamer, and the 5-fold axis once again in the center of a pentamer. By its very nature, all 5-fold axes will be found in the center of each pentamer.

The exact alignment of pentamers and hexamers could influence how the virus decides to oscillate about the symmetry axes. For example, for a $T = 3$ virus we see that at the 2-fold axis (as indicated by the diamond) there is a border between two hexamers. Because this is not a shared vertex between hexamers and rather a side that they have in common, it's possible that this area is very rigid and we will see very little oscillation about the 2-fold axis for $T=3$ viruses.

1.4 Normal Modes

Any oscillating system, whether it be as simple as a few masses and springs or as complex as a virus composed of hundreds of thousands of atoms vibrating in tandem, can be analyzed and decomposed into normal modes. Normal modes are motions in which all the parts of a system oscillate at the same frequency. Normal modes can be thought of as the most fundamental modes of vibration, as any oscillatory motion of a system can be described as a linear combination of its normal modes. That is to say, the normal modes form a basis for oscillatory motion. Therefore, normal modes can be defined as orthogonal to one another. Our analysis of normal modes will start with elementary systems consisting of springs and masses and work out way up to their eventual abstraction and application to viruses.

This paper attempts to develop a more complete understanding and description of viral normal modes that have been calculated. By decomposing these complicated motions that have a seemingly sporadic outward appearance into a set of defined parameters, we provide the tools necessary to compare, describe, and validate the “authenticity” of viral normal modes. The decomposition that is described in this paper makes the use of the spherical harmonic functions,

and more importantly, a special set of linear combinations of Laplace's spherical harmonics.

1.5 SAFs

These special linear combinations, coined as **Symmetry Adjusted Functions** (SAFs), are combinations of spherical harmonics of the same order l that result in spherical functions with defined symmetries, e.g. tetrahedral, octahedral, and icosahedral. For the purpose of decomposing spherical viruses, in-depth analysis of icosahedrally symmetric SAFs will be explored.

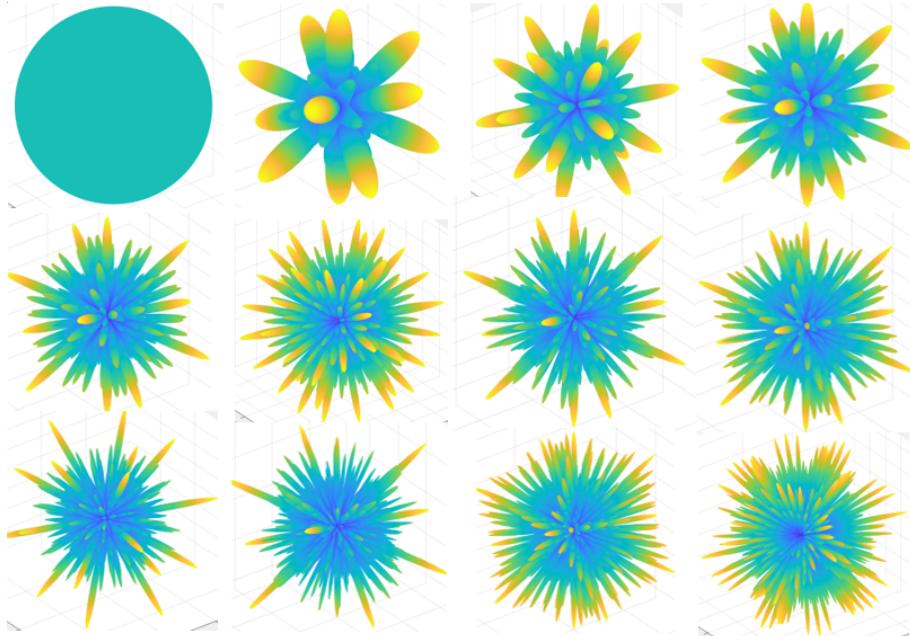


Figure 6: A visual overview of all the icosahedral SAFs plotted in MATLAB. First row: (from left to right) SAF0, SAF6, SAF10, SAF12. Second row: SAF16, SAF18, SAF20, SAF24. Third row: SAF26, SAF28, SAF30, SAF30(2). Spherical harmonics are generated by taking a linear combination of some Real Laplacian spherical harmonic of order l . The input for these functions are polar coordinates, and the output is a radial value that corresponds to the relative radial displacement at that location. The SAFs can only describe the radial oscillations and do not account for twisting and rotating motions that are sometimes present in normal modes calculations.

2 Background

2.1 Normal Mode Analysis

To begin the exploration of decomposing calculated viral eigenmodes into their symmetrically adjusted spherical functions, we need to first develop a clear understanding of what normal modes are and how to solve them. We can examine normal modes starting with the simplest edition of a system that contains multiple oscillating bodies. We start by envisioning 2 masses that are on a horizontal spring, thus cutting the spring into 3 sections.

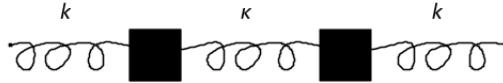


Figure 7: A simple system consisting of 2 masses and 3 springs

If we say that both masses are the same, and that the spring constant for the first and last spring in this system are k , we can denote the middle spring's spring constant as κ . According to Hooke's law, the restoring force of a displaced spring depends on the negative of the spring constant multiplied by the displacement itself.

$$F = -kx \quad (2)$$

If we denote x_1 as the displacement of the first mass, and x_2 as the displacement of the second mass, the force enacted on each mass can be written out as:

$$F_{mass1} = -kx_1 - \kappa(x_1 - x_2) = ma \quad (3)$$

$$F_{mass2} = -kx_2 - \kappa(x_2 - x_1) = ma \quad (4)$$

We can use the displacements of each mass to calculate the force that is applied according to Hooke's law and relate this to Newton's $F = ma$. This relation allows us to compare how the acceleration of each mass changes based on position, which gives us more information about the oscillation of the masses. Because we are going to set both of the force equations equal to each other, we will be solving for a set of differential equations of the second order. So, we can guess solutions of the position functions in the form of

$$x_1(t) = A_1 e^{i\omega t} \quad (5)$$

$$x_2(t) = A_2 e^{i\omega t} \quad (6)$$

in order to describe motion that is harmonic in nature and solve for eigenvalues. The benefits of guessing a solution in the form of $e^{i\omega t}$ are two-fold; first off, using euler's identity,

$$e^{i\theta} = \cos\theta + i * \sin\theta \quad (7)$$

we are able to solve for the most general solution, even extending into the complex plane. Euler's identity is succinct representation of harmonic motion and using the natural constant e makes the derivatives of the position functions very easy to compute. Secondly, having a function in the form of some e^{ikx} allows us to easily calculate the eigenvalues of the second derivative, making this a strong guess to solve this ODE. We can plug these new position function guesses into the set of Hooke's equations.

$$F_{mass1} = [-kA_1 - \kappa(A_1 - A_2)]e^{i\omega t} = m\ddot{x}_1(t) \quad (8)$$

$$F_{mass2} = [-kA_2 - \kappa(A_2 - A_1)]e^{i\omega t} = m\ddot{x}_2(t) \quad (9)$$

When setting the force calculated by Hooke's law to mass times acceleration, the acceleration can actually be written as the second derivative of the position function.

$$\ddot{x}_1(t) = -\omega^2 A_1 e^{i\omega t} \quad (10)$$

$$\ddot{x}_2(t) = -\omega^2 A_2 e^{i\omega t} \quad (11)$$

The second derivative of the position function can therefore be plugged into our equation where $-kx = ma$.

$$[-kA_1 - \kappa(A_1 - A_2)]e^{i\omega t} = -m\omega^2 A_1 e^{i\omega t} \quad (12)$$

$$[-kA_2 - \kappa(A_2 - A_1)]e^{i\omega t} = -m\omega^2 A_2 e^{i\omega t} \quad (13)$$

The overall goal is to solve for a system where all the ω values are the same, indicating that all bodies are oscillating at the same frequency. We can divide out the exponential and bring all the terms to one side.

$$-m\omega^2 A_1 + kA_1 + \kappa A_1 - \kappa A_2 = 0 \quad (14)$$

$$-m\omega^2 A_2 + kA_2 + \kappa A_2 - \kappa A_1 = 0 \quad (15)$$

We can use a system of matrices to take a linear algebra approach to find the solution.

$$\begin{bmatrix} -m\omega^2 + k + \kappa & -\kappa \\ -\kappa & -m\omega^2 + k + \kappa \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (16)$$

The only way that the solution to this set of equations is not trivial, is if A_1 and A_2 are not equal to 0, then the determinant of the first matrix has to be 0, so that the inverse of the matrix does not exist.

$$(-m\omega^2 + k + \kappa)^2 - \kappa^2 = 0 \quad (17)$$

$$-m\omega^2 + k + \kappa = \pm\kappa \quad (18)$$

$$-m\omega_1^2 + k + \kappa = +\kappa \quad (19) \quad -m\omega_2^2 + k + \kappa = -\kappa \quad (20)$$

$$-m\omega_1^2 = +\kappa - \kappa - k \quad (21) \quad -m\omega_2^2 = -\kappa - \kappa - k \quad (22)$$

$$\omega_1^2 = \frac{k}{m} \quad (23) \quad \omega_2^2 = \frac{k+2\kappa}{m} \quad (24)$$

We get two solutions for the eigenvalue ω^2 , which we can plug back into our matrix from equation 14 to numerically solve for the corresponding motions.

$$\begin{bmatrix} -m\left(\frac{k}{m}\right) + k + \kappa & -\kappa \\ -\kappa & -m\left(\frac{k}{m}\right) + k + \kappa \end{bmatrix} \quad (25) \quad \begin{bmatrix} -m\left(\frac{k+2\kappa}{m}\right) + k + \kappa & -\kappa \\ -\kappa & -m\left(\frac{k+2\kappa}{m}\right) + k + \kappa \end{bmatrix} \quad (26)$$

$$\begin{bmatrix} \kappa & -\kappa \\ -\kappa & \kappa \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (27) \quad \begin{bmatrix} -\kappa & -\kappa \\ -\kappa & -\kappa \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (28)$$

$$A_1 = A_2 \quad (29) \quad A_1 = -A_2 \quad (30)$$

Therefore, for the first mode the amplitudes A_1 and A_2 are always equal to each other, which means that the displacement for x_1 and x_2 are always the same because the masses are always moving in the same direction. For the second mode, The displacements are always opposite of each other, which means that the masses are moving in opposite directions. These movements can be encoded in the matrices:

$$\text{for } \omega_1, \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \propto \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (31) \quad \text{for } \omega_2, \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \propto \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad (32)$$

The amazing thing about these two normal modes is that **ANY** oscillatory motion can be written as a linear combination of the two base normal modes. This is because these two matrices, which govern the mass displacement patterns of each normal mode, form a basis for all oscillatory motion. The amplitude vectors are linearly dependent from each other, and therefore they form a basis for 2-D vector space. We can then describe any oscillatory motion as a linear combination of the two position solutions as shown in Equations 4 and 5. Each eigenvalue had it's own solution to the position function, so each mode incorporates both position functions. For m_1 , the first mode, the amplitude vector $[1, 1]$ describes the position functions $x_1(t)$ and $x_2(t)$. Therefore, for ω_1 , the position solutions are described by:

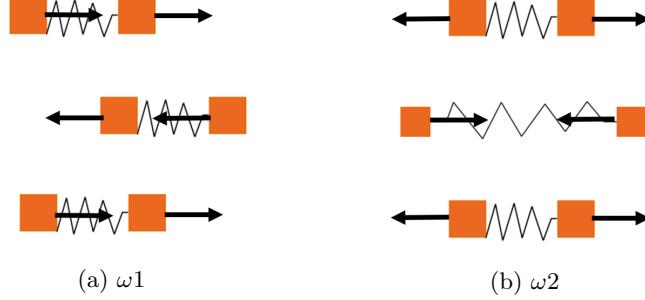


Figure 8: The normal mode solutions that serve as a basis for all oscillatory motion. This means that (a) these two solutions are orthogonal, and (b) these two basic oscillations can be combined to create *any* oscillatory motion.

$$x_1(t) = c_1 e^{i\omega_1 t} \quad (33) \quad x_2(t) = c_2 e^{i\omega_1 t} \quad (34)$$

And for m_2 with eigenvalue ω_2 , the position functions are:

$$x_1(t) = c_2 e^{i\omega_2 t} \quad (35) \quad x_2(t) = -c_2 e^{i\omega_2 t} \quad (36)$$

Because oscillatory motion has to be described by real numbers, we can take the real part of these position functions (the cosine) and take a linear combination of the of each mode to provide the most general solution for oscillatory motion. The constants C_1 and C_2 will represent the scalar used for the linear combination of the two normal modes. The general position functions can then be represented by:

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = C_1 \begin{bmatrix} 1 \\ 1 \end{bmatrix} \cos(\omega_1 t + \phi_1) + C_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} \cos(\omega_2 t + \phi_2) \quad (37)$$

Thus,

$$\begin{aligned} x_1(t) &= C_1 \cos(\omega_1 t + \phi_1) + C_2 \cos(\omega_2 t + \phi_2) \\ x_2(t) &= C_1 \cos(\omega_1 t + \phi_1) - C_2 \cos(\omega_2 t + \phi_2) \end{aligned} \quad (38)$$

To show how we can breakdown solutions of any arbitrary oscillation, let's provide a numerical example. Imagine an oscillatory motion described by the initial amplitudes as $[4, -7]$. This motion can be decomposed into the base modes $b_1 = [1, 1]$ and $b_2 = [1, -1]$. First we must normalize the matrices involved, by dividing by the square root of amplitude dotted with itself.

$$b_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \quad b_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \quad m = \begin{bmatrix} 4/\sqrt{65} \\ -7/\sqrt{65} \end{bmatrix} \quad (39)$$

There are two ways to go about solving for the weights, σ which correspond to the overlap that each base mode has with the example mode, m . We can set

up a system of equations where:

$$\sigma_1 \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} + \sigma_2 \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 4/\sqrt{65} \\ -7/\sqrt{65} \end{bmatrix} \quad (40)$$

$$\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix} = \begin{bmatrix} 4/\sqrt{65} \\ -7/\sqrt{65} \end{bmatrix} \quad (41)$$

$$\begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}^{-1} \begin{bmatrix} 4/\sqrt{65} \\ -7/\sqrt{65} \end{bmatrix} \quad (42)$$

$$\begin{bmatrix} \sigma_1 \\ \sigma_2 \end{bmatrix} = \begin{bmatrix} -0.2631 \\ 0.9648 \end{bmatrix} \quad (43)$$

Here we see that the overlap associated with the first normal mode, b_1 , is equal to -0.2631, and the second overlap is equal to 0.9648. Another way to solve for the weights is to take the dot product of the mode with each normal mode.

$$(m \cdot m)^2 = (m \cdot b_1)^2 + (m \cdot b_2)^2 = 1 \quad (44)$$

$$(m \cdot m)^2 = \sigma_1^2 + \sigma_2^2 = 1 \quad (45)$$

$$1 = \left(\begin{bmatrix} 4/\sqrt{65} \\ -7/\sqrt{65} \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \right)^2 + \left(\begin{bmatrix} 4/\sqrt{65} \\ -7/\sqrt{65} \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \right)^2 \quad (46)$$

$$1 = (-0.2631)^2 + (0.9648)^2 \quad (47)$$

$$1 = 0.0692 + 0.9308 \quad (48)$$

Here we get the same result that $\sigma_1 = -0.2631$, and $\sigma_2 = 0.9648$. The fact that the squared weights add up to 1 confirms the fact that b_1 and b_2 are a basis for any oscillatory motion.

We can check our answer by plugging in σ_1 and σ_2 into Equation 40:

$$-0.2631 \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} + 0.9648 \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 4/\sqrt{65} \\ -7/\sqrt{65} \end{bmatrix} \quad (49)$$

$$\begin{bmatrix} -0.1861 \\ -0.1861 \end{bmatrix} + \begin{bmatrix} 0.6822 \\ -0.6822 \end{bmatrix} = \begin{bmatrix} 0.4961 \\ -0.8682 \end{bmatrix} \quad (50)$$

$$\begin{bmatrix} 0.4961 \\ -0.8682 \end{bmatrix} = \begin{bmatrix} 0.4961 \\ -0.8682 \end{bmatrix} \quad (51)$$

This decomposition process serves as the fundamental basis for the decomposition of virus normal modes that are calculated. By taking the dot product of each calculated mode and the basis modes, we can see how much overlap exists between the normal modes and the SAFs

2.1.1 Eigenmodes

When solving for the eigenmodes of two masses, 2 separate motions are acquired. One eigenvalue, solved to be k/m , corresponds to the motion $[1, 1]$, indicating that the two springs move in the same direction. The second eigenmode, corresponding to a higher eigenvalue of $k + 2k/m$, oscillates at a higher frequency as its corresponding vector is $[1, -1]$, indicating the opposite direction of motion. When the two springs move towards each other, the middle spring gets compressed much faster and therefore the total force exerted is larger, leading to higher frequencies.

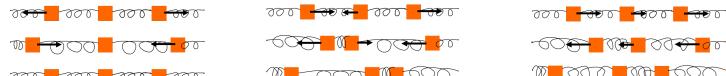
This treatment to solve normal modes can be applied to a system of 3 masses, and then extending to the continuum limit which explores a spring of infinite masses. The method to solve a 3-mass system is the same as solving the 2-mass problem, although this time the eigenmatrix is 3×3 square matrix, rather than a 2×2 .

$$\begin{bmatrix} -\omega^2 + 2\omega_0^2 & -\omega_0^2 & 0 \\ -\omega_0^2 & -\omega^2 + 2\omega_0^2 & -\omega_0^2 \\ 0 & -\omega_0^2 & -\omega^2 + 2\omega_0^2 \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (52)$$

The eigenvalues are solved for in the same fashion as for the 2-mass system, except this time we must solve quadratic equations. We then plug in the solved eigenvalue back into Equation 52 to give us the ratios between A_1 , A_2 , and A_3 .

$$\begin{aligned} \omega_1 \rightarrow \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} &\propto \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \\ \omega_2 \rightarrow \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} &\propto \begin{bmatrix} 1 \\ -\sqrt{2} \\ 1 \end{bmatrix}, \\ \omega_3 \rightarrow \begin{bmatrix} A_1 \\ A_2 \\ A_3 \end{bmatrix} &\propto \begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \end{bmatrix} \end{aligned} \quad (53)$$

These 3 different solutions describe three different normal modes



(a) ω_1 , where the two outside masses move in the same direction as each other and the middle mass stays still. (b) ω_2 , where the outside masses move in one direction, and the middle mass moves in the opposite direction. (c) ω_3 where all three masses move in the same direction while the middle mass moves in the opposite direction.

Figure 9: The normal mode solutions for a system of 3 masses on 4 springs.

As the number of masses on a spring system is increased to infinity, a new solution is introduced which is derived from the second derivative of the displacement with respect to the position. This solution, called the wave equation,

relates the partial second derivative of the displacement (ξ) with respect to time and the partial second derivative of the displacement with respect to position (x).

$$\rho \frac{\partial^2 \xi(x, t)}{\partial t^2} = E \frac{\partial^2 \xi(x, t)}{\partial x^2}. \quad (54)$$

The wave equation, which as per its name defines the propagation of waves, juxtaposes the partial second derivative of the displacement with respect to time, multiplied by the mass density (ρ) and the partial second derivative of the displacement with respect to position multiplied by energy (E).

2.1.2 Wave Equation

The wave equation becomes more complicated with the introduction to membranes, as we are increasing the number of dimensions that are used to describe the displacement. The wave equation becomes exciting and relevant to viruses when we transition to spherical membranes. The solution to the wave equation for spherical membranes can be solved via the Helmholtz equation after the technique of separation of variables was employed. The Helmholtz equation is a useful tool in physics for solving partial second derivatives. If we examine the wave equation in spherical coordinates for a spherical membrane, the solution to the wave equation includes the multiplication of spherical Bessel functions, and the Laplacian spherical harmonics, $Y_l^m(\theta, \phi)$.

$$A(r, \theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l (a_{lm} j_l(kr) + b_{lm} y_l(kr)) Y_l^m(\theta, \phi) \quad (55)$$

The first function that is multiplied by the sum, $a_{lm} j_l(kr) + b_{lm} y_l(kr)$, define the spherical Bessel functions. However, we are rather interested in the second function, $Y_l^m(\theta, \phi)$, which define the Laplacian spherical harmonics, which account for the radial displacement of virus normal modes. These spherical harmonics will be explored more in depth in Section 2.2

2.1.3 Calculations

Two different sets of virus normal modes were calculated, and they will be referred to as the Hammond and Rizzolo mode calculations. The Hammond modes were calculated by Rob Hammond[2] using a super cluster at the University of Michigan, and the Rizzolo modes were calculated by Skylar Rizzolo[3] on the Jigwe system located at Kalamazoo College.

In accordance with the Helmholtz Equation, the normal modes for viruses were solved under the assumption that viral capsids behave as if they were a thin-shelled homogeneous mass distribution. This assumption is a fundamental step in deriving the Laplacian spherical harmonics as a component to the solution to virus eigenmodes. The validity of this assumption will be assessed by the normal mode decomposition, which will indicate how much the vibrations align with the motions prescribed by the spherical harmonics.

For the Hammond calculations, calculated in Rob Hammond's SIP, the software was set to run at 8 residues per block. Residue in this case means amino acid so the software considered every 8 amino acids per protein to be a 'block' acting like a mass on a complicated spring system. The Rizzolo modes were calculated with 1 block per protein, which is a significant loss in resolution. Instead of every 8 amino acids being treated as a block, each protein itself is defined as a block with some mass.

It is suspected that this loss in resolution will affect the SAF decomposition, as lower resolution will behave in a more uniform fashion, and will not have as much dramatic radial stretching.

2.2 Spherical Harmonics

So, what exactly are the spherical harmonics, and what is their role? Spherical harmonics are the functions that are defined on the surface of a sphere, similarly to how sine and cosine are functions defined on the surface of a circle. In the same way that sine and cosine are orthogonal and form a basis for oscillating functions via a Fourier series, the spherical harmonics are also orthogonal and form a basis for functions on a sphere. The special set of spherical harmonics that arise in the solution to the wave equation are called Laplace's spherical harmonics, because they satisfy Laplace's equation.

$$\nabla^2 f = 0 \quad (56)$$

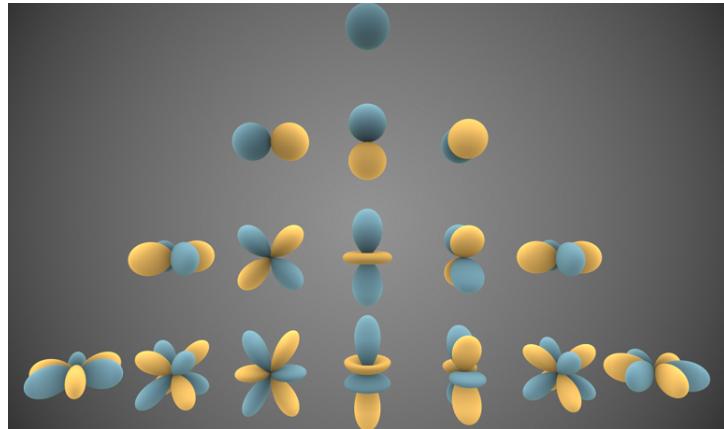


Figure 10: The real Laplacian spherical harmonics[8], denoted $Y_l^m(\theta, \phi)$, where each row corresponds with order l and columns are organized by degree m . The real Laplacian spherical harmonics do not display icosahedral symmetry, and therefore we can not simply decompose virus normal modes into the base spherical harmonics themselves.

2.2.1 Laplacian

Laplace's equation states that the divergence of the gradient ($\nabla^2 f$) is equal to 0. These functions are used notably in the field of quantum chemistry to describe electron orbitals. These spherical harmonics, however, do not display the same type of symmetry found in viruses. In order to find a set of spherical harmonics that do exhibit icosahedral symmetry, we are going to take a linear combination of the spherical harmonics of the same order l . In Figure 10, the order of a spherical harmonic is organized by row, with the first row corresponding to degree 0 and increases while descending rows. Each column is described by the degree, m . The middle column corresponds to $m = 0$, and each column increments by +1 (to the right) or -1 (to the left). For example, $Y_2^{-1}(\theta, \phi)$ represents the 2nd spherical harmonic in the 3rd row. The number of nodes increases as the order increases, and this means that the first $l = 0$ corresponds to a spherical harmonic that displays uniform motion. This by itself obeys icosahedral symmetry and corresponds with a so-called breathing mode vibration of the molecule, where its eigenmode is dominated by the uniform contracting and expanding of a virus. This is the heart of what this research is, breaking down the calculated modes into fundamental functions, giving us an objective comparison of two modes that qualitatively appear rather different.

3 Methods

3.1 Overview

In order to make things more digestible, here I provide a quick overview of the steps that are taken when building SAFs and using them to decompose virus modes. All commands and code written has been archived with Dr. Wilson.

1. Using linear combinations provided by Figure 11, SAFs are built by summing spherical harmonics of the same order but varying degree.
2. Virus objects are initialized by downloading the asymmetric unit (AU) from the VIPER database, running a few commands to extract the coordinates, then generating a virus object by calling its constructor.
3. Introduce the two different coordinate systems. The VDB coordinate system is the default orientation in which viruses downloaded from the VIPER database come in. The SAF coordinate system is the orientation that spherical harmonics have been calculated in with MATLAB. A rotation matrix will be built in order to rotate between the two orientations.
4. The default ‘cartesian to polar’ function in MATLAB outputs three parameters (θ, ϕ, r). The function that generates spherical harmonics takes the same input names, but they correspond to different angles. Figures 20 & 21 display this difference.
5. Virus SAFs are generated by outputting a radial value for each atom, which corresponds to its relative displacement. The polar coordinates of each atom are used when building the virus SAFs.
6. Virus normal modes are decomposed into these orthonormal SAFs by taking the dot product of the mode with each SAF. As demonstrated in Section 2.1, the weights of the basis vectors can be found by dotting the oscillation with the basis vectors. In this case, the SAFs represent the basis vectors and the modes are the oscillation vectors.

3.2 SAFs

In order to get icosahedrally symmetric spherical harmonics, we have to find linear combinations of harmonics of the same order whose solutions exhibit icosahedral symmetry. Luckily, this work has already been done over two decades ago by Prandl et al.

Based on the instructions provided by Figure 11, the SAF of order 6 (denoted SAF6) is composed of: Y_6^0, Y_6^{+5} , and Y_6^{-5} . The calculation and manipulation of

M\L	6	10	12	16	18	20
0	$\sqrt{11}$	$\sqrt{13 \cdot 19}$	$3 \cdot \sqrt{7 \cdot 17}$	$2^3 \cdot \sqrt{5 \cdot 19 \cdot 31}$	$\sqrt{5 \cdot 11 \cdot 17 \cdot 23}$	$\sqrt{5 \cdot 7 \cdot 23 \cdot 29}$
± 5	$\pm \sqrt{7}$	$\pm \sqrt{3 \cdot 11 \cdot 19}$	$\pm \sqrt{2 \cdot 11 \cdot 13}$	$\pm \sqrt{3 \cdot 5 \cdot 13 \cdot 17 \cdot 31}$	$\pm 2 \cdot 3 \cdot \sqrt{2 \cdot 5 \cdot 19}$	$\pm \sqrt{2 \cdot 11 \cdot 17 \cdot 19 \cdot 29}$
± 10		$\sqrt{11 \cdot 17}$	$\sqrt{3 \cdot 13 \cdot 19}$	$-\sqrt{2 \cdot 7 \cdot 17 \cdot 23 \cdot 31}$	$\sqrt{3 \cdot 7 \cdot 11 \cdot 19}$	$41 \cdot \sqrt{17 \cdot 19}$
± 15				$\pm 3 \cdot \sqrt{17 \cdot 23 \cdot 29}$	$\pm \sqrt{19 \cdot 29 \cdot 31}$	$\pm 2 \cdot \sqrt{2 \cdot 11 \cdot 19 \cdot 31}$
± 20						$\sqrt{11 \cdot 13 \cdot 31 \cdot 37}$
Norm	$\frac{1}{5}$	$\frac{1}{5^2 \cdot \sqrt{3}}$	$\frac{1}{5^2 \cdot \sqrt{5}}$	$\frac{1}{2 \cdot 5^3 \cdot \sqrt{3 \cdot 5}}$	$\frac{1}{5^3 \cdot \sqrt{5}}$	$\frac{1}{5^4 \cdot \sqrt{3}}$

M\L	22	24	26	28
0	$2 \cdot \sqrt{3 \cdot 5 \cdot 11 \cdot 19 \cdot 31 \cdot 37}$	$7 \cdot \sqrt{5 \cdot 13 \cdot 23 \cdot 29}$	$2 \cdot 3 \cdot \sqrt{3 \cdot 13 \cdot 29 \cdot 31 \cdot 41}$	$2^4 \cdot 3 \cdot \sqrt{7 \cdot 31 \cdot 37 \cdot 43}$
± 5	$\pm \sqrt{7 \cdot 13 \cdot 23 \cdot 31 \cdot 37}$	$\pm 2 \cdot 59 \cdot \sqrt{2 \cdot 3 \cdot 11}$	$\pm 23 \cdot \sqrt{5 \cdot 7 \cdot 11 \cdot 23 \cdot 41}$	$\pm \sqrt{5 \cdot 11 \cdot 13 \cdot 29 \cdot 37 \cdot 43}$
± 10	$-\sqrt{2 \cdot 7 \cdot 17 \cdot 23 \cdot 29 \cdot 37}$	$19 \cdot \sqrt{2 \cdot 19 \cdot 31}$	$2^3 \cdot \sqrt{2 \cdot 5 \cdot 7 \cdot 19 \cdot 23 \cdot 41}$	$-\sqrt{2 \cdot 3 \cdot 5 \cdot 13 \cdot 17 \cdot 23 \cdot 29 \cdot 43}$
± 15	$\pm 103 \cdot \sqrt{23 \cdot 29}$	$\pm 2 \cdot 3 \cdot \sqrt{11 \cdot 31 \cdot 37}$	$\pm 139 \cdot \sqrt{3 \cdot 5 \cdot 23 \cdot 37}$	$\pm 29 \cdot \sqrt{2 \cdot 5 \cdot 23 \cdot 29 \cdot 41}$
± 20	$\sqrt{13 \cdot 19 \cdot 23 \cdot 29 \cdot 41}$	$\sqrt{31 \cdot 37 \cdot 41 \cdot 43}$	$-89 \cdot \sqrt{5 \cdot 37 \cdot 43}$	$-2^2 \cdot 3 \cdot \sqrt{2 \cdot 5 \cdot 13 \cdot 29 \cdot 41 \cdot 47}$
± 25			$\pm 7 \cdot \sqrt{2 \cdot 17 \cdot 37 \cdot 43 \cdot 47}$	$\pm \sqrt{7 \cdot 17 \cdot 29 \cdot 41 \cdot 47 \cdot 53}$
Norm	$\frac{1}{5^5 \cdot \sqrt{2 \cdot 3}}$	$\frac{1}{5^3}$	$\frac{1}{5^6 \cdot \sqrt{2 \cdot 3}}$	$\frac{1}{2 \cdot 5^6 \cdot \sqrt{3}}$

M\L	30 - 1	30 - 2
0	$3 \cdot 12251 \cdot \sqrt{11 \cdot 13 \cdot 23 \cdot 29}$	$2^3 \cdot \sqrt{17 \cdot 19 \cdot 31 \cdot 37 \cdot 41 \cdot 43 \cdot 47 \cdot 53 \cdot 59}$
± 5	$\pm 2 \cdot 5 \cdot 4639 \cdot \sqrt{3 \cdot 17 \cdot 23 \cdot 31}$	$\pm 2 \cdot 3 \cdot \sqrt{3 \cdot 11 \cdot 13 \cdot 19 \cdot 29 \cdot 37 \cdot 41 \cdot 43 \cdot 47 \cdot 53 \cdot 59}$
± 10	$2^3 \cdot 17 \cdot \sqrt{5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 31 \cdot 37}$	$3^2 \cdot 41 \cdot \sqrt{5 \cdot 7 \cdot 23 \cdot 29 \cdot 41 \cdot 43 \cdot 47 \cdot 53 \cdot 59}$
± 15	$\pm 241 \cdot \sqrt{3 \cdot 5 \cdot 13 \cdot 31 \cdot 37 \cdot 41 \cdot 43}$	$\pm 2^4 \cdot 31 \cdot \sqrt{3 \cdot 5 \cdot 11 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 47 \cdot 53 \cdot 59}$
± 20	$2 \cdot 5 \cdot \sqrt{2 \cdot 7 \cdot 11 \cdot 23 \cdot 31 \cdot 37 \cdot 41 \cdot 43 \cdot 47}$	$3 \cdot 2161 \cdot \sqrt{2 \cdot 7 \cdot 13 \cdot 17 \cdot 19 \cdot 29 \cdot 53 \cdot 59}$
± 25	$\pm \sqrt{2 \cdot 3 \cdot 13 \cdot 17 \cdot 23 \cdot 31 \cdot 37 \cdot 41 \cdot 43 \cdot 47 \cdot 53}$	$\pm 2 \cdot 101 \cdot 151 \cdot \sqrt{2 \cdot 3 \cdot 11 \cdot 19 \cdot 29 \cdot 59}$
± 30		$71 \cdot 233 \cdot 4793 \cdot \sqrt{7 \cdot 11}$
Norm	$\frac{1}{5^5 \cdot \sqrt{71 \cdot 233 \cdot 4793}}$	$\frac{1}{2 \cdot 5^7 \cdot \sqrt{3 \cdot 71 \cdot 233 \cdot 4793}}$

Figure 11: A summary of the linear combinations for even Symmetry Adapted Functions[1]. The headers of each column tell us the order (L) of the SAF, and the row headers tell us the corresponding degree (M) of the Laplacian spherical harmonic to be added. The positive and negative coefficients are arranged so that the sign that appears on top corresponds to the positive degree, and the sign on the bottom corresponds to the negative degree. For example, we would build SAF6 by summing $\sqrt{11}(Y_6^0) - \sqrt{7}(Y_6^{+5}) + \sqrt{7}(Y_6^{-5})$

these now symmetrically adjusted harmonics allows us to describe the vibration modes of icosahedral viruses.

SAFs were generated in Matlab using a custom script ‘buildHarmonics.m,’ which builds the appropriate spherical harmonics and then linearly combines them, and then they were plotted using ‘plotSAF.m’

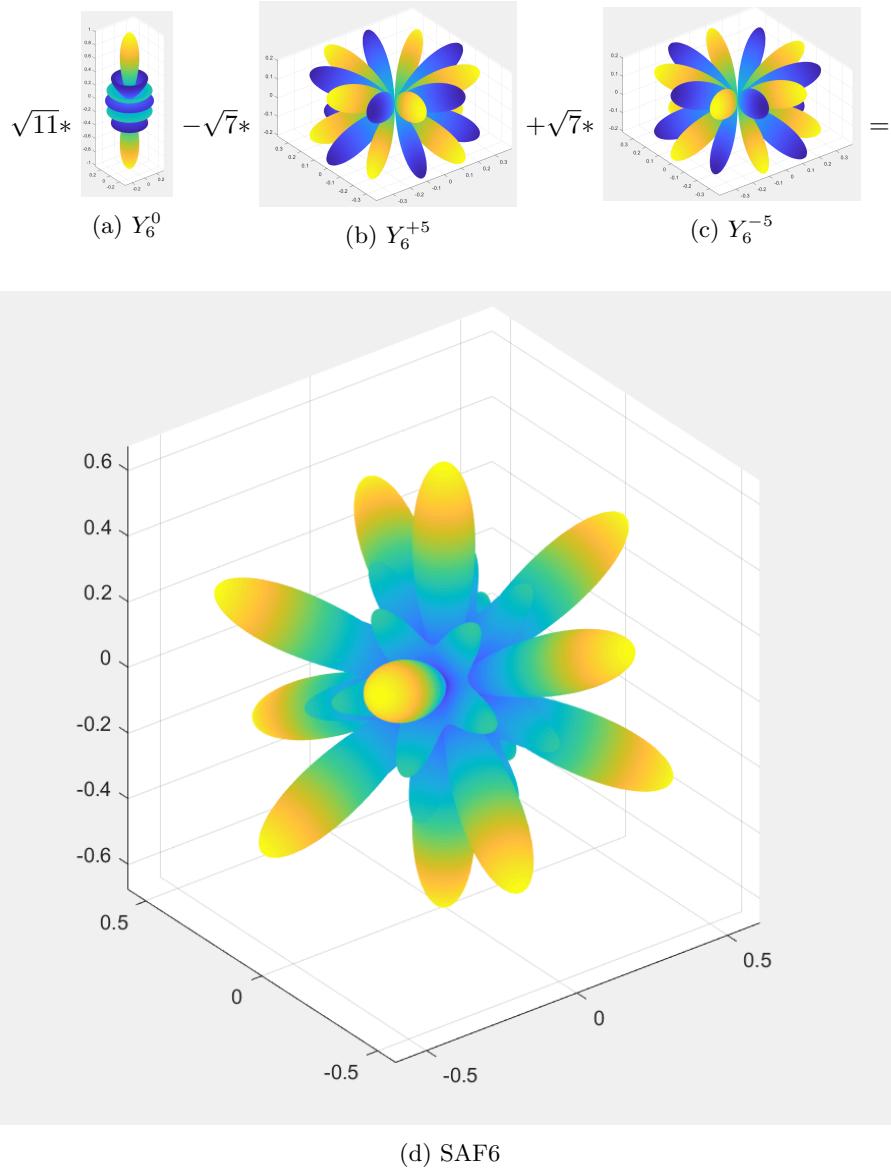


Figure 12: The linear combination to build SAF6. We observe that the 5-fold axis rising out of the z-axis is accounted for by the spherical harmonic of order 6 and degree 0. The combination of the spherical harmonics of order 6 and degree +5, -5 form the rest of the essential symmetry axes. SAF6 has the most radial displacement occur at the 5-fold axis, and has some slight oscillation about the 3-fold axis as well.

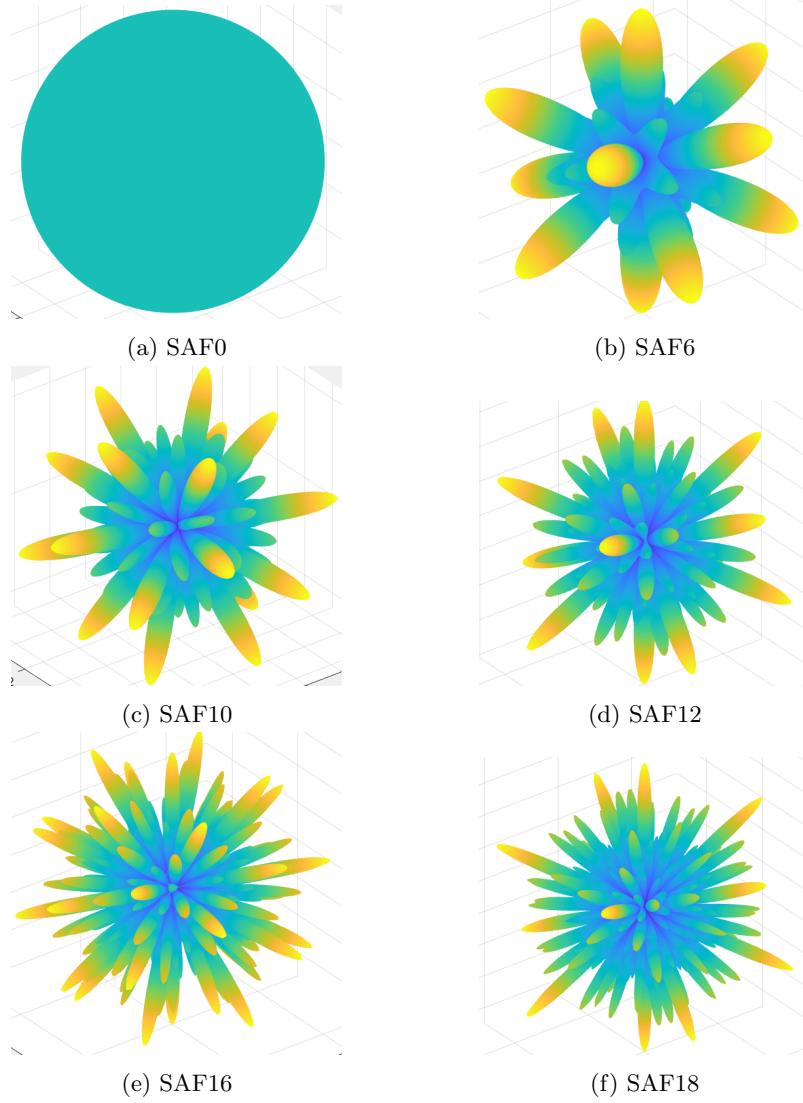


Figure 13: The first 6 SAFs. These SAFs are important to distinguish due to the frequency in which they show up in our virus normal mode decomposition. As we increase the order of the SAF, the spikiness increases. We also observe that the symmetry axis that show the highest radial displacement changes for each SAF, some being 5-fold dominant (e.g. SAF6, SAF12), and some being 3-fold dominant(e.g. SAF10).

It's important to distinguish these SAFs based on their presence in the eigen-mode decomposition. In general, only the lower order SAFs ($l \leq 18$) actually contribute to the overall virus's vibration with a large weight. This trend can

be explained by the shape and 'spikiness' of the generated SAFs, as increasing the order drastically increases how spiky the SAFs appear, which changes the strain on the system. The spikier the SAF, the more the bonds between amino acids are stretched out as the difference in radial motion of surrounding proteins is much more dramatic.

As we will see in the results section, the vast majority of the decomposition of virus's vibration modes can be accounted for by these first six SAFs. The follow up question then is: *why?* As aforementioned, it's clear from Figure 13 that as we increase the order l , the spikiness increases. The correlation between the spikiness and the vibrations that are calculated by the software has to do with *how* the normal mode software is treating the amino acids and proteins.

The lower the resolution (residue per block/protein per block), the more uniform radial motion will be present in the computation of the normal modes. When entire proteins are treated as a single unit, the more extreme radial motions prescribed by higher order SAFs are not even possible, as their differences in motion are so great that it would require individual proteins to stretch in different directions. Obviously, this is impossible when each protein is treated as a single mass. Therefore, the lower the resolution, the less representation we expect to see for the higher order SAFs.

We can analyze and classify the SAFs by looking at where the highest radial displacement takes place. In Figure 13a, SAF0 is defined by uniform radial displacement, and vibrations that follow this motion will be described as a breathing mode, because the motion is a series of contractions and expansions. the next SAF, as shown in Figure 13b, has order 6 and has max radial displacement around the 5-fold axes. SAF10 has some radial motion at and around the 5-fold axes, but the most prominent displacements take place at the 3-fold axes. SAF12 shows similarities to SAF6, as they both have max radial displacement around the 5-fold axes. As follows the trend, SAF12 is spikier than SAF6 even if they are pulsing at the same symmetry axis. SAF16 is curious because it displays large radial displacement at both the 5-fold axes and 3-fold axes, although it appears that the 3-fold relative displacement is slightly larger than the 5-fold. SAF18 has the opposite behavior, as there is dominant 5-fold extension and some 3-fold displacement.

Table 1: SAF Modes Have Max Radial Displacement About Symmetry Axes

SAF	max Δr
0	uniform
6	5-fold
10	3-fold
12	5-fold
16	5-fold, 3-fold
18	5-fold

The displacement around symmetry axes may be affected by the T-number of the virus, as different T-numbers have different structural features that line up along the symmetry axes.

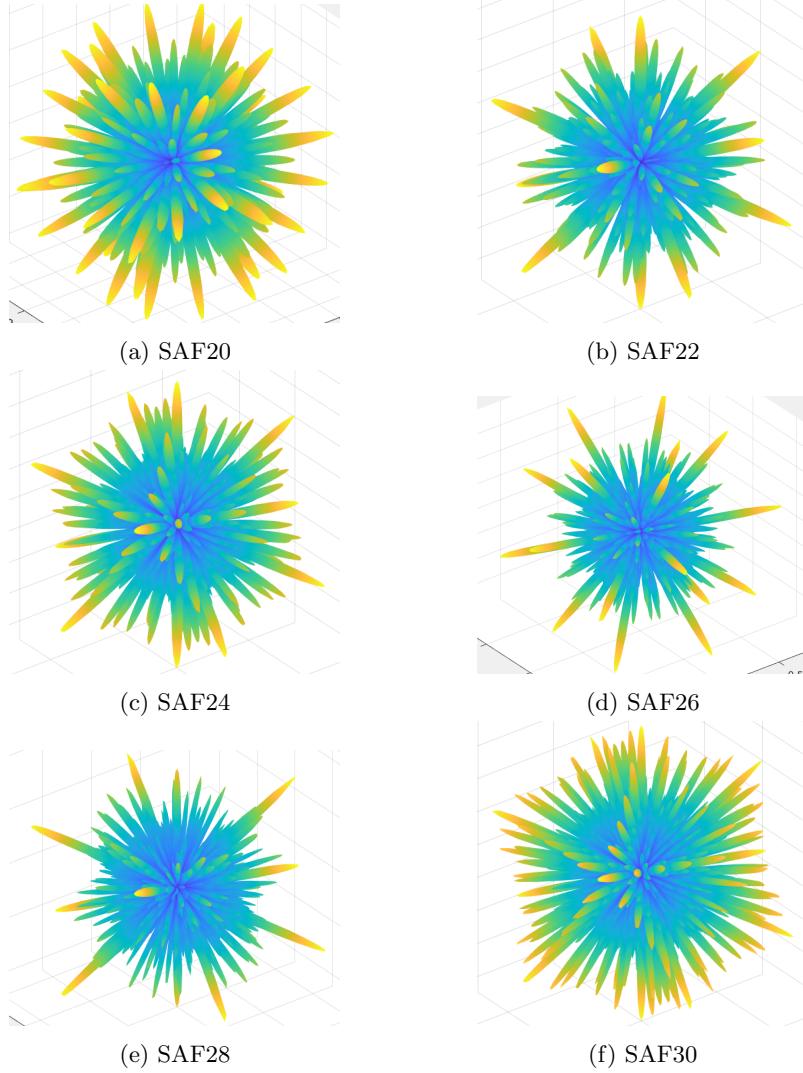


Figure 14: The next 6 SAFs. These SAFs showed up much less often during the decomposition of virus normal modes due to their sporadic, spiky shape. The larger the incline for each peak, the more opposite motion that occurs in a very limited space. This opposing motion translates to large protein deformations in the case of real viruses. Because the resolution at which the normal modes were calculated was not terribly high, we observe much more relaxed movement patterns.

These next 6 SAFs describe much more sporadic radial displacements, and show up much less during the decomposition of virus vibrations than the pre-

vious 6 SAFs. This can be explained by the incredible amount of strain and protein deformation that is predicted by these higher order SAFs.

3.3 Initialization

In order to begin the decomposition of a virus's calculated eigenmodes into individual SAFs, we need to first calculate the predicted radial movement of each atom if it were to follow an SAF. Importing the coordinates for a specific virus's asymmetric unit(AU) from the Virus Particle Explorer database (VIPERdb) into MATLAB as a user-defined virus object allowed me to manipulate the calculations for a large number of viruses. If I wanted to calculate the SAF modes for the full capsid, commands written by Dr. Wilson were used to build a full icosahedral shell from the AU. Instantiation of a virus object calculated a number of important characteristics facilitating the decomposition of its eigenmodes. When fed cartesian coordinates, spherical coordinates were automatically calculated by MATLAB and denoted as (θ, ϕ, R) , where θ denotes the azimuth, the angular displacement from the positive x-axis; ϕ denotes the elevation, the angular displacement from the x-y plane; and R evidently denotes the radius.

3.3.1 Initialization example

In order to instantiate a virus object (defined in the classdef virus m-file) in MatLab, we need the virus's atom coordinates, which can be found by downloading the pdb file from VIPERdb. The virus can be located with its pdb id, provided in Table 1 located in the results section. Once you are on the virus's particle explorer page, we need to download the virus's asymmetric unit (AU).

The screenshot shows the VIPERdb interface for the Canine Panleukopenia Virus (1c8d). The main page displays the virus's structure and key parameters. A red box highlights the 'Download' section, specifically the 'Download VIPER (AU) coordinates (52 kB)' link. The page includes tabs for 'Biodata', 'Illustrations', 'Fold-a-Virus', '3D View', 'CapsidMaps (d4 Explorer)', 'Annotations', 'MSA', and 'Related Viruses'. On the right, there is a 3D 'Surface View' of the virus and a 'Primary Sequence' section. At the bottom, there are links to SCOPe, CATH, CASTP, and MMDB.

Figure 15: An example downloading the AU of Canine Panleukopenia Virus, pdb id: 1c8d, from the VIRUS database.

This will download a .vdb file, and we can go ahead and simply change the file extension to .pdb. This will allow us to run Dr. Wilson's makeicos.pl

Perl script which will perform the 60 rotation operations on the AU to build a full icosahedral capsid. We then want to run the extract_coords.pl on the full_pdbid.pdb file that was created in order to easily import just the XYZ coordinates into MatLab.

```
perl makeicos.pl 1c8d.pdb
perl extract_coords.pl full_1c8d.pdb
```

These commands will give us a pdb file named xyz.full_1c8d. We want to open up MatLab and import this file into an array whose size is # atoms x 3, and then input this array into the constructor of our virus to create a new virus.

```
XYZ = importdata('xyz.full_1c8d.pdb')
```

We can then finally instantiate a virus by typing in the following to MatLab:

```
v1c8d = virus('1c8d', 'vdb', XYZ)
```

It's important to let the virus know which coordinate system that you inputted the coordinates in ('saf' or 'vdb'). These two coordinate systems will be discussed in more detail below, but the 'vdb' coordinate system is the default format provided by the virus AU files through the VIPERdb, and the 'saf' system is aligned with SAFs default orientation. Before calculating the SAFs for each virus coordinate, its vdb coordinates will be properly rotated.

W	x	y	z	V	x	y	z	U	k	x	y	z
1	0	1	ϕ	1	$1/\phi$	0	ϕ	1	1	0	0	ϕ
2	0	-1	ϕ	2	$-1/\phi$	0	ϕ	2	$1/2$	ϕ	1	ϕ^2
3	ϕ	0	1	3	1	1	1	3	$1/2$	$-\phi$	-1	ϕ^2
4	$-\phi$	0	1	4	-1	-1	1	4	$1/2$	ϕ	-1	ϕ^2
5	1	ϕ	0	5	1	-1	1	5	$1/2$	$-\phi$	1	ϕ^2
6	-1	$-\phi$	0	6	-1	1	1	6	$1/2$	1	ϕ^2	ϕ
7	1	$-\phi$	0	7	0	ϕ	$1/\phi$	7	$1/2$	-1	$-\phi^2$	ϕ
8	-1	ϕ	0	8	0	$-\phi$	$1/\phi$	8	$1/2$	-1	ϕ^2	ϕ
9	ϕ	0	-1	9	ϕ	$1/\phi$	0	9	$1/2$	1	$-\phi^2$	ϕ
10	$-\phi$	0	-1	10	$-\phi$	$-1/\phi$	0	10	$1/2$	$-\phi^2$	ϕ	1
11	0	1	$-\phi$	11	ϕ	$-1/\phi$	0	11	$1/2$	ϕ^2	$-\phi$	1
12	0	-1	$-\phi$	12	$-\phi$	$1/\phi$	0	12	$1/2$	$-\phi^2$	$-\phi$	1
				13	0	ϕ	$-1/\phi$	13	$1/2$	ϕ^2	ϕ	1
				14	0	$-\phi$	$-1/\phi$	14	1	0	ϕ	0
				15	1	1	-1	15	1	0	$-\phi$	0
				16	-1	-1	-1	16	1	ϕ	0	0
				17	1	-1	-1	17	1	$-\phi$	0	0
				18	-1	1	-1	18	$1/2$	ϕ^2	ϕ	-1
				19	$1/\phi$	0	$-\phi$	19	$1/2$	$-\phi^2$	$-\phi$	-1
				20	$-1/\phi$	0	$-\phi$	20	$1/2$	ϕ^2	$-\phi$	-1
				21	$1/\phi$	$-\phi^2$	ϕ	21	$1/2$	$-\phi^2$	ϕ	-1
				22	$1/\phi$	1	ϕ^2	22	$1/2$	1	ϕ^2	$-\phi$
				23	$1/\phi$	-1	$-\phi^2$	23	$1/2$	-1	$-\phi^2$	$-\phi$
				24	$1/\phi$	-1	ϕ^2	24	$1/2$	-1	ϕ^2	$-\phi$
				25	$1/\phi$	1	$-\phi^2$	25	$1/2$	1	$-\phi^2$	$-\phi$
				26	$1/\phi$	ϕ	1	26	$1/2$	ϕ	1	$-\phi^2$
				27	$1/\phi$	$-\phi$	-1	27	$1/2$	$-\phi$	-1	$-\phi^2$
				28	$1/\phi$	$-\phi$	1	28	$1/2$	$-\phi$	1	$-\phi^2$
				29	$1/\phi$	ϕ	-1	29	$1/2$	ϕ	-1	$-\phi^2$
				30	1	0	0	30	1	0	0	$-\phi$

Figure 16: VDB coordinate system, where W indicates the 5-fold axes, V indicates the 3-fold axes, and U designates the 2-fold axes

The instantiating of a virus object carries out several important steps in order to decompose the Normal Modes. The first thing that it does is to call the `installModes` function, which searches through the `modes` folder for new (Rizzolo) or old (Hammond) modes for the virus you’re building, and it places them into cells called `NEW_MODES` and `OLD_MODES`, respectively. These modes are naturally provided in ‘vdb’ coordinates, so the instantiating process next rotates coordinates from ‘vdb’ to ‘saf’, in preparation for the SAF calculations. The virus also generates spherical coordinates from the Cartesian coordinates provided.

3.4 SAFs Calculations and Coordinate Systems

3.4.1 SAF Coordinates

As previously discussed, spherical harmonics themselves are not icosahedrally symmetrical, therefore we need to calculate the symmetrically adjusted spherical functions (SAFs) that exhibit icosahedral symmetry. There are 13 even SAFs that have been calculated by Prandl et al, and they are formed by taking a linear combination of Laplace's spherical harmonics for a given order where the coefficients and the degree are given in Figure 11.

Let's take a look at SAF of order 6, which is constructed by adding $\sqrt{11}(Y_6^0) - \sqrt{7}(Y_6^5) + \sqrt{7}(Y_6^{-5})$

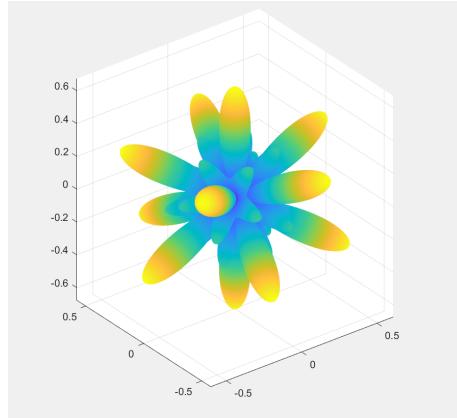


Figure 17: SAF6, the 2nd saf that exhibits icosahedral symmetry (The first is the perfect sphere, order = 0)

It's important to compare the spatial arrangement of the calculated SAFs, and the virus objects themselves. We observe that the 5-fold axis for SAF6 lies along the z-axis, the 2-fold axis is along the y-axis, and the 3-fold axis lies somewhere in between the 5-fold and 2-fold.

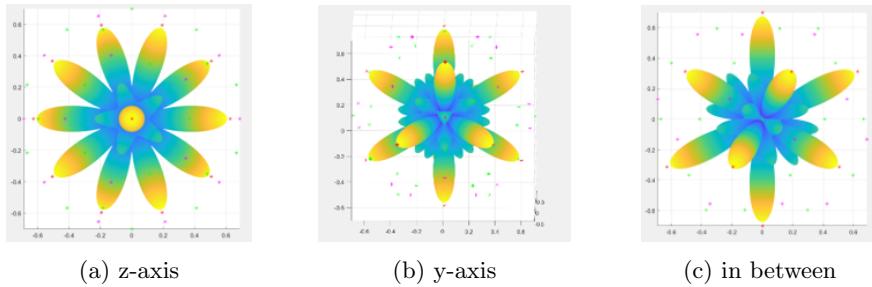


Figure 18: SAF6 shows icosahedral symmetry. (From left to right) 5-fold on the z-axis, 2-fold on the y-axis, and 3-fold axis somewhere in between

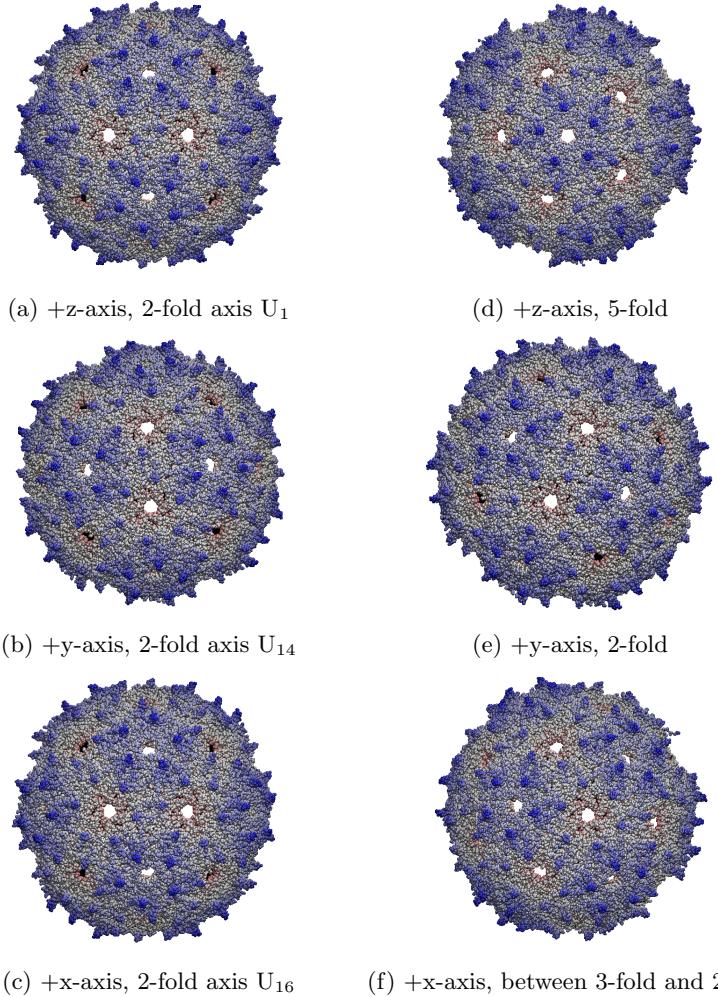


Figure 19: (on the left) A virus in the standard vdb orientation, whose symmetry axes are defined by the vectors W , V , and U . The x , y , and z axes line up with three different 2-fold axes in vdb orientation. (on the right) A virus that has been rotated to the saf orientation. As shown in Figure 18, the z -axis lines up with the 5-fold symmetry axis, the y -axis corresponds to a 2-fold axis, and the x -axis is found between a 3-fold and 2-fold axis.

However, the viper data base coordinates for a given virus follow the coordinate system of U , V , and W , where the 5-fold axis is the first W coordinate.

$$W_1 = \begin{bmatrix} 0 \\ 1 \\ \phi \end{bmatrix} \quad (57)$$

We want to rotate the virus's coordinates to line up with the calculated spherical harmonics, so we want to multiply the vdb coordinates by a rotational matrix that rotates the 5-fold axis W_1 to the positive z-axis.

$$\text{SAF}_{5\text{-fold}} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad (58)$$

The rotational matrix about the x-axis can be found by solving for theta, the angle between W_1 and the z-axis, and plugging into the matrix:

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \quad (59)$$

Once this rotation matrix has been applied to the virus, the vector produced from W_1 has successfully been rotated to the z-axis. Still there are other symmetry axes on the virus that don't line up with the SAFs symmetry axes. One of the resulting vectors - which corresponds to a 2-fold axis - lands on $[1, 0, 0]$. Because the 2-fold axis for any SAF is on the y-axis (Figure 18), we have to rotate this intermediate coordinate mapping 90° about the z-axis. The resulting rotational matrix, B, is then:

$$B = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (60)$$

Thus, the complete rotational matrix to start from vdb coordinates and end up with the same symmetry axes as SAF (saf coordinates) is $A^*B = C$.

$$C = \begin{bmatrix} 0 & -0.8507 & 0.5257 \\ 1 & 0 & 0 \\ 0 & 0.5257 & 0.8507 \end{bmatrix} \quad (61)$$

The rotational matrix, C, successfully aligns virus axes and SAF peaks, as demonstrated by the presence of symmetry markers in Figure 19. Therefore, we can say that the rotational matrix C is a linear transformation from vdb to saf coordinates.

3.4.2 Virus SAF Generation

Once the coordinates have been rotated to the saf system, the SAF value at each individual atom can be properly calculated. These SAF values are calculated by first generating the appropriate spherical harmonics involved in SAF formation at every single one of the virus's atoms, in spherical coordinates. Then these different spherical harmonics of the same order l are linearly combined to give us SAF instructions at each atom.

The spherical coordinates of each atom were used in the calculation of the spherical harmonics, as the Laplace's spherical harmonics take the input of

theta and phi ($Y_l^m(\theta, \phi)$). The function used to calculate Laplace's spherical harmonics, harmonicY.m, uses the same parameters that MATLAB outputs during the translation between cartesian and polar coordinates, however θ and ϕ denote different angles. The harmonicY function denotes the azimuth angle as ϕ and treats the inclination angle as θ . The harmonicY function is called by:

```
Y = harmonicY(l, m, theta, phi)
```

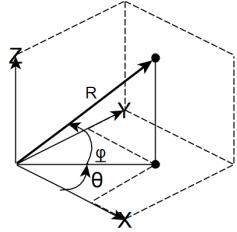


Figure 20: Polar coordinate representation in MATLAB

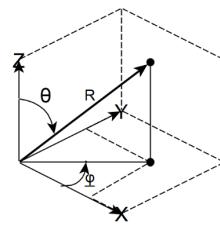


Figure 21: Polar coordinates for the input of HarmonicY

This is important because when calculating a spherical harmonic for a virus, we need to adjust the polar coordinates that are inputted into the equation.

$$\begin{aligned}\theta_H &= \frac{\pi}{2} - \phi_M \\ \phi_H &= \theta_M\end{aligned}\tag{62}$$

Where the subscript H denotes variables of the harmonicY function and subscript M denotes the variables from MATLAB's perspective

So then for each atom i , we calculate the radial displacement, $Y_l^m(\theta_i, \phi_i)$, by calling:

```
Y = harmonicY(l, m, pi/2 - phiMi, thetaMi)
```

We can then use the custom function

```
buildHarmonics.m
```

to perform the appropriate linear combination based on the SAF degree, to yield the relative radial displacement for a given atom. Each virus object has a built in function

```
virus = buildVSAF(virus)
```

Which builds a linear combination for every single SAF at every single atom coordinate. This function completes the calculation for the relative radial displacement for each atom, as defined by the icosahedrally symmetric combinations of Laplace's real spherical harmonics. In essence, the generation of SAF instructions at each atom is like generating our own normal modes, if they were to be 100% described by the SAFs. We can use these artificial modes to test for the overlap between the *actual* calculated modes.

Once the SAFs have been generated, we want to test for their orthogonality in order to corroborate the hypothesis that the SAFs form a basis for the calculated eigenmodes. We normalize the calculated SAFs before testing for orthogonality, in order to minimize the numerical error when processing the calculation for hundreds of thousands of atoms. Taking the sum of the dot products for a virus' calculated SAFs and then checking if the answer is (close) to 0 indicates orthogonality as the dot product definition includes cosine, and cosine of $\pi/2$ is 0.

3.5 Decomposition

After confirming that the SAFs are orthogonal to each other, the next step is to decompose the calculated eigenmodes themselves that indicate a virus' true motion. The calculated eigenmodes are also normalized to ensure that the maximum dot product value is 1. It's also important to note that the calculated modes were calculated in vdb coordinates, so we need to apply the rotational matrix C to each mode in order to line up in saf coordinates. To decompose the calculated modes into the simulated SAF modes to determine their influence towards the total motion, we must take the dot product of a given mode with every single SAF. The results denoted as σ indicate the amount that a given mode contributes to the eigenmode, and the σ values will be referred to as the weight associated with each SAF. For $a = \#$ of atoms, i representing the i^{th} atom, and j as the mode number,

$$\left(\sum_{i=1}^a \mathbf{M}_j \cdot \mathbf{M}_j\right)^2 = \left(\sum_{i=1}^a \mathbf{M}_j \cdot \text{SAF0}\right)^2 + \left(\sum_{i=1}^a \mathbf{M}_j \cdot \text{SAF6}\right)^2 + \left(\sum_{i=1}^a \mathbf{M}_j \cdot \text{SAF10}\right)^2 + \dots = 1 \quad (63)$$

$$\left(\sum_{i=1}^a \mathbf{M}_j \cdot \mathbf{M}_j\right)^2 = \sigma_{\text{SAF0}}^2 + \sigma_{\text{SAF6}}^2 + \sigma_{\text{SAF10}}^2 + \dots = 1 \quad (64)$$

The sum of each mode dotted with itself (which, after normalization, is equal to 1) is equal to the sum of the mode dotted with each of the basis SAFs. This decomposition provides us with overlap values, or weights, that are associated with each SAF, showing us how much a normal mode resembles the SAFs.

3.6 Visualizing

In order to qualitatively inspect the amount that a predicted SAF mode coincides with a calculated eigenmode, vmd animations were built using MATLAB.

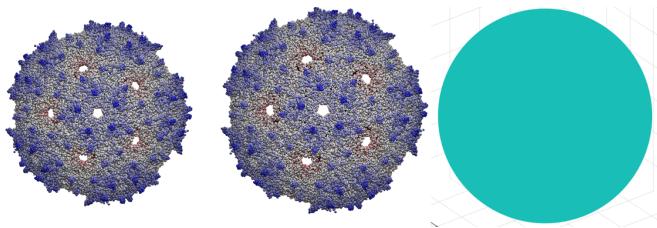


Figure 22: Artificial mode SAF0

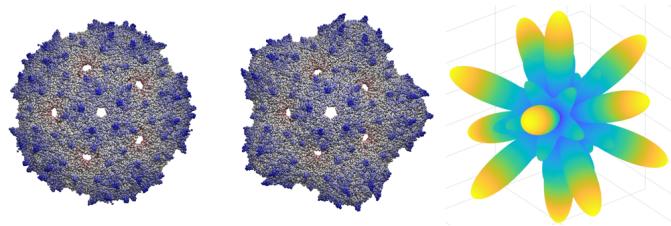


Figure 23: Artificial mode SAF6

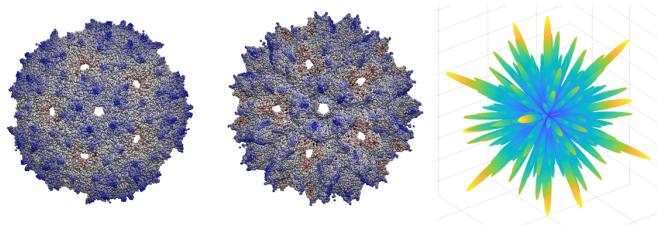


Figure 24: Artificial mode SAF22

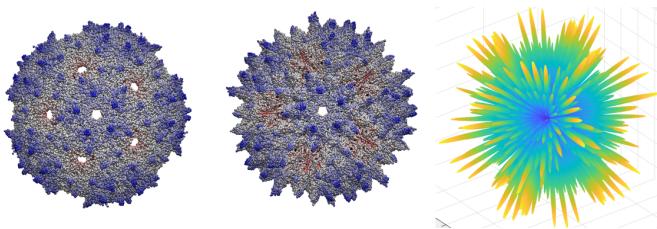


Figure 25: Artificial mode SAF30

Taking Bacteriophage MS2 for example, in order to create a vmd animation I had to add the SAF radial motion component x number of times to visualize the displacement of a virus if its mode were 100% composed of a SAF. Then, each frame stored in the format of an .xyz file was formed by displacing the original virus by 10 au, in 1 au increments. Therefore, a total of 21 frames were created by adding and subtracting SAF displacements to the virus' capsid.

Exporting a vmd command to play the movie into several windows batch files allowed for easy qualitative comparison of SAFs and eigenmodes.

Several viruses underwent this decomposition analysis, including both the full capsid and the isolated au treatment. The au treatment was carried out by building virus objects composed of only the first 1/60th atoms, and taking the first 1/60th eigenmode values, pertaining to the instructions for movement of the au. As stated before, this treatment was in an attempt to reduce the amount of numerical error present after hundreds of thousands of MATLAB calculations.

4 Results

Eigenmode data that was previously calculated by the super cluster at University of Michigan as a part of Rob Hammond's SIP was decomposed and compared with modes that were calculated this year on the Jigwe system by Skylar Rizzolo. One of the Linear algebra packages that is used in the diagonalization and calculation of eigenmatrices was updated since the modes were calculated by Rob Hammond, and this new linear algebra pack affected the calculated omega values for the same viruses. The calculated eigenmodes are also visually different, as they provide different instructions for the oscillations. It is suspected that some of the Hammond calculated modes are bunk, and in order to numerically compare the Rizzolo calculated modes to the Hammond ones, they must be decomposed into their SAFs. There were 24 total viruses whose calculated normal modes were decomposed into SAFs. Most of the viruses had 10 eigenmodes that were calculated and decomposed, but one of them, 1f8v had 20 calculated modes.

Table 2: Shows the viruses whose modes were decomposed. PBID was used to find the virus asymmetric unit on the Viper database and will be used for their identification in this paper.

PDB-ID	Full Name	T-number	# of Modes	Hammond	Rizzolo
1c8d	Canine Panleukopenia Virus	1	10	X	
1dnv	Parvovirus (Denosivirus) from <i>Galleria Mellonella</i>	1	10	X	
1dzl	Human Papilloma Virus 16 L1 Capsid	1	10	X	
1m1c	L-A Virus	1	10	X	
1m06	Bacteriophage a3	1	10	X	
1s58	B19 Parvovirus Capsid	1	10	X	
1x9p	Human Adenovirus 2 Penton Base	1	10	X	
2df7	Bursal Disease Virus VP2	1	10	X	X
3hag	Hepatitis E Virus-like Particle	1	10	X	
3ide	IPNV Subviral Particle	1	10	X	
3p0s	<i>Bombyx mori</i> densovirus	1	10	X	
4qc8	Pathogenic Bovine Parvovirus-1	1	10	X	
1auy	Turnip Yellow Mosaic Virus (TYMV)	3	10	X	X
1cwp	Cowpea Chlorotic Mottle Virus (CCMV)	3	10	X	X
1f8v	Pariacoto Virus (PAV)	3	20	X	X
1qbe	Bacteriophage Q- β Capsid (QB)	3	10	X	X
2ms2	Bacteriophage MS2 (MS2)	3	10	X	X
2tbv	Tomato Bushy Stunt Virus (TBSV)	3	10	X	
1hxs	Poliovirus (Mahoney Strain)	pT3	10		X
1ohf	Nudaurelia Capensis ω Virus	4	10	X	X
1qgt	Human Hepatitis B Viral Capsid	4	10		X
1sva	Simian Virus 40	7d	10	X	X
3j6r	Human Papillomavirus Type 16	7d	10	X	X
3esk	HK97 Prohead ii	7l	10	X	

The T-number of the virus is of particular interest during the decomposition of its eigenmodes because the higher the T-number the more the virus' capsid resembles a perfect sphere. Whether or not there is a correlation between the T-number and overlap weights of SAFs is studied.

Not all of the viruses had both Hammond and Rizzolo normal modes calculated, and therefore only 9 out of the 24 viruses were able to be compared side by side. The decompositions, however, were still performed for all viruses giving us

the tools to quickly determine which normal modes corresponded to breathing, radial, and twisting motions. One of the first viruses whose Hammond modes I decomposed, Bacteriophage MS2 (MS2), had 10 calculated eigenmodes to break down. The normal modes contain the displacement vectors for each individual atom of the virus, and the first two normal modes for MS2 encode the following motions:

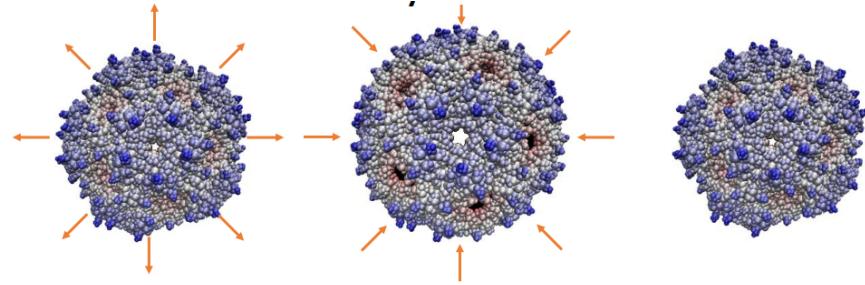


Figure 26: The first mode for Bacteriophage MS2, denoted as the breathing mode for its relatively uniform radial displacements. This real, calculated mode looks similar to the artificial mode SAF0 as shown in Figure 22. This visual observation will later be corroborated by the decomposition which shows that the first mode has large overlap with SAF0.

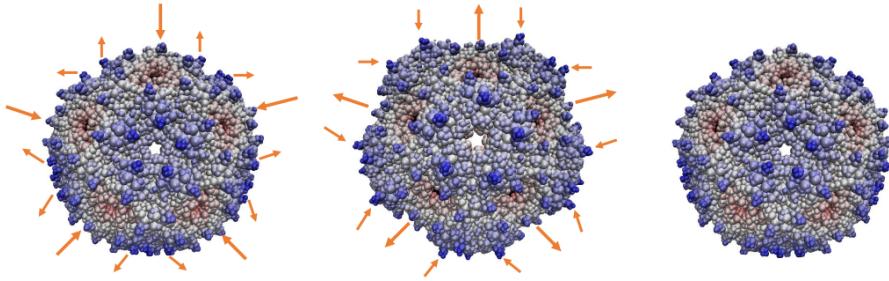


Figure 27: The second mode for MS2. This mode strongly resembles that of SAF6, as shown in Figure 23. The decomposition will show a strong overlap between the second mode for MS2 and the artificial mode predicted by SAF6, providing numerical support to what we observe.

The first 10 Hammond modes for MS2 were decomposed into the following table.

SAF\Mode	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_10
SAF0	0.95	0.29	-0.01	0.00	0.02	0.04	0.01	-0.03	0.01	-0.89
SAF6	0.16	0.91	-0.02	-0.02	0.02	0.02	0.02	-0.01	0.06	0.21
SAF10	0.34	0.12	0.02	0.00	0.80	0.04	0.10	-0.04	-0.14	0.38
SAF12	0.02	-0.13	0.02	0.02	-0.28	-0.09	0.12	-0.13	-0.81	-0.09
SAF16	-0.08	-0.17	-0.01	0.01	0.08	-0.06	0.47	0.09	0.27	-0.05
SAF18	0.08	-0.12	-0.01	0.01	-0.03	0.00	-0.15	0.15	0.16	0.07
SAF20	-0.05	0.01	-0.01	0.00	-0.03	0.00	0.05	-0.10	0.08	-0.01
SAF22	0.00	-0.06	0.01	-0.05	-0.06	0.00	-0.16	-0.05	0.05	0.02
SAF24	0.01	-0.13	-0.01	0.02	-0.02	0.03	-0.08	0.05	0.05	0.02
SAF26	0.03	0.02	0.01	-0.01	0.08	0.01	-0.12	-0.04	-0.01	0.04
SAF28	0.01	-0.07	0.00	-0.02	-0.07	0.00	-0.03	-0.04	0.05	0.00
SAF30	0.02	-0.08	0.00	0.02	0.00	0.01	0.00	0.02	0.10	0.00
SAF31	-0.08	-0.03	0.00	-0.02	-0.11	-0.01	0.08	0.03	-0.04	-0.10

Figure 28: A typical decomposition of viral eigenmodes, using MS2’s Hammond modes as an example. Mode number corresponds to ascending ω (or frequency) values. The column headers indicate the mode that is being decomposed, and the row headers correspond to the SAF mode that is contributing to the calculated mode motion. Mode 1 shows strong overlap with SAF0 ($\sigma_{SAF0} = .95$). Mode 2 shows strong overlap with SAF6 ($\sigma_{SAF6} = .91$).

The first 10 eigenmodes of Bacteriophage MS2 are dotted with the radial displacement vectors determined by each SAF and the resulting sum represents the overall contribution that an SAF has towards the virus’ normal modes. This value, henceforth referred to as the weight of an SAF, gives us the data to quickly compare different virus normal modes. For the first eigenmode in Figure 28 we see that SAF0 has a weight of 0.95 attached to it. This means that the overall motion of the calculated eigenmode greatly resembles that of SAF0, an event in which there is perfectly uniform radial displacement. SAF6, however, does not have uniform radial translation, and instead we observe peaks around the 5-fold and 3-fold axes, and neutral dividing lines around the 2-fold axes in which there is no atomic displacement. The second calculated eigenmode of 2ms2 does not show significant overlap with SAF0 but does show significant overlap with SAF6. Carrying a weight of 0.91, we expect the visualization of the second eigenmode to strongly resemble an SAF6 oscillation, which it indeed does as per Figures 23 & 27. The negative sign that appears for certain SAFs can be explained by the initial orientation of motion and has no effect on the overlap as the initial direction of oscillation is arbitrary.

After decomposing the virus eigenmodes, the modes where an SAF had a dominant weight attached were highlighted and classified by SAF. For example, in Figure 29 we see that the first eigenmode for the old calculations, M₁, is made up of mainly SAF6. This Hammond mode is comparable with the Rizzolo mode 1, which is also largely composed of SAF6. An interesting result of the decomposition of eigenmodes into spherical harmonics is the fact that the SAFs

Row	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_10	Row	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_10	
SAF6	0.04	-0.02	0.04	-0.11	0.00	0.09	-0.01	0.01	-0.03	0.17	0.00	0.02	0.00	0.03	0.01	0.00	0.12	0.00	0.03	0.17	0.02	
SAF6	0.02	0.01	-0.02	0.09	-0.01	0.01	-0.03	0.17	0.00	0.02	0.00	0.01	0.11	0.06	-0.01	-0.07	0.09	-0.01	0.23	-0.03		
SAF10	0.04	-0.07	0.27	-0.78	-0.01	0.31	-0.03	0.14	0.02	-0.09	0.01	0.09	0.22	0.67	-0.02	-0.11	-0.04	-0.21	0.09			
SAF12	-0.02	-0.04	-0.04	0.32	-0.01	0.75	-0.03	0.34	-0.03	-0.06	0.00	0.01	0.03	-0.37	0.16	0.13	0.01	0.50	0.07	0.08	0.04	
SAF16	-0.02	0.01	0.01	0.13	0.00	0.09	0.01	0.09	0.01	0.06	0.00	0.01	0.00	0.01	0.01	0.01	0.05	0.01	0.15	0.17	0.09	
SAF18	-0.03	0.10	0.03	0.02	0.00	0.09	0.11	0.09	-0.24	0.03	0.09	0.01	0.11	-0.04	0.03	0.01	0.03	0.01	0.06	0.05	0.05	
SAF20	0.03	0.03	0.00	0.02	0.00	-0.01	0.02	-0.07	0.00	0.02	0.02	0.00	0.04	-0.05	0.06	-0.08	0.01	0.00	0.01	-0.04	0.04	-0.03
SAF22	-0.06	0.05	-0.01	0.04	0.00	0.00	0.04	0.01	-0.06	0.02	0.00	0.00	0.08	-0.05	-0.06	-0.06	-0.03	0.00	0.05	-0.04	0.05	0.05
SAF24	-0.09	0.02	0.01	-0.02	0.00	0.00	-0.01	0.01	-0.01	0.00	0.00	0.00	0.08	-0.03	0.03	-0.01	0.01	-0.02	-0.04	0.08	-0.12	-0.07
SAF26	-0.09	-0.03	0.04	-0.11	0.00	-0.01	-0.01	-0.02	0.00	-0.03	0.00	0.00	-0.09	0.03	0.04	-0.04	0.07	0.01	-0.02	-0.04	0.01	0.01
SAF28	-0.06	0.01	-0.02	0.05	0.00	0.08	-0.02	-0.04	-0.02	0.01	0.01	0.00	-0.04	-0.01	0.02	-0.03	0.01	-0.03	0.01	-0.02	-0.06	
SAF30	0.01	-0.01	-0.01	0.01	0.01	-0.11	0.00	0.00	-0.03	0.01	0.01	0.01	0.03	-0.01	-0.05	-0.01	-0.08	0.02	-0.01	0.02	0.01	0.02
SAF31	0.02	-0.05	-0.01	0.07	0.00	-0.02	-0.03	-0.02	-0.04	0.00	0.00	0.04	-0.03	0.00	-0.01	-0.01	0.02	-0.02	0.00	0.00	0.00	

Figure 29: Hammond (left) and Rizzolo (right) eigenmode decompositions of Turnip Yellow Mosaic Virus (TYMV). Some modes are highlighted to show SAF dominance, and to easily identify similar patterns of motions between calculations. For example, the SAF10 dominant Hammond mode 4 corresponds to the Rizzolo SAF10 dominant mode 5.

only account for radial displacement for an oscillation. Thus, any normal mode vibration that is composed of twisting and stretching will show very little to no overlap with the spherical harmonics. This element introduces a fundamental limitation in this current project, as the methodology outlined above provides a useful but incomplete picture of a virus’ oscillations. The decomposition of old eigenmodes has brought into question the legitimacy and accuracy of Hammond’s previous calculations. We suspect that an accurate output of eigenmodes contains discrete modes that are largely composed of individual SAFs. That is to say, we expect there to be one breathing mode, one SAF6 mode, one SAF10 mode, etc. The repetition of modes could be an indication of similar movement around the symmetry axes of different modes, or it could present a flaw in the previous algorithm causing the computation to output multiple degenerate modes. This is best illustrated with the decomposition of Human Papilloma Virus.

Row	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_10	Row	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_10	
SAF6	-0.21	0.25	0.03	-0.16	0.71	0.03	-0.01	-0.54	-0.13	0.06	-0.19	0.01	-0.05	0.00	0.02	-0.03	-0.03	0.01	0.56	0.03		
SAF10	0.10	0.49	-0.01	-0.49	-0.12	0.02	-0.02	0.15	0.06	0.06	0.03	0.26	-0.04	0.07	-0.01	-0.01	-0.08	0.13	0.21	-0.01		
SAF12	-0.07	-0.20	0.11	0.30	0.02	-0.02	0.15	0.06	0.06	0.03	0.03	-0.19	-0.02	-0.07	0.17	-0.03	0.07	-0.44	0.49	-0.37	-0.01	
SAF16	-0.17	0.20	-0.26	-0.21	0.21	-0.20	-0.13	0.14	-0.08	0.20	-0.03	0.23	0.01	-0.12	0.00	0.28	-0.10	0.01	0.00	0.33	0.00	
SAF18	-0.06	-0.07	-0.05	0.07	0.11	-0.02	-0.06	0.02	0.14	-0.06	0.03	0.12	-0.03	-0.04	0.05	0.06	0.16	0.00	-0.01	0.07	0.03	
SAF20	0.02	-0.12	-0.07	0.18	0.11	-0.22	-0.10	0.03	-0.01	-0.01	0.01	0.11	0.01	0.06	0.16	-0.02	0.00	0.11	-0.19	0.24	-0.01	
SAF22	-0.05	-0.05	0.03	0.11	0.04	-0.02	0.00	0.02	0.04	-0.04	0.03	-0.10	0.01	0.06	-0.15	-0.06	0.00	-0.07	0.02			
SAF24	0.09	0.08	0.00	-0.13	0.00	-0.09	0.03	0.03	-0.05	-0.10	0.07	0.13	0.00	-0.10	-0.01	-0.01	0.04	0.06	-0.02	0.04	-0.01	
SAF26	-0.08	-0.03	-0.13	0.00	0.03	-0.11	0.06	0.12	0.00	-0.09	0.17	0.04	0.00	-0.02	-0.09	0.13	0.07	0.02	-0.01	0.08	0.01	
SAF28	0.17	-0.08	0.16	0.05	-0.18	0.14	0.12	-0.10	-0.05	0.17	0.10	0.00	0.01	-0.13	-0.18	0.02	0.00	0.02	-0.19			
SAF30	0.07	-0.09	0.05	0.05	-0.07	0.05	0.08	-0.01	0.01	0.06	-0.11	-0.01	0.01	0.04	-0.11	-0.06	-0.01	0.04	-0.08	-0.07	0.00	
SAF31	-0.13	-0.08	-0.14	0.10	0.11	-0.13	0.05	0.14	0.09	-0.13	0.03	-0.11	0.01	0.03	-0.05	0.15	-0.01	-0.03	-0.01	0.06	0.00	

Figure 30: The SAF decomposition of Human Papilloma Virus Type 16. Hammond modes are on the left and Rizzolo modes appear on the right. The Hammond modes outputs a variety of different mode frequencies that have strong overlap with SAF0. This indicates that they are all breathing modes with fairly uniform radial displacements. The Rizzolo modes show a wider distribution of SAF dominance.

The Hammond calculated eigenmode spits out five normal modes that can be classified as breathing modes, whereas the Rizzolo calculated modes decompose prominently into four different SAFs. This indicates that multiple degenerate eigenmodes were not properly filtered out during the calculation of the Hammond modes. In order to begin the comparison of Hammond calculations

to Rizzolo calculations, modes were separated and defined by the most dominant SAF contributor. For example, in Figure 30, the Hammond modes of M_1, M_3, M_6, M_8 , and M_{10} all have SAF0 contribute the most to the eigenmode motion. However, we see much more diversity in the Rizzolo mode calculations, with M_5 being influenced by SAF0, M_4 by SAF6, M_1 by SAF10, and finally M_8 corresponding to SAF12. Normal modes were assigned this classification to compare modes that represent the same motion rather than modes that have the same omega ranking, as in this case Hammond M_1 (SAF0 dominant) clearly does not match up with Rizzolo M_1 (SAF10 dominant).

4.1 Virus Analysis

4.1.1 Turnip Yellow Mosaic Virus (1auy, T = 3)

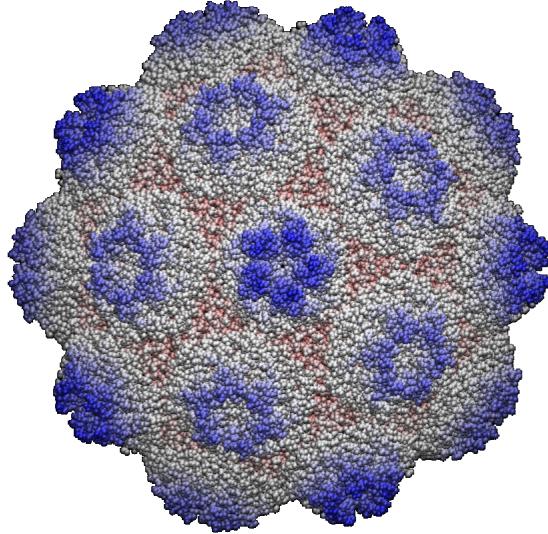


Figure 31: The viral capsid of Turnip Yellow Mosaic Virus (TYMV), viewed from the 5-fold axis. The protruding features are evenly distributed about the protein shell.

In order to begin the analysis and comparison of the Hammond and Rizzolo modes, hereby noted as HM_i and RM_i , respectively, we first need to examine which SAFs exhibit dominant weights for each of the modes. As seen in Figure 29, the highlighted values show which SAF contributes the most to the overall motion of each mode. Defining the modes in this fashion, as opposed to only looking at the mode number, allow us to quickly analyze which *motions* line up with each other instead of looking at an arbitrary order spit out by the calculation software. In the case of Figure 32, we see that HM_2 (shown in dark red) has SAF0 (as indicated by the x-axis) contributing the largest influence on the normal mode, with an associated weight of just under 1. The decomposition for HM_1 involves SAF6 contributing the most, HM_4 is associated with SAF10, HM_6 lines up with SAF12, and finally HM_8 shows SAF16 having the highest influence. Each SAF has a color assigned to it so we can easily find which mode number is associated with that SAF. For example, in Figure 32 we see that SAF0 is dark red, SAF6 is light red, SAF10 is orange, SAF12 is yellow, and SAF16 is green. The modes which are grey or black is used to indicate that there is no dominant SAF contributing to the overall mode motion - this indicates to us that these modes have almost no radial motion and are composed

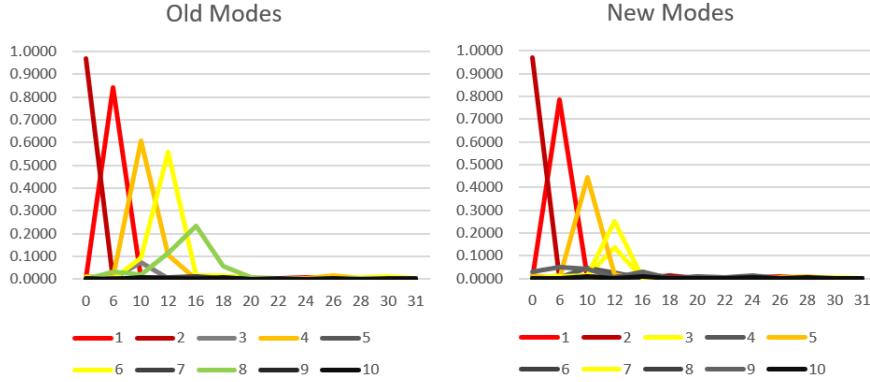


Figure 32: Comparison of the weights associated with each SAF for the Hammond mode computations (left) and the Rizzolo mode computations (right). SAF dominant modes of the same order are assigned the same color so that we can easily identify which mode number correspond to similar motions.

entirely of twisting and rotating. Each of these graphs are extrapolated from the decomposition data as seen in Figure 29 and 30.

So how can my decomposition give us information on the “goodness” of the Hammond and Rizzolo calculations? As discussed above in reference to Figure 29 and 30, we expect legitimate mode calculations to be able to be decomposed into an even distribution of SAF dominate modes. For example, we see in Figure 32 that there is an even distribution of SAF influence for the Hammond modes. This cascading effect shows us that there is only one unique mode that is largely influenced by each SAF, and this lack of overlap proves that there are not multiple, degenerate modes being spit out by the software.

Both the Hammond mode and Rizzolo mode calculations meet this criteria of having unique, non-degenerate modes being calculated. The only discrepancy arises for the Rizzolo modes, in which we have two modes, both RM_3 and RM_7 , that can largely be described by SAF12. Whether these two modes represent the same, degenerate motion is tough to say without visually analyzing the two motions, but the weights would indicate that each eigenmode, RM_3 and RM_7 , have core structural similarity in their vibrations.

Another interesting discrepancy between the Hammond modes and the Rizzolo modes has to do with the value of the weights associated with the SAFs of order > 6 . The Rizzolo calculated modes, which were computed at a lower resolution, showed less overlap with SAF10 and SAF12 compared to the overlaps for the Hammond modes. This result is consistent with the hypothesis that lower resolution will show less agreement with the “spikier”, higher order SAFs

Table 3: SAF Dominance for Hammond and Rizzolo Modes for Turnip Yellow Mosaic Virus

SAF	HM_i	RM_i
0	2	2
6	1	1
10	4	5
12	6	3,7
16	8	-
18	-	-

4.1.2 Cowpea Chlorotic Mottle Virus (1cwp, T = 3)

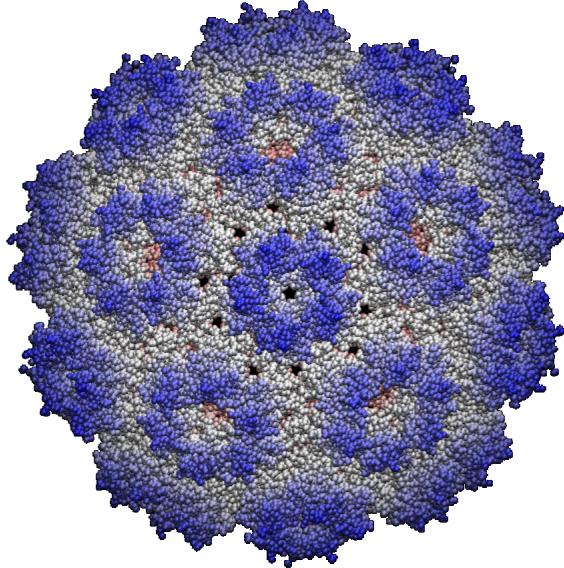


Figure 33: The viral capsid of Cowpea Chlorotic Mottle Virus (CCMV), viewed from the 5-fold axis. The arrangement of protrusions is very similar to that of TYMV, but it appears that the proteins in CCMV are not as tightly packed around the symmetry axes as the proteins in TYMV.

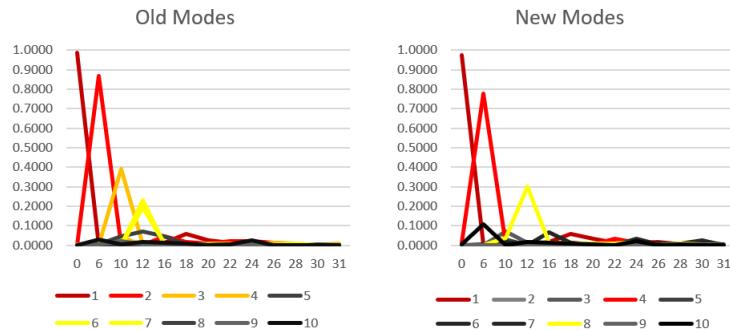


Figure 34: Comparison of the weights associated with each SAF for the Hammond mode computations (left) and the Rizzolo mode computations (right) for CCMV. Background modes that show very little SAF overlap are color coded with either grey or black to show their insignificance during normal mode decomposition.

Cowpea Chlorotic Mottle Virus, another T=3 virus, shows similar mode decomposition patterns to those of Turnip Yellow Mosaic Virus. We see that there is clearly a SAF0 dominant breathing mode for each calculation (HM_1 and RM_1), and the Hammond modes show a clear distribution of SAF dominance up until SAF12. Both of the breathing modes show a very high σ value, which is a reflection of the uniformity of the capsid, and how much the structure is similar. In the case of the Hammond modes, we see slightly more degenerate overlap than we would expect based on the results for TYMV. We see that there is overlap for both SAF10 and SAF12, with $HM_{3,4}$ being influenced by SAF10 and $HM_{6,7}$ being categorized by SAF12.

On the contrary, we see absolutely no degenerate overlap calculated by the Rizzolo modes. This indicates that each mode accounts for unique vibrations. However, we see that there is a only small number of SAF dominate motions being calculated, only managing to account for SAF0, SAF6, and SAF12. There is a little bit of background noise that arises, but no weights that are sufficiently large to be able to categorize the motions as SAF dominant.

Table 4: SAF Dominance for Hammond and Rizzolo Modes for Cowpea Chlorotic Mosaic Virus

SAF	HM_i	RM_i
0	1	1
6	2	4
10	3,4	-
12	6,7	8
16	-	-
18	-	-

The summary table allows us to quickly see which mode numbers correspond to each other in terms of their SAF dominance. As we can see, the mode number does not necessarily correspond to the motion that it prescribes, and my decomposition allows us to quickly describe and compare that motion.

The absence of a SAF10 dominant mode for the Rizzolo mode calculations indicates that there is minimal movement around the 3-fold axes, given that SAF10 has maximum radial displacement around the 3-fold axes. If we look at the structure of 1cwp as seen in Figure 33, the hexamers find themselves at the 3-fold axes as per Figure 5. The absence of a SAF10 dominant mode could suggest that the hexamers are more rigid and don't oscillate as much as compared to the pentamers around the 5-fold axes.

4.1.3 Pariacoto Virus (1f8v, T = 3)

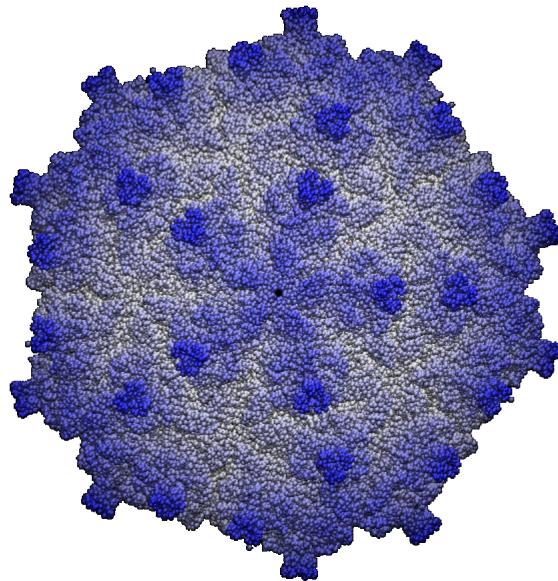


Figure 35: Viral capsid of Pariacoto Virus (PAV), viewed from the 5-fold axis. This virus, while also $T = 3$, shows very different protrusion patterns than the previous two viruses. Most notably is the lack of an enclosed protein pentagon around the 5-fold axes.

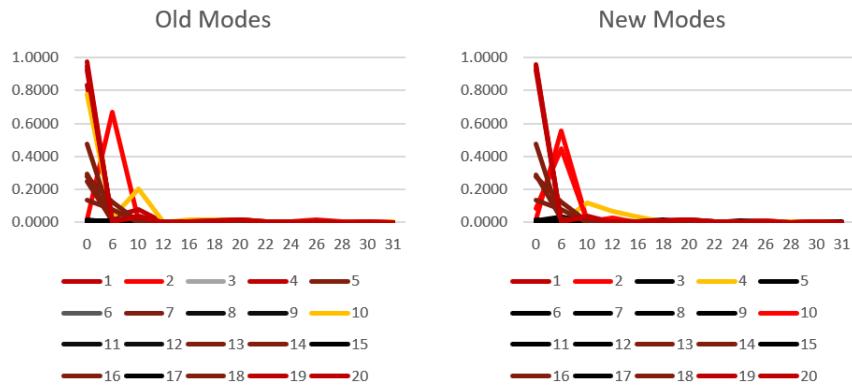


Figure 36: Comparison of the weights associated with each SAF for the Hammond mode computations (left) and the Rizzolo mode computations (right) for PAV. PAV had 20 normal modes that were decomposed, as opposed to 10.

The Pariocoto Virus, yet another T=3 virus whose modes were decomposed, showed completely different decomposition patterns than its predecessors. Right off the bat, the most striking difference is the number of modes that were calculated, and the sheer concentration of SAF dependency on lower order SAFs. We see that all of the SAF dependant modes are concentrated within degree 10, amounting to only 3 different SAFs contributing to overall vibrations. The decomposition indicates that there are a large amount of degenerate eigenmodes being calculated, suggesting an error in the manner in which the vibrations were calculated. Examining the breathing mode, SAF0, we see that there are *ten* modes that are degenerate for this degree. $HM_1, HM_4, HM_5, HM_7, HM_{13}, HM_{14}, HM_{16}, HM_{18}, HM_{19}$, and HM_{20} all exhibit dependence on SAF0, which means they all prescribe the motion of the breathing mode. Obviously, there is an error in the way in which these modes are being produced, as we expect a unique distribution of eigenmodes.

We run into the same exact problem with the Rizzolo mode calculations, which indicates that the difference in computation method did not have an effect on the legitimacy of eigenmodes for PAV. Tight concentration around the lower SAF degrees and ridiculous degeneracy indicate that there are a wide range of normal modes that are *not* being calculated by the software. All of the further degeneracies are outlined below in a summary table.

Table 5: SAF Dominance for Hammond and Rizzolo Modes for Pariacoto Virus.

SAF	HM_i	RM_i
0	1,4,5,7,13,14,16,18,19,20	1,13,14,16,18,19,20
6	2	2,10
10	10	4
12	-	-
16	-	-
18	-	-

The presence of one or two degenerate modes does not necessarily indicate a problem with the calculation, but it is evident in both the summary table and the comparison graphs that something is wrong with how the modes are being calculated. This result is surprising given that the other T=3 viruses showed a decent spread of SAF dependence, whereas the Pariacoto Virus modes are showing tight concentration and repeated degeneracy of the breathing mode.

The failure of the normal modes to line up with large SAF influence proves that the capsid for PAV can not be accurately described as a thin-shelled homogeneous distribution. This observation is corroborated by the relative thickness of the Pariacoto Virus.

4.1.4 Nudaurelia Capensis ω Virus (1ohf, T = 4)

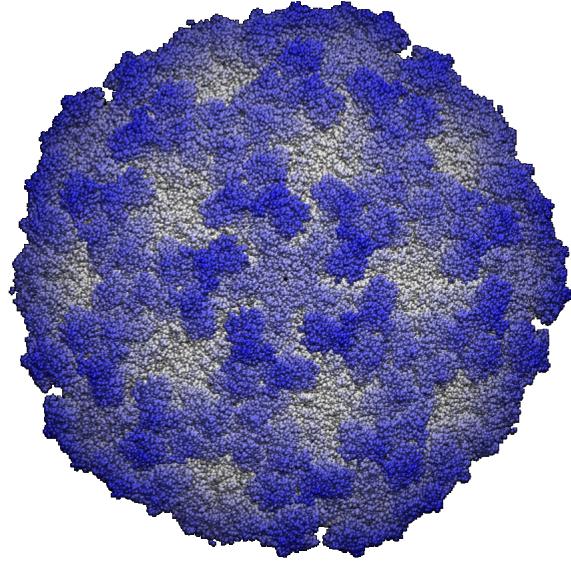


Figure 37: Nudaurelia Capensis ω Virus's (NCV) capsid. NCV is the first T=4 virus analyzed. The conformation of proteins and the orientation of symmetry axes looks similar to the previous T=3 viruses that were decomposed.

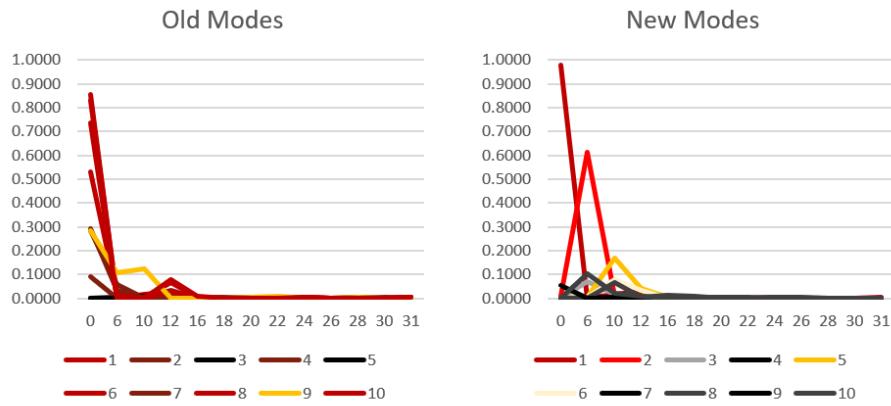


Figure 38: Comparison of the weights associated with each SAF for the Hammond mode computations (left) and the Rizzolo mode computations (right) for NCV. The Hammond modes are a great example of a poor mode distribution; the decomposition indicates that there are multiple breathing modes being calculated at different frequencies.

The decomposition of the normal modes for NCV, a T=4 virus, do not show promising results. The Hammond modes that were calculated showed a lot of overlap with SAF0, thus there were many breathing modes that were found as a solution to the normal modes. Like the previous virus, there was a lot of degenerate solutions found, and they were all very concentrated on the lower degree SAFs. Any weight beyond SAF0 was incredibly weak indicating very little overlap. The Hammond calculations failed to provide a strong case that a thin-shelled homogeneous mass distribution is an appropriate way to describe NCV.

The Rizzolo modes' decomposition painted a different picture for the accuracy of the normal mode solutions. There is a better distribution of SAF dominance, and there are no degeneracies for the breathing mode. Having only one unique solution for the breathing mode is already a good indication that there will be other modes that can be described by the spherical harmonics. As such, SAF6 and SAF10 do show good, discrete agreement with the Rizzolo modes. There is a lot of low weight background noise, which tells us that many modes incorporate twisting and rotations as a way to disperse energy.

Table 6: SAF Dominance for Hammond and Rizzolo Modes for Naudaurelia Capensis ω Virus.

SAF	HM _i	RM _i
0	1,2,4,6,7,8,10	1
6	-	2
10	9	5
12	-	-
16	-	-
18	-	-

The newer mode calculations proved that Naudaurelia Capensis ω Virus can be accurately described as a thin-shelled distribution, and that it was simply a failure on the part of the old algorithm to provide unique modes. This T=4 virus also exhibits almost no SAF10 influence, suggesting very little vibrations about the 3-fold axes. If we consider the arrangement of the proteins for a T=4 virus in Figure 5, we see that the 3-fold axis lies on the union of 3 different hexamers. It makes sense that this point would be rigid and not susceptible to large radial displacement, hence the low representation of SAF10.

4.1.5 Bacteriophage Q β Capsid (1qbe, T = 3)

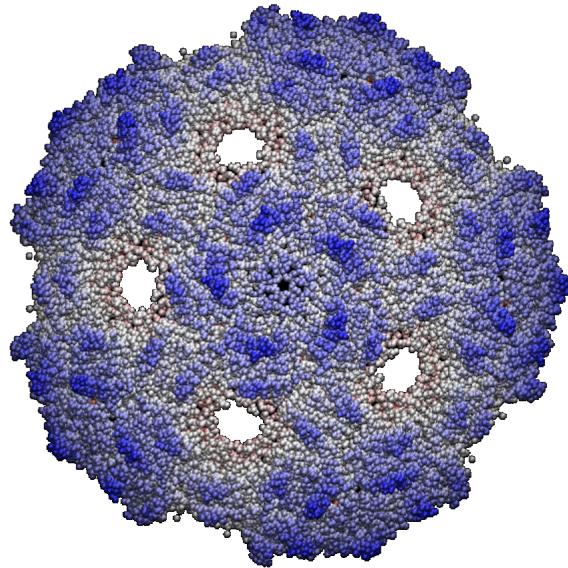


Figure 39: Viral capsid of Bacteriophage Q β (BQ). We observe large gaps in the protein outer shell around the 3-fold axes. The absence of mass at these points could potentially alter the oscillatory motion and reduce 3-fold dominant SAF overlap.

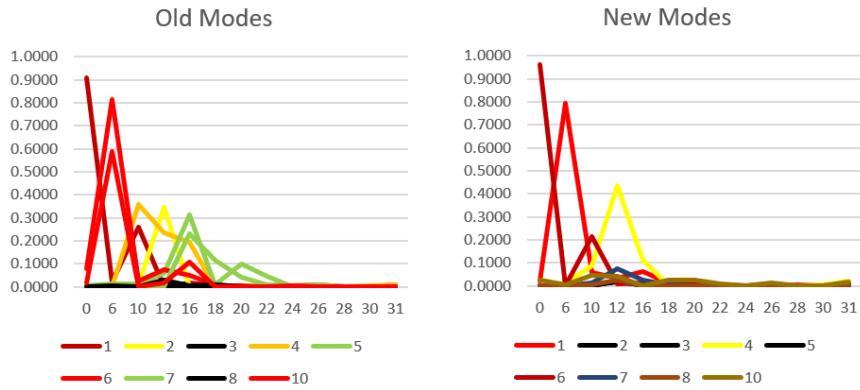


Figure 40: Comparison of the weights associated with each SAF for the Hammond mode computations (left) and the Rizzolo mode computations (right) for BQ. There is a lot of SAF representation for the Hammond modes, as evident by the multiple colors that are displayed.

The Hammond mode calculations show a lot of SAF coverage, but they unfortunately are all over the place in terms of degenerate modes calculated. Certain SAFs have single dominance, whereas other SAFs share their influence across multiple different mode calculations. The overall distribution and strong representation indicates that BQ can be classified as following the thin capsid postulate. The modes that share dominance could be a fault on the part of the calculation method, where some redundant modes are not being filtered out.

The Rizzolo mode calculations show a tighter concentration of SAF presence, but they have the benefit of having unique SAF dominance. SAF0, SAF6, and SAF12 are all well represented, but the Hammond mode calculations show dominance up until SAF16, and even some inclusion of the second set of SAFs, at order $l = 20$

An interesting parallel between the two sets of modes is the decomposition of the breathing mode. The Hammond breathing mode, HM_1 and its counterpart RM_6 show a strikingly similar decomposition pattern. HM_1 has a large weight associated with SAF0 ($\sigma > 0.90$), and then bounces up again at SAF10 ($\sigma \sim 0.25$). This behaviour is analogous with RM_6 , whose SAF0 has a weight of around 0.95 and pops up again for SAF10 at 0.20. These two modes evidently represent the same motion, despite being assigned a different order for solving and having slightly different σ values for each SAF. Having the slight influence from SAF10 indicates large radial vibrations about the 3-fold axes. This makes sense as there is a hole in the capsid of Bacteriophage Q β about the 3-fold axis, so the surrounding proteins might be more flexible without that added boundary condition.

Table 7: SAF Dominance for Hammond and Rizzolo Modes for Bacteriophage Q β Capsid.

SAF	HM_i	RM_i
0	1	6
6	6,10	1
10	4	-
12	2	4
16	5,7	-
18	-	-

The lack of representation of higher degree SAFs for the Rizzolo modes can be explained by the reduced resolution employed when assigning the frequency of blocks per protein. An interesting result observed in both Bacteriophage Q β and Cowpea Chlorotic Mottle Virus is the absence of SAF10 influence for the Rizzolo modes, even though a SAF10 dominant mode showed up in both of the Hammond mode sets. Both of these viruses are T=3, which may have some correlation with the flexibility around the 3-fold axes. Or maybe the resolution has some correlation with the ability to account for 3-fold oscillations.

4.1.6 Simian Virus 40 (1sva, T = 7d)

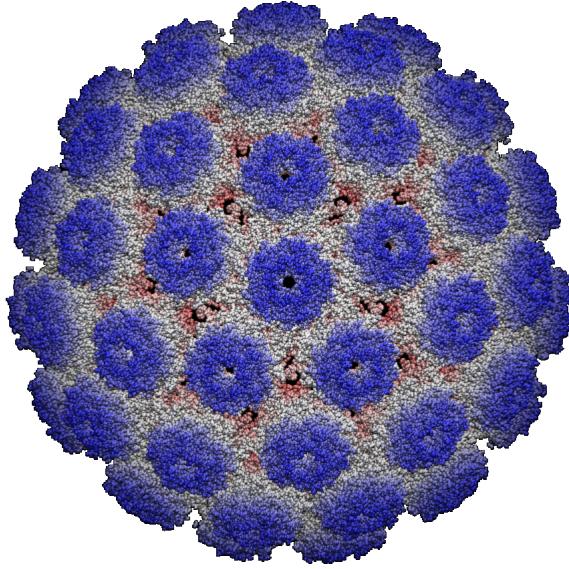


Figure 41: Viral capsid of Simian Virus 40 (SV40). SV40 has the most uniform organisation of polymers about the capsid, and it is even visually difficult to distinguish between hexamers and pentamers. This peculiarity could translate into having a strong breathing mode overlap with SAF0.

1sva, the first T=7 virus explored, has a polarized decomposition pattern between the Hammond and Rizzolo modes. Starting with the assessment with the Hammond modes, it's safe to say that the normal mode calculations are awful. Tight concentration, multiple degeneracies, and overall low overlap plague the normal modes with red flags. The σ values associated with SAF0 don't even exceed 0.70, whereas normally the breathing mode shows weights of ≈ 0.90 , or nearly 1.0 for RM₁. Even the second most dominant SAF, $l = 6$. shows extremely weak overlaps with the highest weight just hovering over 0.40. For nearly all of the viruses analyzed, both SAF0 and SAF6 have showed consistently high overlap.

The Rizzolo modes, on the contrary, show legitimate results. There is a discrete, spread out distribution of SAF dominance, extending even up until SAF16, which has been a rare occurrence for the lower definition, Rizzolo modes. These results show that even a higher triangulation number virus can be properly described as a thin-shelled capsid.

Table 8: SAF Dominance for Hammond and Rizzolo Modes for Simian Virus 40.

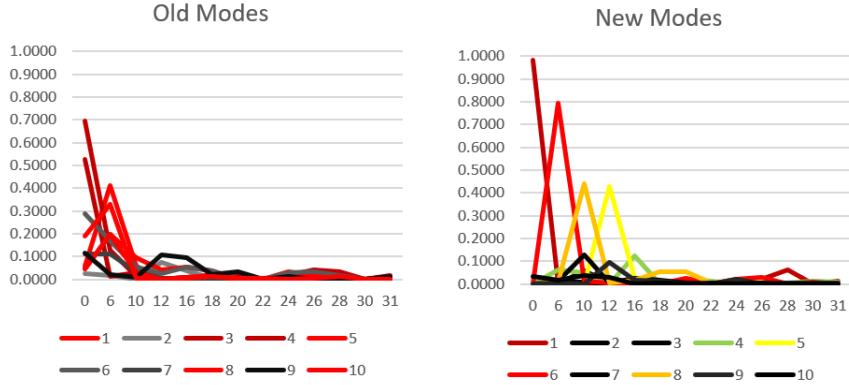


Figure 42: Comparison of the weights associated with each SAF for the Hammond mode computations (left) and the Rizzolo mode computations (right) for SV40. The Hammond modes show large degeneracy while the Rizzolo modes have a nice distribution.

SAF	HM_i	RM_i
0	3,4	1
6	1,5,8,10	6
10	-	8
12	-	5
16	-	4
18	-	-

What's not surprising for SV40 is the inclusion of SAF10, made up of 3-fold oscillations, given the relatively uniform appearance of the virus as per Figure 41. This uniform arrangement of proteins explains why the breathing mode is *so* close to 1.0, which would represent **perfectly** uniform radial displacement. There is very little different radial motion to distribute energy throughout the vibrations given how similar each region on the capsid is. In conclusion, high T-number viruses can accurately be modeled using a thin-shelled homogeneous mass distribution.

4.1.7 Bursal Disease Virus VP2 (2df7, T = 1)

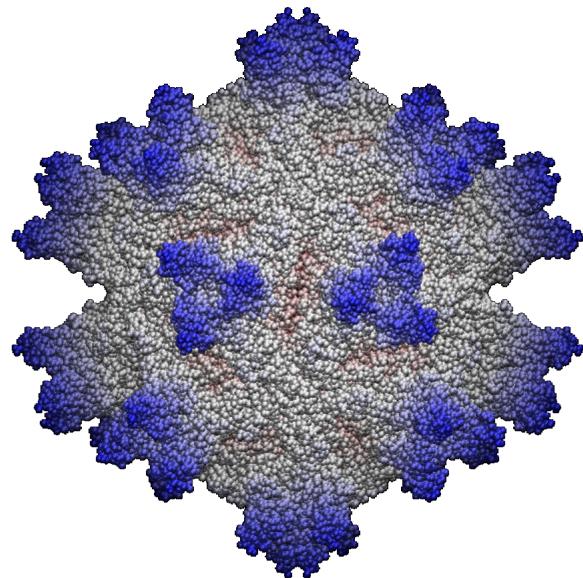


Figure 43: Viral capsid of Bursal Disease Virus VP2 (BDV) from the 2-fold axis. Being the first T=1 virus decomposed, the structure is quite different from the previous viruses. We see that the most prominent protrusions occur around the 3-fold axes.

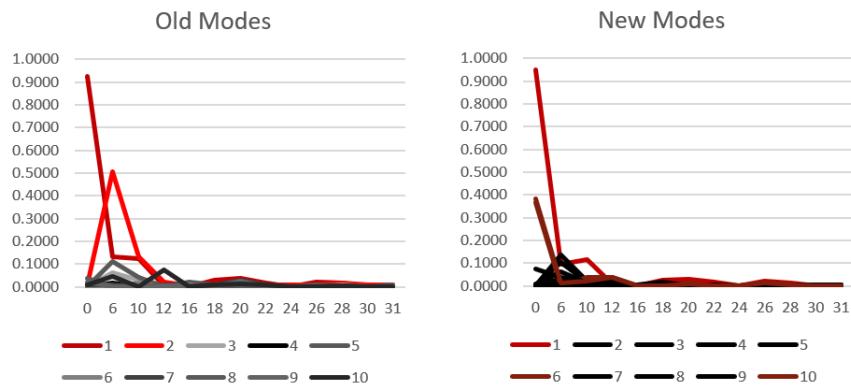


Figure 44: Comparison of the weights associated with each SAF for the Hammond mode computations (left) and the Rizzolo mode computations (right) for BDV.

Bursal Disease Virus VP2 is the first T=1 virus to be analyzed, and the results of the decomposition can provide new information about how the arrangement of the 60 proteins affect the normal mode calculations. The decomposition of the Hammond modes reveals that only the very first two SAFs, SAF0 and SAF6 show large overlap values. After SAF6, the weights drop off steeply. 80% of the modes that were calculated showed almost no overlap with any of the SAFs.

We see a similar trend with the Rizzolo modes, except the representation is even more disappointing, only having a few degenerate breathing modes and not extending beyond degree 0. Both of these decompositions suggest that this T=1 virus is not properly characterized by a thin-shell homogeneous mass distribution. There are a number of reasons that could possible explain these poor results.

The first, most convincing argument is directly tied with the resolution and T-number. T=1 viruses only have 60 proteins, and therefore the resolution for a given virus is 3 times lower than a T=3 virus, for example. The Rizzolo modes have even *lower* resolution than the Hammond modes, and the disappearance of SAF6 shows how the remarkably low resolution doesn't allow for even the least dramatic protein deformations to arise in the normal mode calculations.

Table 9: SAF Dominance for Hammond and Rizzolo Modes for Bursal Disease Virus VP2.

SAF	HM _i	RM _i
0	1	1,6,10
6	2	-
10	-	-
12	-	-
16	-	-
18	-	-

Another possible contributing factor is the relative thickness of the viral capsid. The thicker the capsid, the less likely it is going to oscillate in an ideal fashion.

4.1.8 Bacteriophage MS2 (2ms2, T =3)

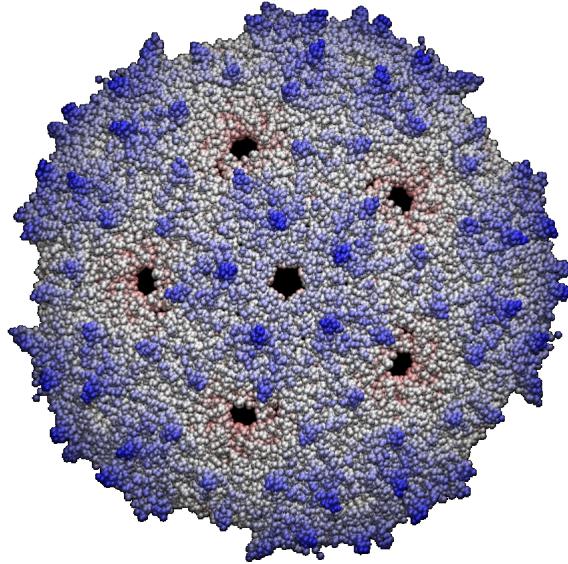


Figure 45: Viral capsid of Bacteriophage MS2 (MS2), viewed from the 5-fold axis. The pentamers and hexamers are not uniformly raised like for Cowpea Chlorotic Mottle Virus or Turnip Yellow Mosaic Virus. Like in Bacteriophage Q β , there are holes at the 3-fold axes, and an extra hole at the 5-fold axes.

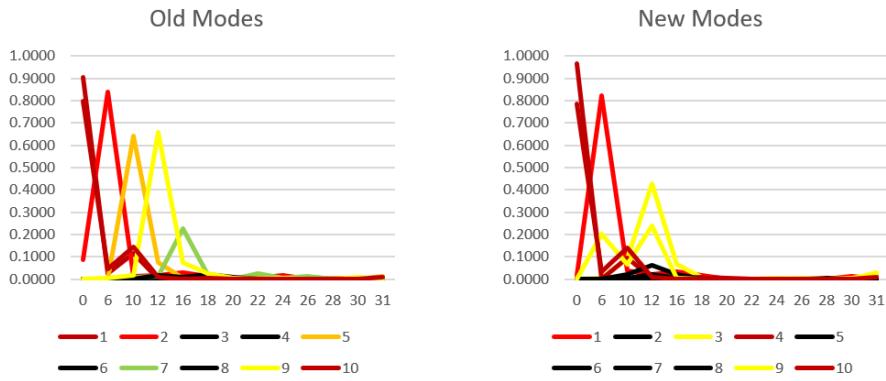


Figure 46: Comparison of the weights associated with each SAF for the Hammond mode computations (left) and the Rizzolo mode computations (right)

The Hammond calculated modes of Bacteriophage MS2 show a beautiful, almost perfectly unique distribution of SAF dominant modes, with a range

extending to SAF16. The weights associated with each SAF are even quite large, with both $\sigma_{\text{SAF}10}$ and $\sigma_{\text{SAF}12} > 0.60$. It's interesting to note that the breathing mode for both the Hammond and Rizzolo calculated modes ($\text{HM}_{1,10}$ and $\text{RM}_{4,1}$) display the same $\text{SAF}0 \rightarrow \text{SAF}10$ bouncing pattern. This again indicates the oscillation around the 3-fold axes during the breathing mode. This normal mode pattern is similar to the breathing mode of Bacteriophage $Q\beta$. If we compare Figure 39 and Figure 45, the images of the Bacteriophage $Q\beta$ and Bacteriophage MS2, respectively, we notice that both of these viruses have holes in the capsid around around the 3-fold axis. It would appear that the discontinuation of the viral capsid has ramifications for (at least) the breathing mode's unique radial displacement pattern.

The Rizzolo calculated modes display a smaller range of SAF dominance, and also exhibits the curious trend found in T=3 viruses where the Rizzolo calculated modes show an absence of an SAF10, 3-fold dominant mode. This might be explained by the lower resolution, and maybe the blocks that are generated are so far from the holes that there isn't even any mass object able to radiate along the 3-fold axes.

Table 10: SAF Dominance for Hammond and Rizzolo Modes for Bacteriophage MS2.

SAF	HM_i	RM_i
0	1,10	4,10
6	2	1
10	5	-
12	9	3,9
16	7	-
18	-	-

4.1.9 Human Papilloma Virus Type 16 (3j6r, T = 7d)

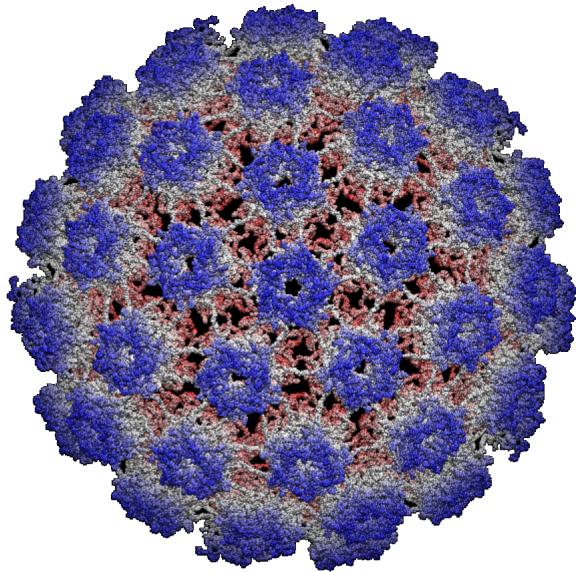


Figure 47: Viral capsid of Human Papilloma Virus 16 (HPV16), viewed from the 5-fold axis. HPV16 has a curious structure with thin proteins “bridging” the gap between polymers. This structural component could affect the normal modes for this virus.

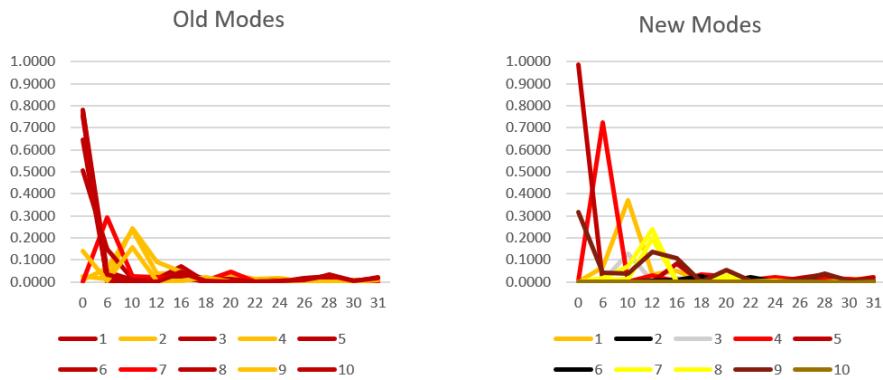


Figure 48: Comparison of the weights associated with each SAF for the Hammond mode computations (left) and the Rizzolo mode computations (right) for HPV16.

Human Papilloma Virus Type 16, the second T=7 that shows up in my

analysis, is the last virus who had both Hammond and Rizzolo normal modes. The Hammond modes paint a familiar picture of repeated degenerate modes and a small distribution of SAF degrees. Nearly all the representation is limited to the first 3 SAFs. It's important to comment on the fact that the weights themselves are quite low.

The Rizzolo modes provide a different story about the distribution of normal modes calculated. Already there is a wider distribution of discrete SAF influence on the modes. One of the expected results that was not present in the Hammond mode set is the incredibly high $\sigma_{\text{SAF}0}$ value, nearly reaching 1.0. This can be explained by the relative consistency and even spacing of the protein units on the capsid structure. Apart from the high SAF0 influence, the other weights associated with the rest of the SAFs are not that strong. This could possibly be explained by the specific protein linking of HPV16. In Figure 47, we see an unorthodox linking between the hexamers and pentamers. This bridged linking could reduce the amount of radial displacement that takes placing, acting as a sort of glue between the polygon protein structures.

Table 11: SAF Dominance for Hammond and Rizzolo Modes for Human Papilloma Virus Type 16.

SAF	HM _i	RM _i
0	1,3,5,6,8,10	5
6	7	4
10	2,4,9	1
12	-	7,8
16	-	-
18	-	-

5 Conclusion

The normal modes for viruses, in which all of its atoms are oscillating at the same frequency, can be thought of as a very complicated system of masses and springs. Each atom has its own mass, and the bonds between atoms can be modeled as springs. The solution for normal modes of spherical viruses incorporates a sum of Laplacian spherical harmonics, based on the assumption that a viral capsid can be approximated by a thin-shelled homogenous mass distribution. However, the regular spherical harmonics by themselves do not exhibit the same type of symmetry as spherical viruses. In order to achieve icosahedral symmetry, we need to linearly combine specific spherical harmonics of the same order l , which generates the Symmetry Adapted Functions (SAFs). The SAFs are orthogonal functions that can be combined to build any function that is icosahedrally defined. These SAFs are not identical, however, and they predict different patterns of motion. Some SAFs predict max radial displacement about the symmetry axes, while others describe perfectly uniform contraction and expansion. As the order of the SAF is increased, the gradient between close regions increases. In other words, the SAF appears much spikier.

The calculated normal modes for viruses are organized as a series of displacement instructions (in the form of a 3 dimensional vector) for every single atom. For a typical virus of 200,000 atoms, this means that the normal mode data is stored in the form of a 200,000 x 3 matrix. If we want to compare different calculated normal modes, we can build VMD animations based on the displacement instructions and compare them qualitatively. However, until now there has been no robust way to decompose and describe the normal modes with a set of parameters, which facilitates comparison and analysis. Instead of comparing two 200,000 x 3 matrices, the decomposition outlined in this paper allows us to compare the overlap weights of each mode with 13 SAFs. Because the SAFs are orthogonal, icosahedral functions, they can be used as a basis for icosahedral normal modes. Simply taking the calculated mode matrix and dotting it with an ‘artificial’ mode predicted by a given SAF yields the overlap associated with that SAF. These overlap weights provide crucial details about the composition of a normal mode. If two different modes have the same exact decomposition but different mode number, we know that the two motions are exactly alike. One crucial piece of information regarding the SAFs is that they can only account for radial displacements. That is to say, if a mode is predominantly composed of twisting and stretching, the overlap values will be alarmingly low.

There were two normal mode sets that were decomposed and compared in this paper. The first set, denoted the Hammond modes, was calculated by Rob Hammond for his SIP. The second set, denoted the Rizzolo modes, was calculated by Skylar Rizzolo for his SIP. Despite the fact that some of these calculated modes used the same viruses as input, they were outputting different motions. In order to see which modes describe similar motion, despite having a different mode number, both sets of modes were decomposed into the 13 SAFs and then compared. Modes that were visually similar when undergoing qualitative analysis showed similar decomposition patterns. Modes that were composed only of contraction and expansion showed strong overlap with SAF0, the breathing mode. We believe that an accurate set of calculated modes contains modes that show broad representation of SAFs, has strong overlap values with the SAFs, and does not contain multiple degenerate motions (e.g. having 4/10 modes being the breathing mode). Viruses that have convincing mode sets prove that they can be accurately approximated as a thin-shelled homogenous mass distribution. Viruses that showed poor overlap and tight distributions suggest that they cannot be described well by a thin-shelled model.

References

- [1] W. Prandl, P. Schiebel, and K. Wulf. *A Recursive Algorithm for the Generation of Symmetry-Adapted Functions: Principles and Applications to the Icosahedral Group*. Institut für Kristallographie, Universität Tübingen, Charlottenstrasse 33, D-72070 Tübingen, Germany
- [2] R. Hammond. *The Mechanical Properties of Spherical Viral Capsids*. Department of Physics, Kalamazoo College, Kalamazoo, Michigan, United States of America
- [3] S. Rizzolo. *Exploring Viral Maturation Pathways Using Normal Mode Flexible Fitting Software*. Department of Computer Science, Kalamazoo College, Kalamazoo, Michigan, United States of America
- [4] D. Wilson. *Protruding Features of Viral Capsids Are Clustered on Icosahedral Great Circles*. Department of Physics, Albion College, 611 E. Porter St., Albion, Michigan, United States of America
- [5] Viral Zone https://viralzone.expasy.org/1057?outline=all_by_protein
- [6] G. Firsove <http://microproject.jjht.org/wp-content/uploads/2012/12/Virus-Shapes.jpg>
- [7] L. Poudel *Deciphering of packing signal hypothesis in bacterophage RNA recognition by the MS2 capsid protein in virus assembly*
- [8] Wikipedia https://en.wikipedia.org/wiki/File:Spherical_Harmonics.png