

# Contents

<b>1</b>	<b>Radial Density Function</b>	<b>2</b>
1.1	Calculation of Distances with Periodicity . . . . .	2
1.2	Adding Noise For Atom Vibration . . . . .	3
1.3	Cubane Example . . . . .	3
1.3.1	Cubane Radial Density Functions . . . . .	4
1.4	Experimental and Theoretical RDFs for Known Structures . . . . .	6
1.4.1	Ga As . . . . .	6
1.4.2	In As . . . . .	7
1.4.3	Si Lattice . . . . .	7
<b>2</b>	<b>Smoothing Analysis</b>	<b>8</b>
<b>3</b>	<b>Noise Analysis</b>	<b>11</b>
3.1	Peak Counts . . . . .	11
<b>4</b>	<b>Recognition Using Eigenfaces</b>	<b>13</b>
4.1	Mean Image . . . . .	13
4.2	Variance Explained by Principal Components . . . . .	14
4.3	Eigenfaces . . . . .	15
4.4	Data in Eigenspace . . . . .	18
4.4.1	Eigenspace Outliers . . . . .	23
4.5	Experimental Image Recognition . . . . .	28
4.5.1	3 Principal Components . . . . .	28
4.5.2	10 Principal Components . . . . .	39
4.5.3	128 Principal Components . . . . .	50
4.6	Synthetic Experimental Image Recognition . . . . .	61
<b>5</b>	<b>Recognition Using Sparse Representations</b>	<b>63</b>
5.1	Experimental Image Recognition . . . . .	63
5.2	Synthetic Experimental Image Recognition . . . . .	74

# 1 Radial Density Function

## 1.1 Calculation of Distances with Periodicity

Suppose a large chemical structure has uncountably many atoms but they follow a periodic pattern of  $n$  atoms every  $p$  Angstroms. The atom locations within a period are given by  $a_1, a_2, \dots, a_n$  where  $a_i \in \mathbb{R}^3$ . The radial density function is the distribution of pairwise distances between these atoms.

The distances  $d$  between atoms  $a_i$  and  $a_j$  where  $i \neq j$ , atom  $a_i$  has been displaced by  $x$ , and atom  $a_j$  has been displaced by  $y$  per the periodicity is

$$\begin{aligned} d^2 &= \langle a_i + x - (a_j + y), a_i + x - (a_j + y) \rangle \\ &= \langle a_i - a_j, a_i - a_j \rangle + \langle x - y, x - y \rangle + 2\langle a_i - a_j, x - y \rangle \end{aligned}$$

where  $x = (k_1 p, k_2 p, k_3 p)$  for  $k_i \in \mathbb{Z}$  and  $y = (l_1 p, l_2 p, l_3 p)$  for  $l_i \in \mathbb{Z}$ . Here  $\langle x, y \rangle$  denotes the inner product between  $x$  and  $y$ .

Suppose  $D$  is a random variable that samples at random the distances,  $d$ , in the chemical structure. The radial density function is the probability density function of this random variable. This function can be estimated empirically via a histogram.

The histogram is then normalized by the volume of a spherical shell.

$$\begin{aligned} &\frac{4}{3}\pi(r + \Delta r)^3 - \frac{4}{3}\pi r^3 \\ &= \frac{4}{3}(3r^2\Delta r + 3r(\Delta r)^2 + (\Delta r)^3) \\ &\approx 4\pi r^2\Delta r \end{aligned}$$

where  $\Delta r$  tends to zero.

For a histogram with frequency,  $f$ , for bin  $[d_i, d_{i+1}]$ , we replace  $f$  with  $f/d_i^2$ . And then normalize the histogram so that the sum over all bins is one.

## 1.2 Adding Noise For Atom Vibration

Due to the vibrations of the molecules, the radial density function will not be just the equilibrium positions. We can approximate this fluctuation in distances via a Gaussian filter or Weierstrass transform.

$$F(x) = \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} f(y) e^{-\frac{(x-y)^2}{4t}} dy$$

Given that the density function is only defined for a finite number of distances, we use a discrete version of the transform making sure to keep the sum of the weights equal to one.

$$F(d_k) = \frac{\sum_{d_i=d_0}^{d_n} f(d_i) \exp\left(-\frac{(d_k-d_i)^2}{4t}\right)}{\sum_{d_i=d_0}^{d_n} \exp\left(-\frac{(d_k-d_i)^2}{4t}\right)}$$

where  $d_0$  is the minimum distance and  $d_n$  is the maximum distance.

## 1.3 Cubane Example

As an example of the above, below are the calculations for cubane ( $C_8H_8$ ).

Here are the coordinates of the elements in cubane in Angstroms.

Element, x, y, z

C, 1.2455, 0.5367, -0.0729

C, 0.9239, -0.9952, 0.0237

C, -0.1226, -0.7041, 1.1548

C, 0.1989, 0.8277, 1.0582

C, 0.1226, 0.7042, -1.1548

C, -0.9239, 0.9952, -0.0237

C, -1.2454, -0.5367, 0.0729

C, -0.1989, -0.8277, -1.0582

H, 2.2431, 0.9666, -0.1313  
H, 1.6638, -1.7924, 0.0426  
H, -0.2209, -1.2683, 2.0797  
H, 0.3583, 1.4907, 1.9059  
H, 0.2208, 1.2681, -2.0799  
H, -1.6640, 1.7922, -0.0427  
H, -2.2430, -0.9665, 0.1313  
H, -0.3583, -1.4906, -1.9058

### 1.3.1 Cubane Radial Density Functions

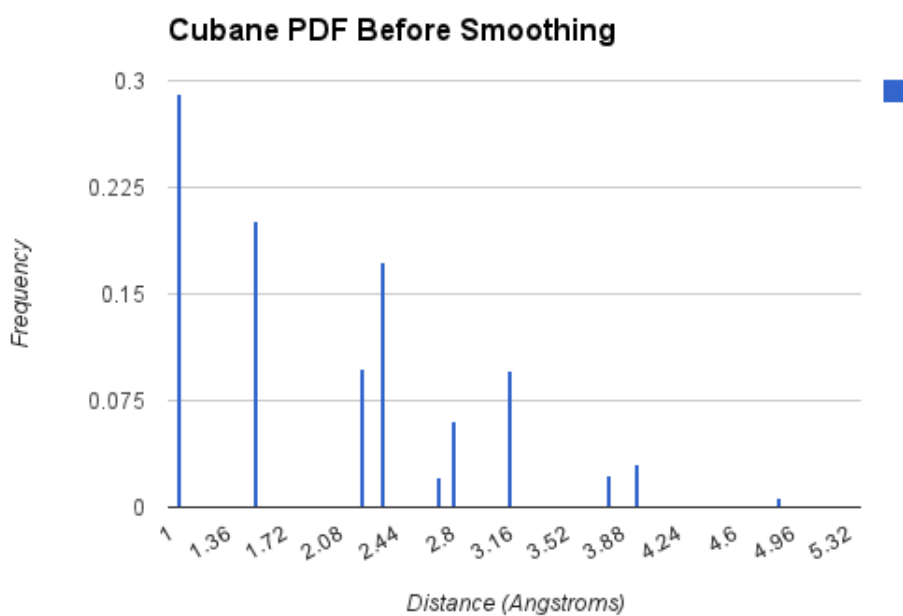


Figure 1: Before Smoothing

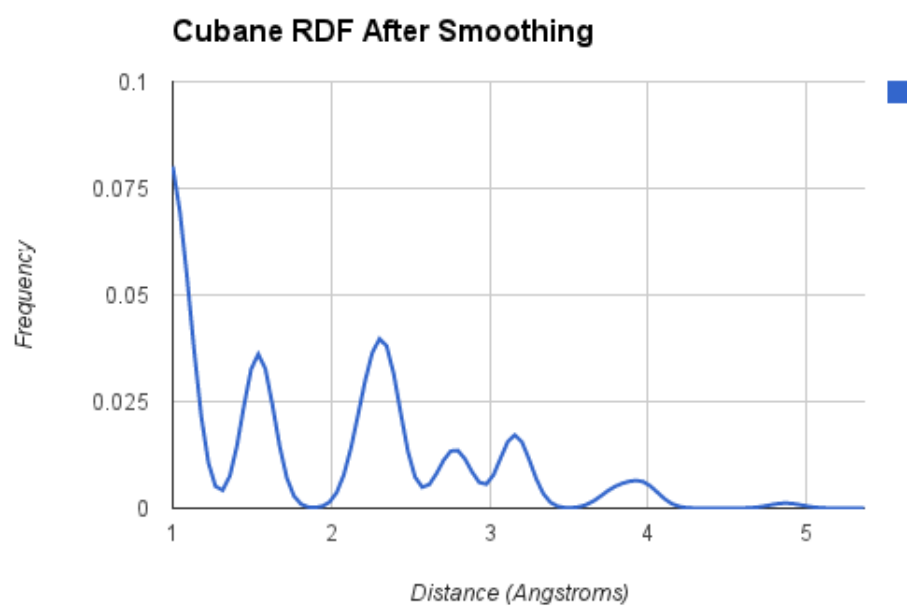


Figure 2: After Smoothing

## 1.4 Experimental and Theoretical RDFs for Known Structures

For some structures, we are able to theoretically calculate the RDF from atom locations and also have the experimental RDF from Xray scattering. These known matches provide some insight into understanding how the experiments and theory align. The RDF comparison are shown below.

Outside of these structures, there are not many other known matches. There are a few reasons for this. First, if a structures is already known at the atomic level then there is no need to run an xray diffraction experiment. Second, if a structure is periodic as in a lattice, the atomic structure can be determined by xray diffraction which is easier and cheaper than xray scattering.

### 1.4.1 Ga As

Experimental Data: Pair Distribution Functions Analysis, Valeri Petkov

Calculated Data: Maria Chan

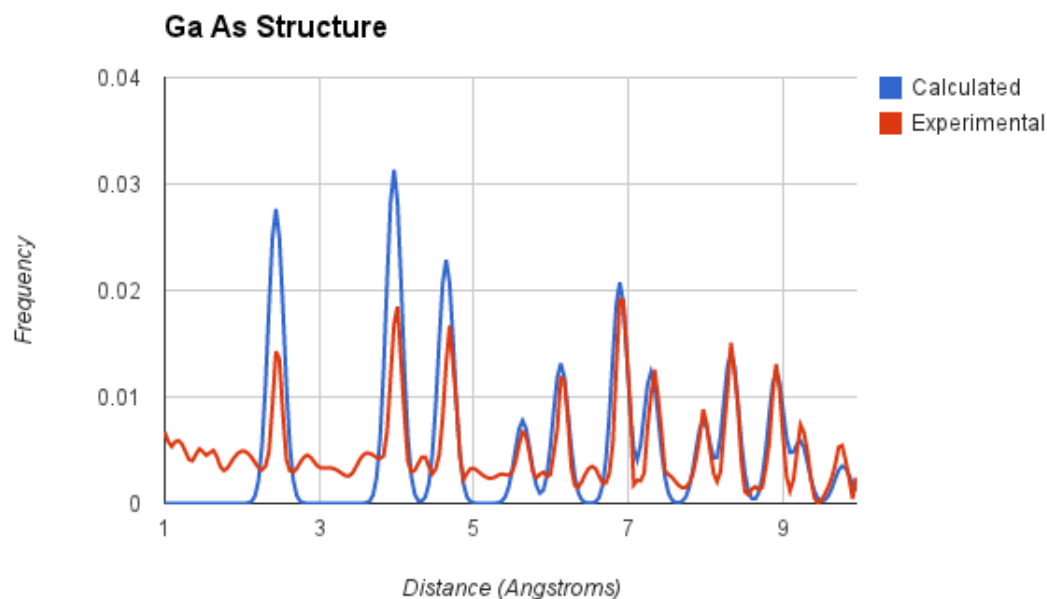


Figure 3: Ga As

### 1.4.2 In As

Experimental Data: Pair Distribution Functions Analysis, Valeri Petkov

Calculated Data: Maria Chan

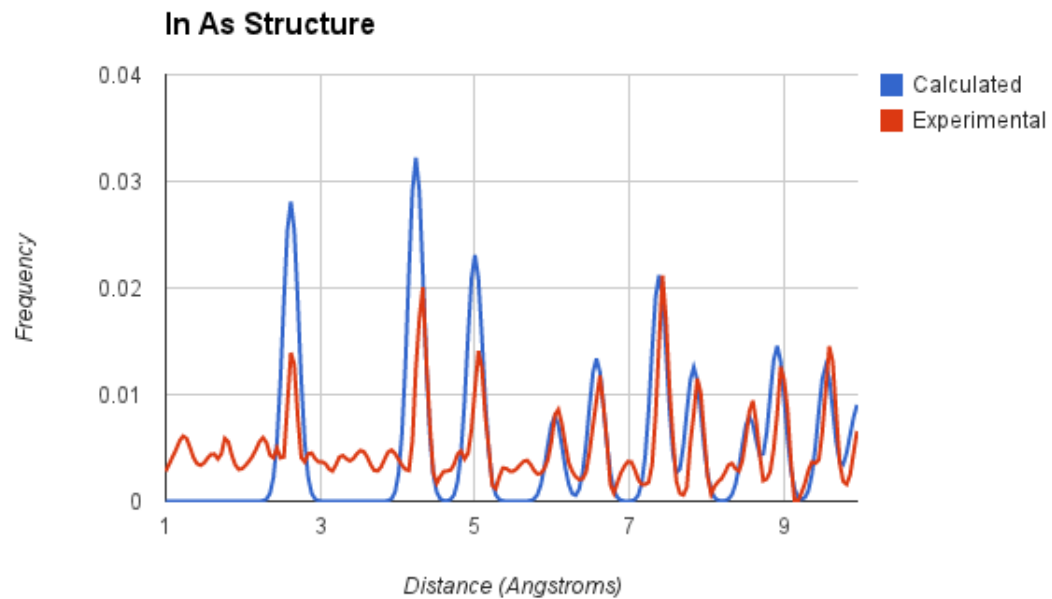


Figure 4: In As

### 1.4.3 Si Lattice

Experimental Data: J. AM. CHEM. SOC. VOL. 133, NO. 3, 2011, P: 503-512

Calculated Data: <http://materialsproject.org/materials/mp-149/>

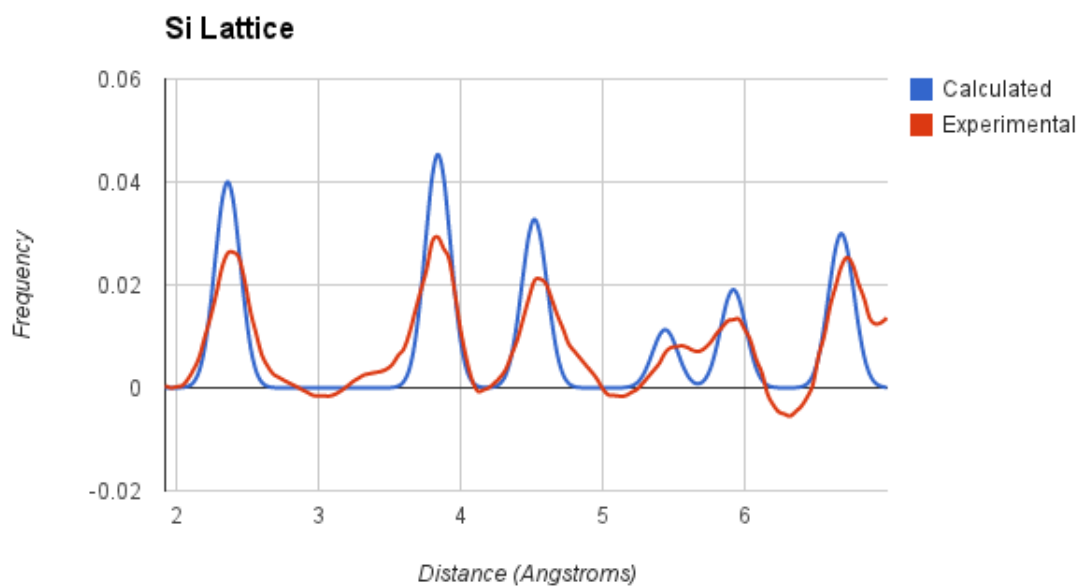


Figure 5: Si Lattice

## 2 Smoothing Analysis

smooth image and then normalize image solve for smoothing coefficient that results in the minimum l2 norm between smoothed SiLiCalc10001 and SiLiExpt1

minimum at 0.0092



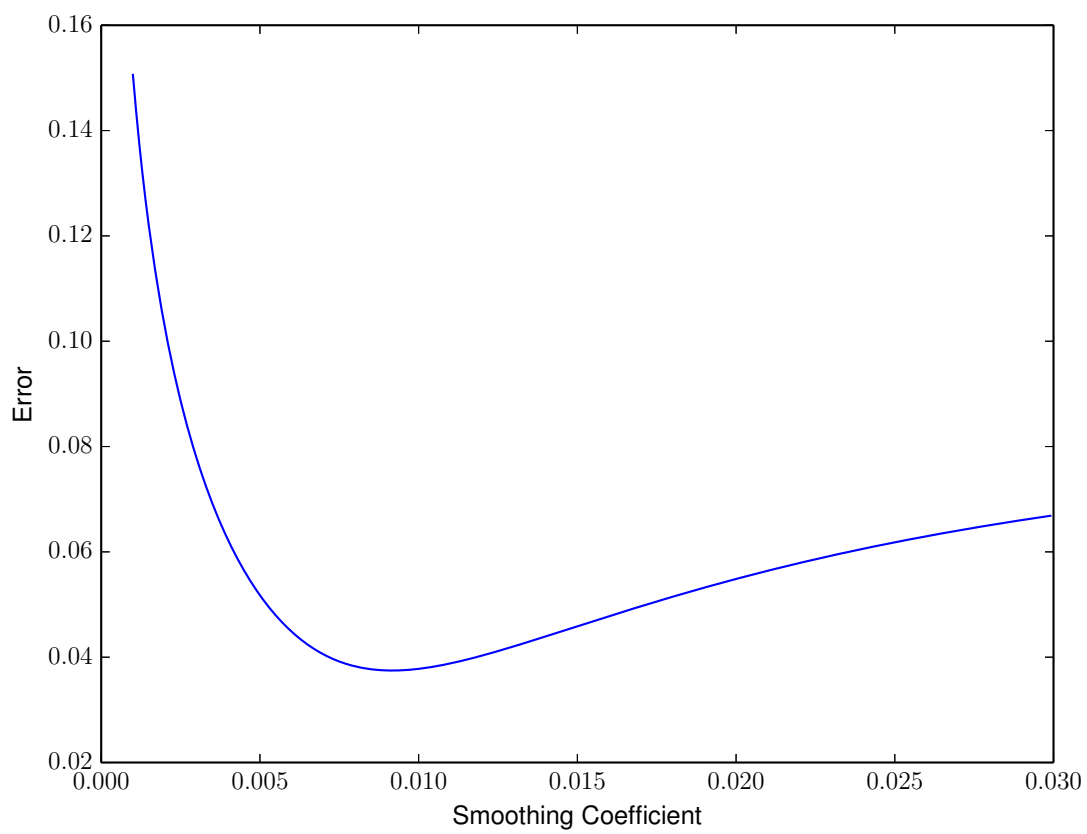


Figure 6: Smoothed - Expt Error vs Smoothing Coefficients

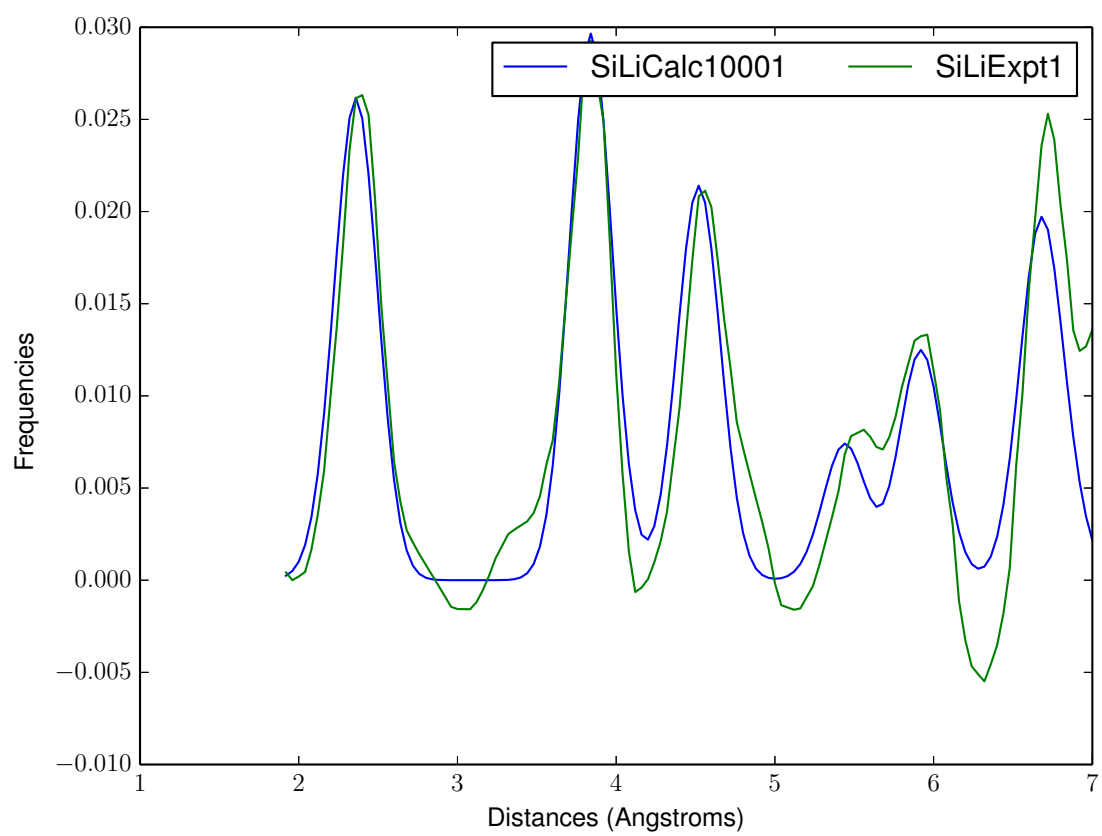


Figure 7: Smoothed SiLiCalc10001 vs SiLiExpt1

### 3 Noise Analysis

#### 3.1 Peak Counts

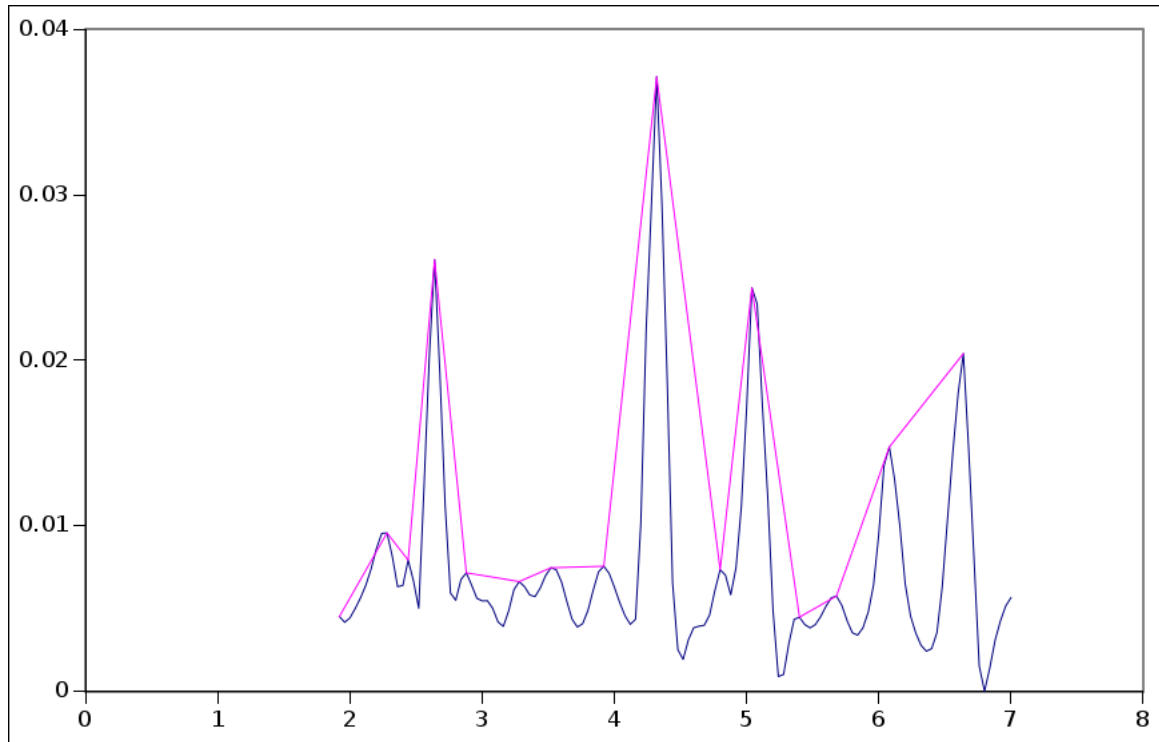


Figure 8: InAs Expt, Max 7 Angstroms

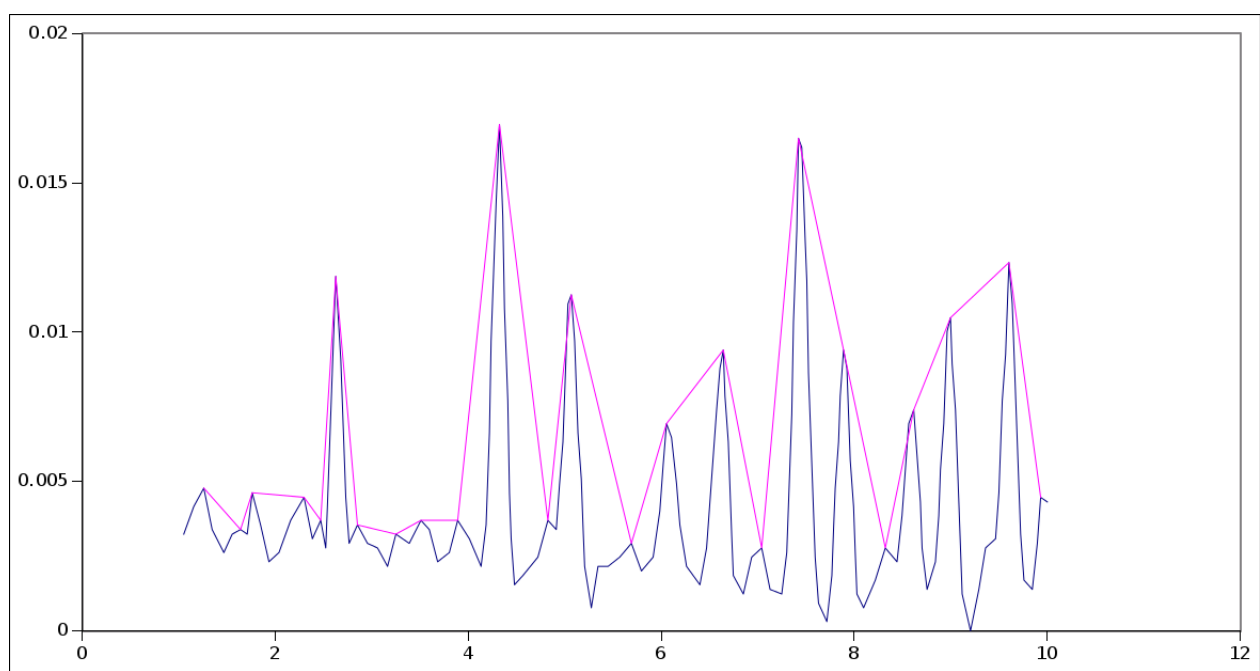


Figure 9: InAs Expt, Max 10 Angstroms

## 4 Recognition Using Eigenfaces

### 4.1 Mean Image

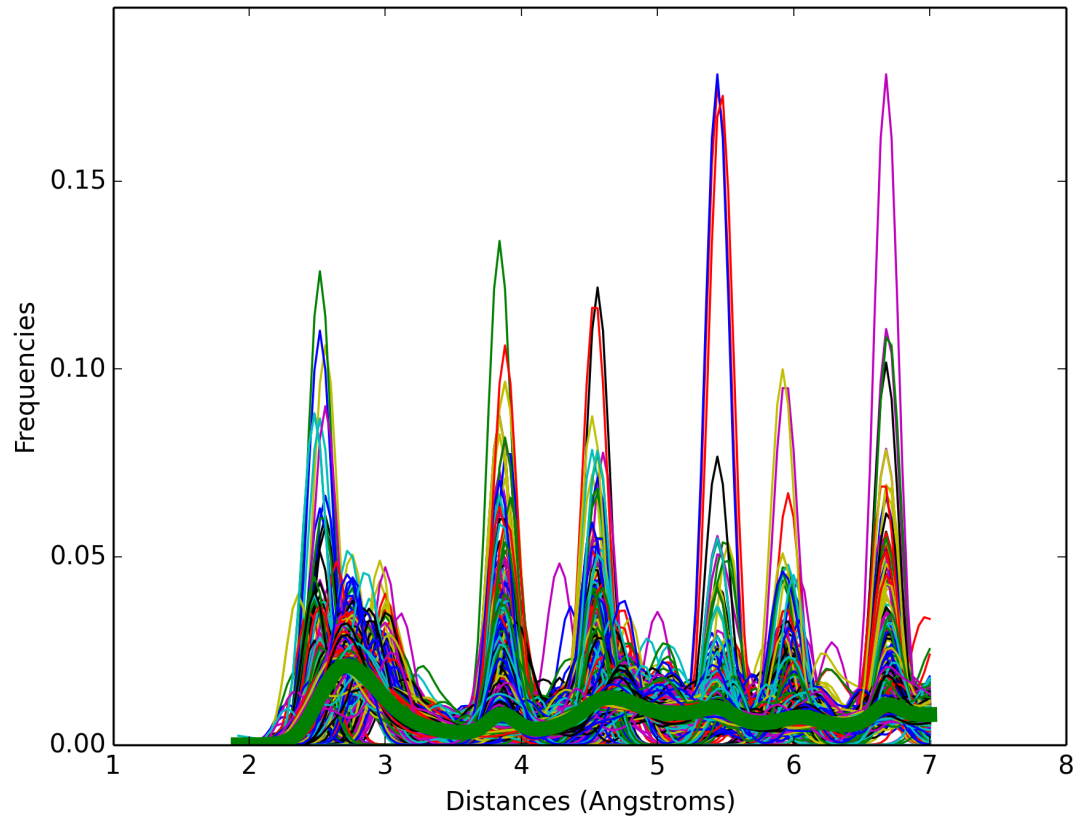


Figure 10: All Calculated Images with Mean

## 4.2 Variance Explained by Principal Components

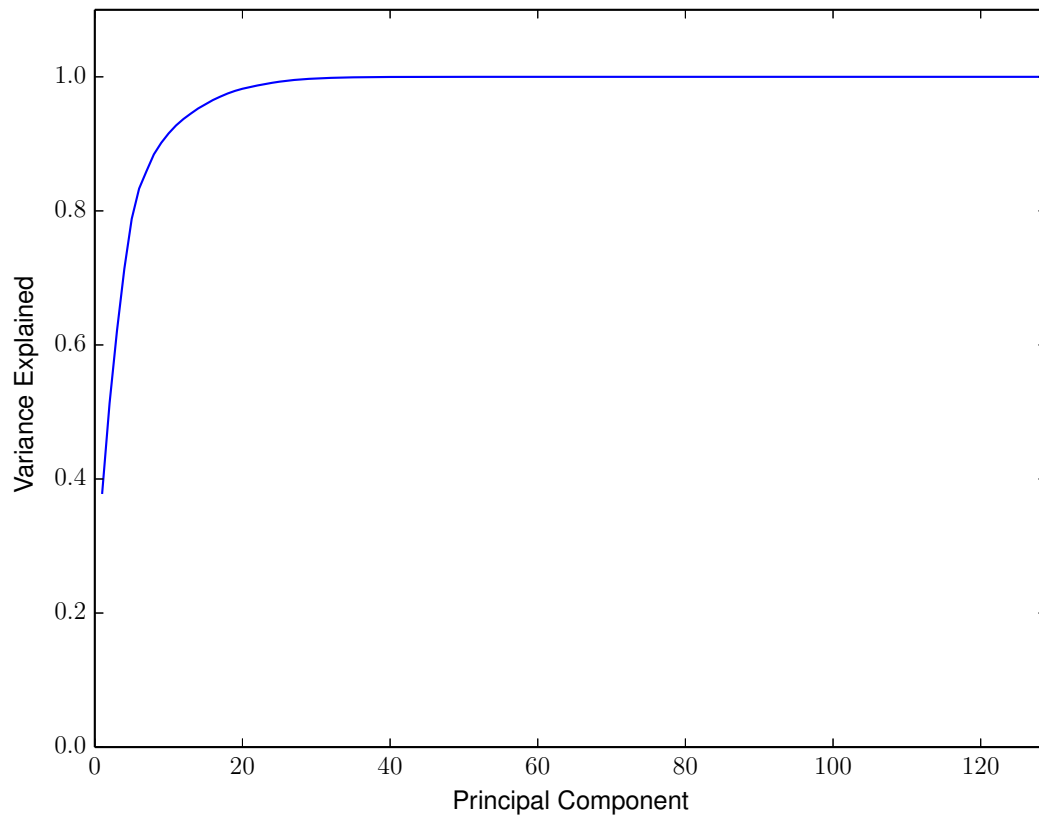


Figure 11: Cumulative Variance Explained by Principal Components

### 4.3 Eigenfaces

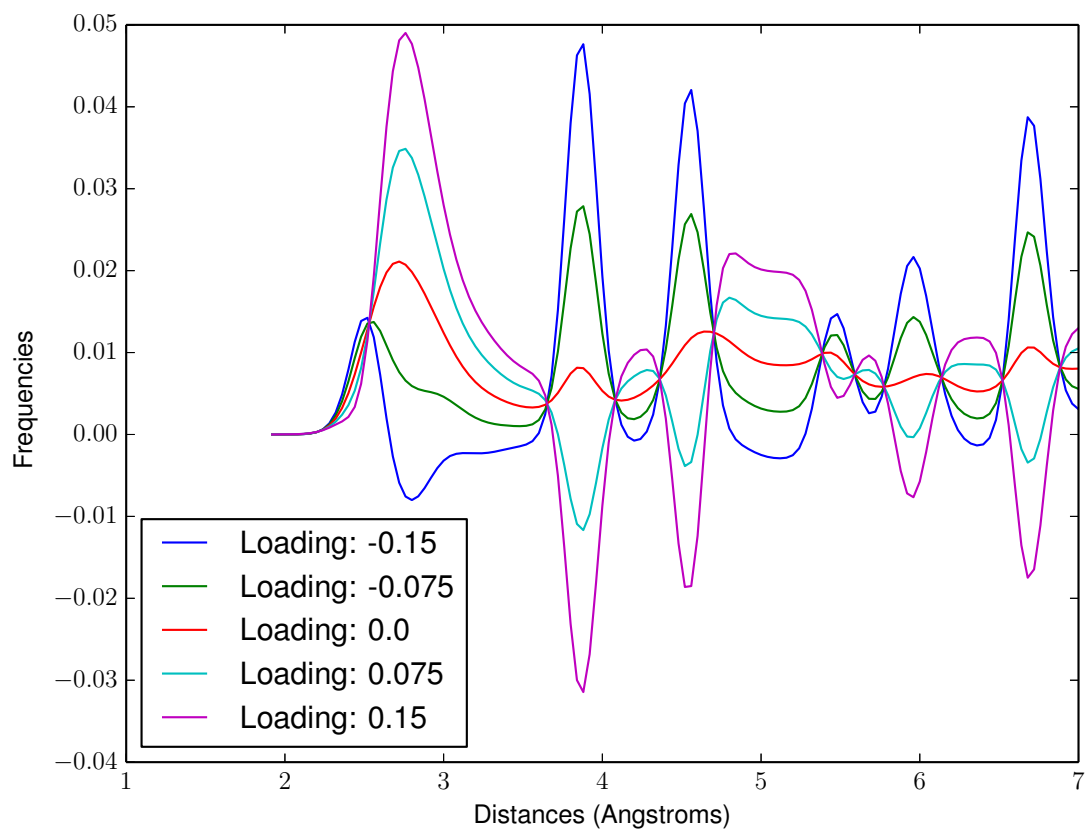


Figure 12: First Eigenface

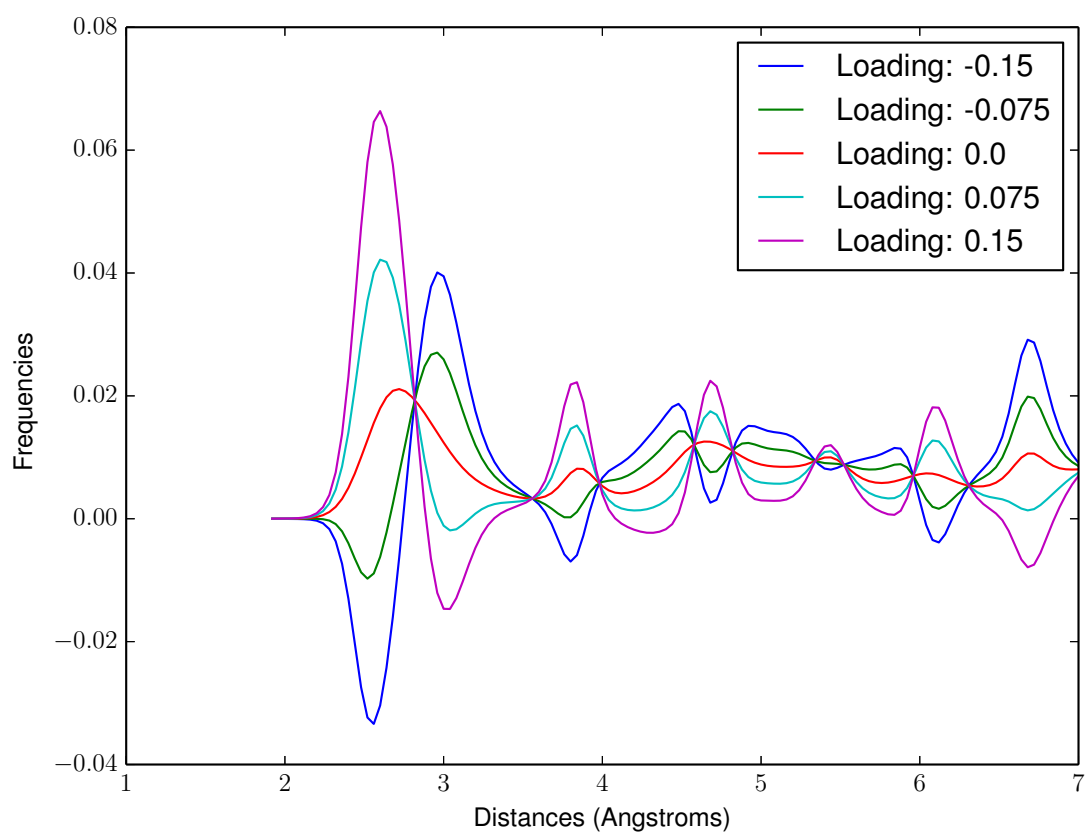


Figure 13: Second Eigenface



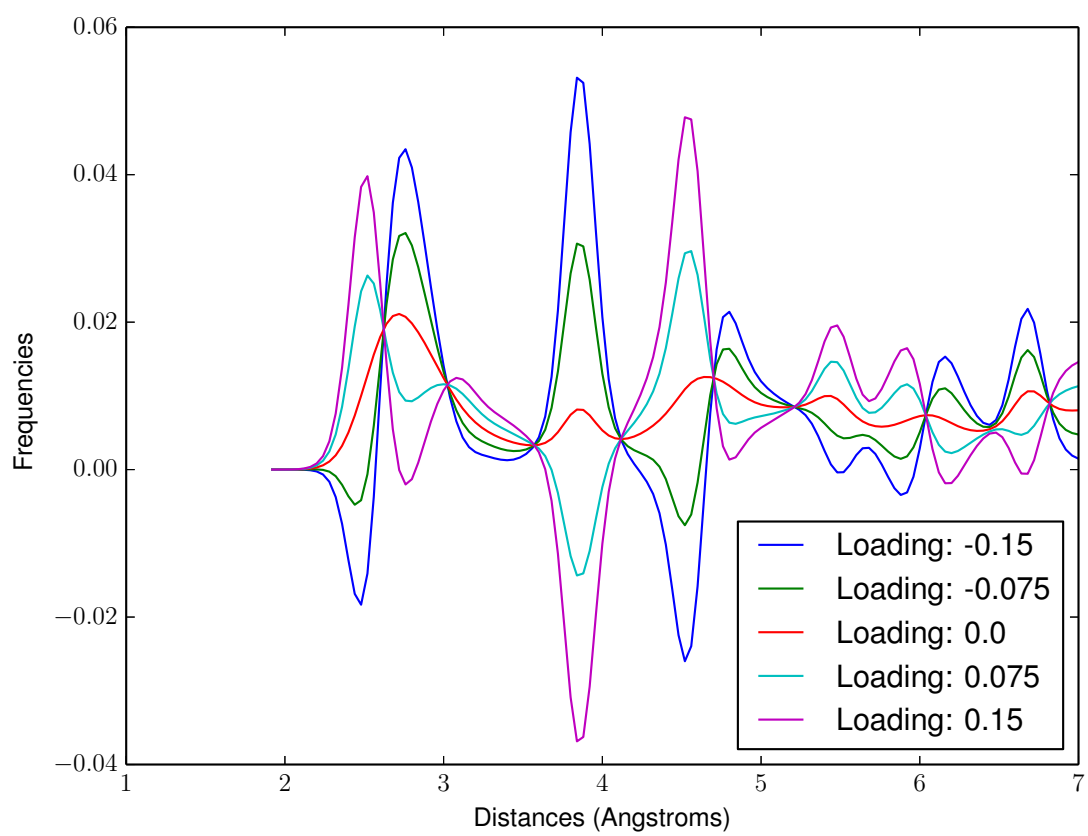


Figure 14: Third Eigenface

## 4.4 Data in Eigenspace

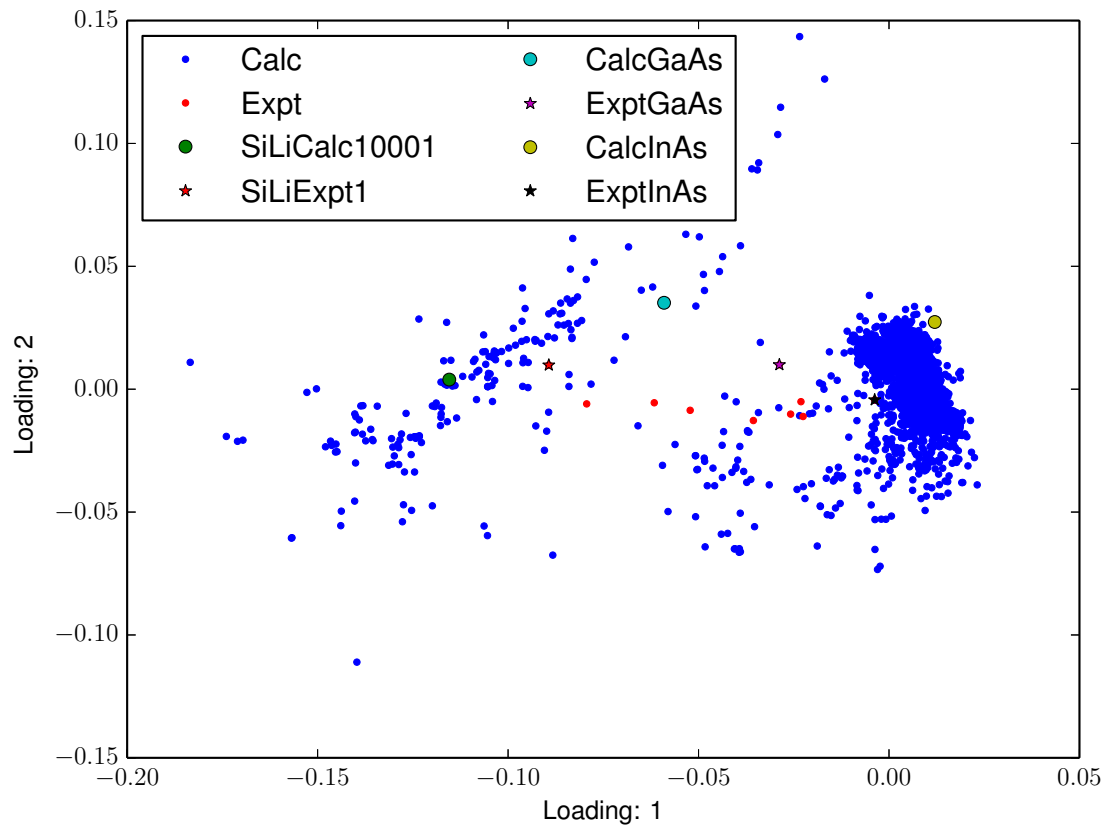


Figure 15: Loading 1 vs Loading 2

<b>Label</b>	<b>Loading 1</b>	<b>Loading 2</b>
SiLiExpt1	-0.0893	0.00982
SiLiExpt2	-0.0794	-0.006
SiLiExpt3	-0.0616	-0.0055
SiLiExpt4	-0.0522	-0.0086
SiLiExpt5	-0.0356	-0.0128
ExptGaAs	-0.0288	0.00994
SiLiExpt7	-0.0258	-0.0101
SiLiExpt6	-0.0231	-0.0051
SiLiExpt8	-0.0226	-0.0111
ExptInAs	-0.0038	-0.0044

Table 1: Experimental Data Sorted by Loading 1

<b>Label</b>	<b>Loading 1</b>	<b>Loading 2</b>
SiLiExpt5	-0.0356	-0.0128
SiLiExpt8	-0.0226	-0.0111
SiLiExpt7	-0.0258	-0.0101
SiLiExpt4	-0.0522	-0.0086
SiLiExpt2	-0.0794	-0.006
SiLiExpt3	-0.0616	-0.0055
SiLiExpt6	-0.0231	-0.0051
ExptInAs	-0.0038	-0.0044
SiLiExpt1	-0.0893	0.00982
ExptGaAs	-0.0288	0.00994

Table 2: Experimental Data Sorted by Loading 2

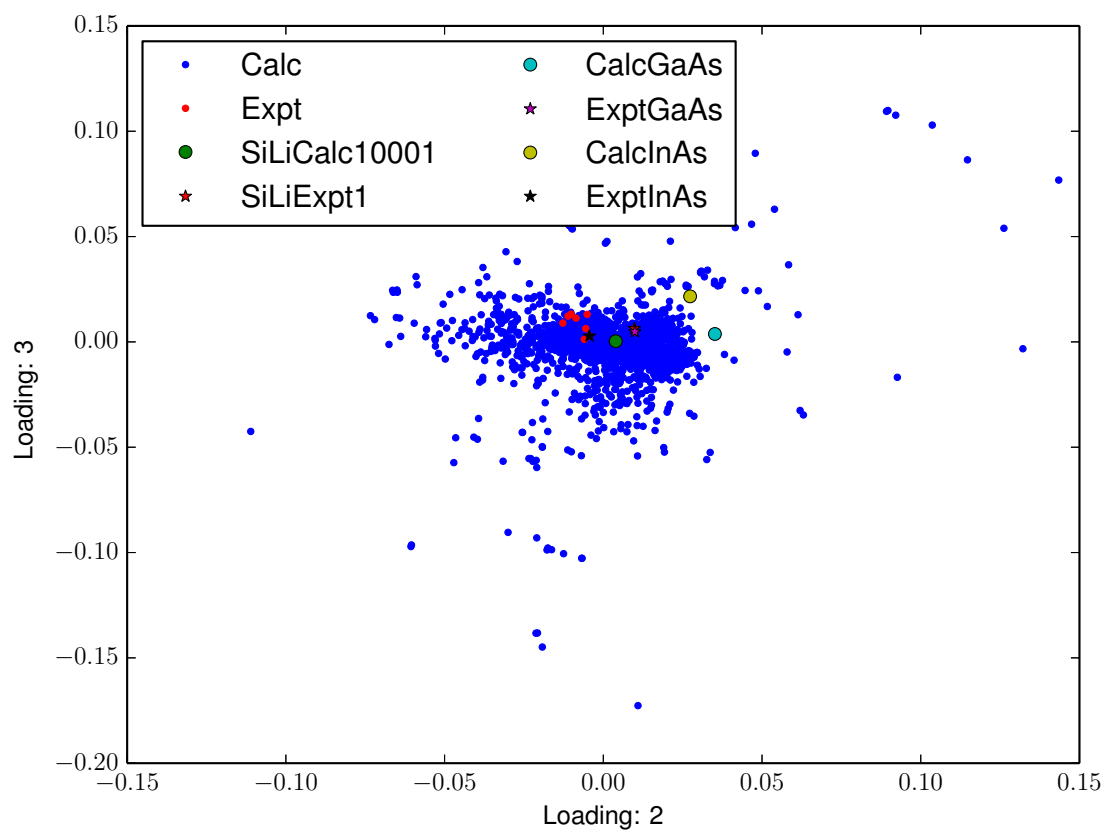


Figure 16: Loading 2 vs Loading 3

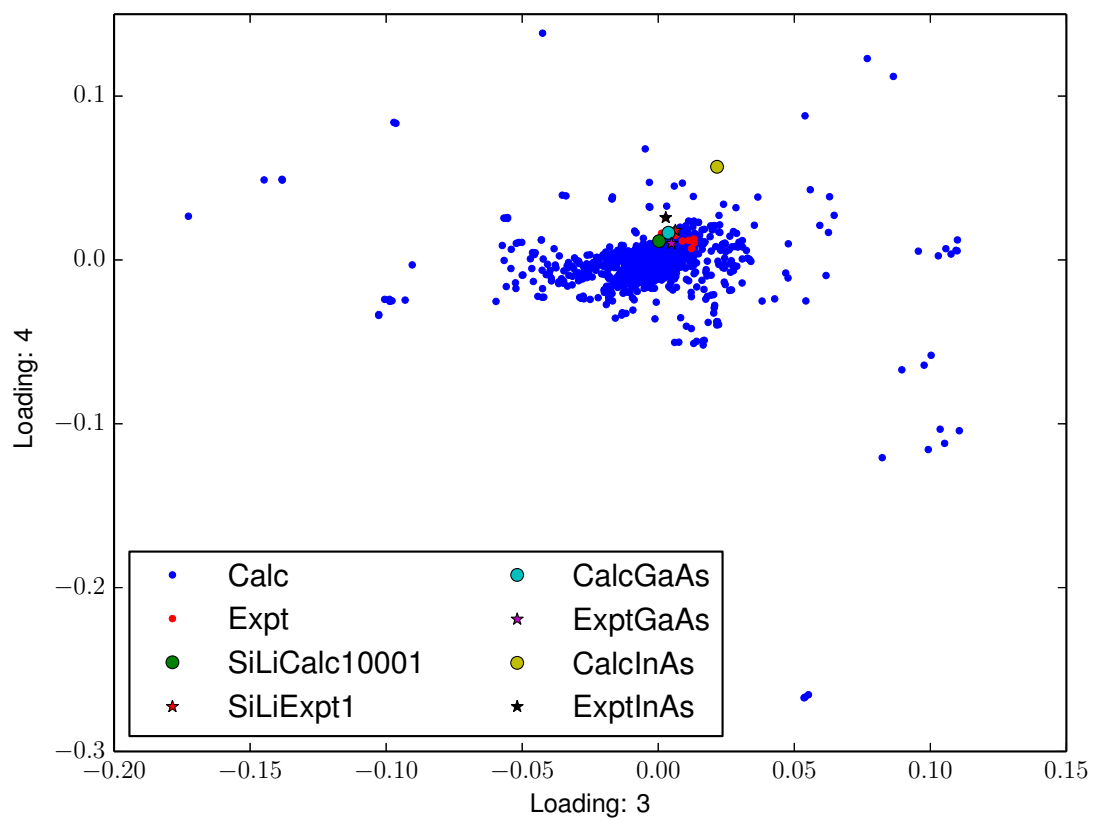


Figure 17: Loading 3 vs Loading 4

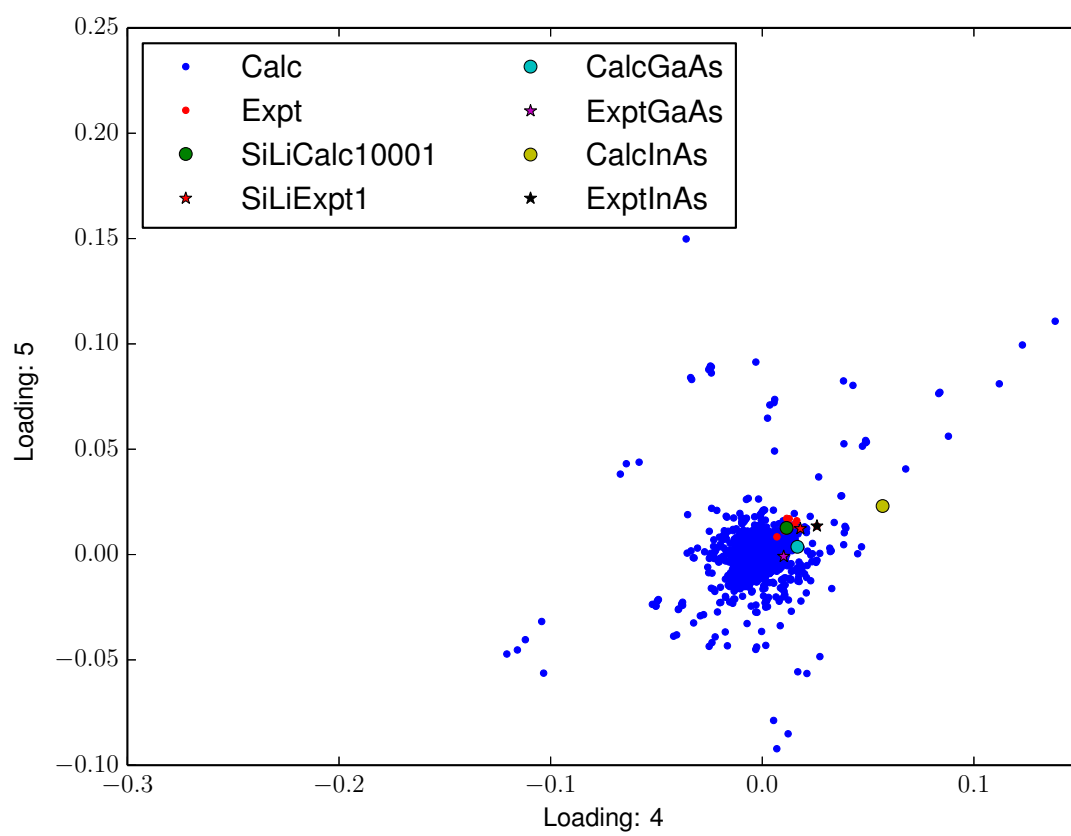


Figure 18: Loading 4 vs Loading 5

### 4.4.1 Eigenspace Outliers

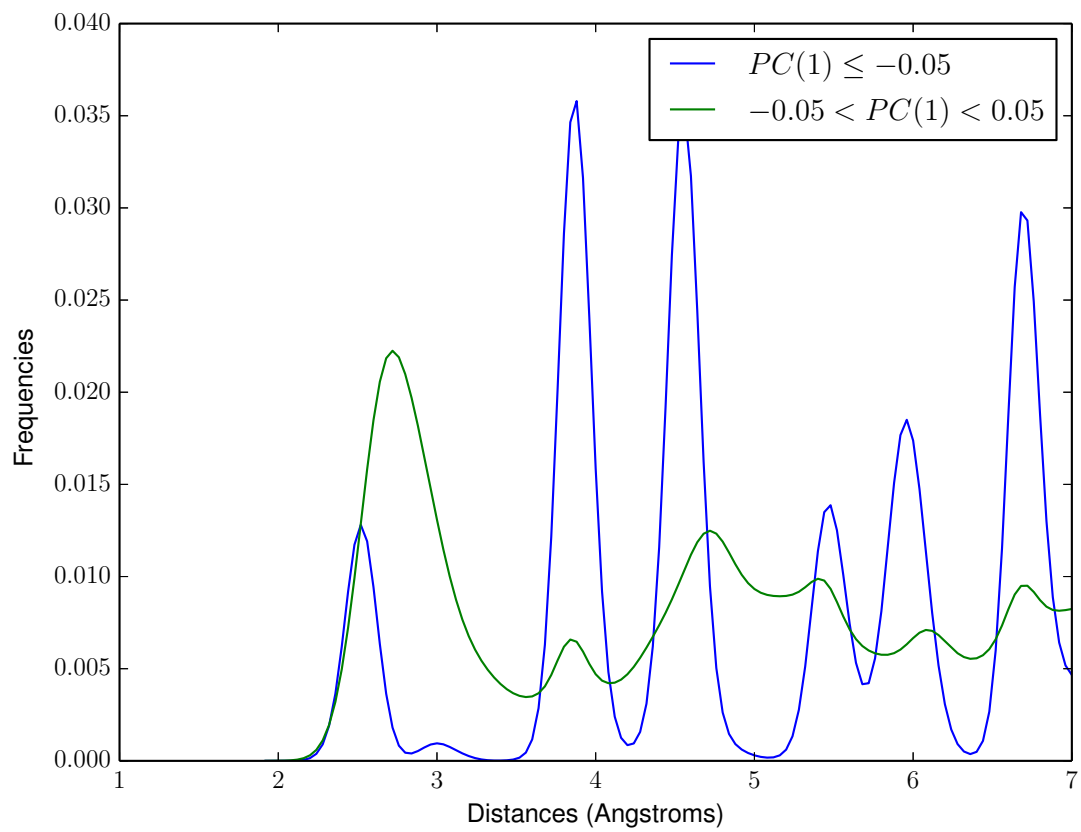


Figure 19: First Principal Component Outliers

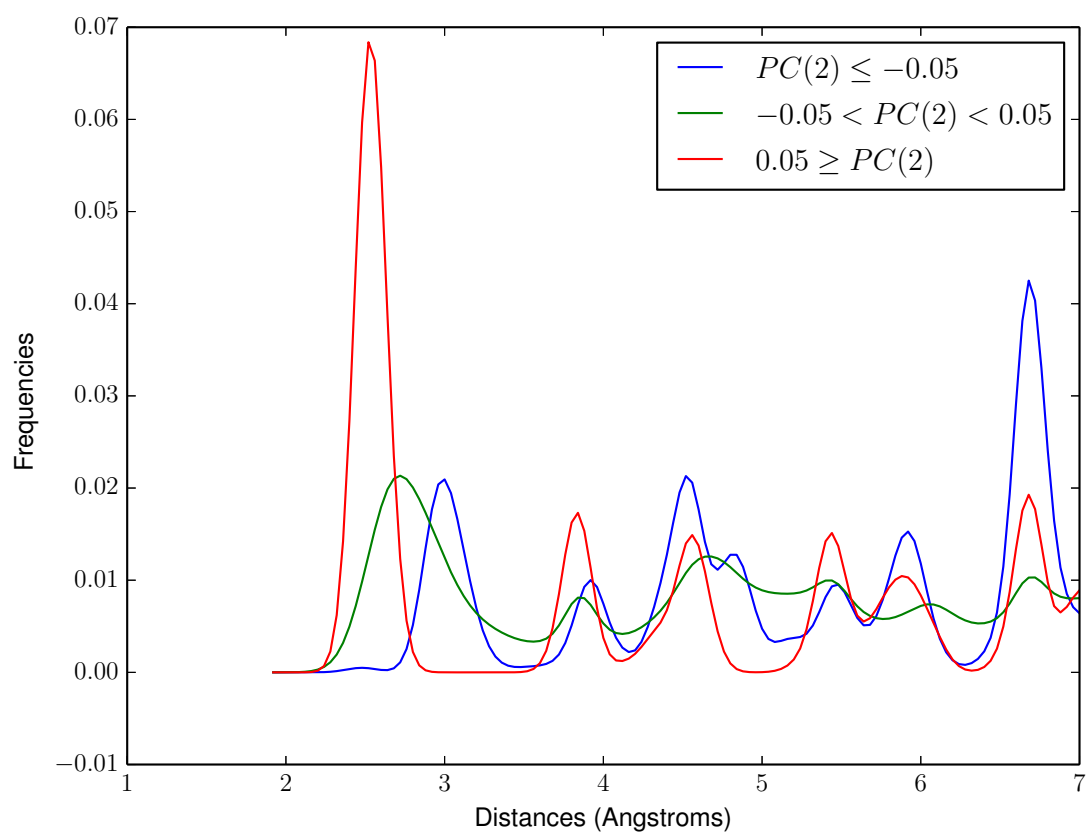


Figure 20: Second Principal Component Outliers



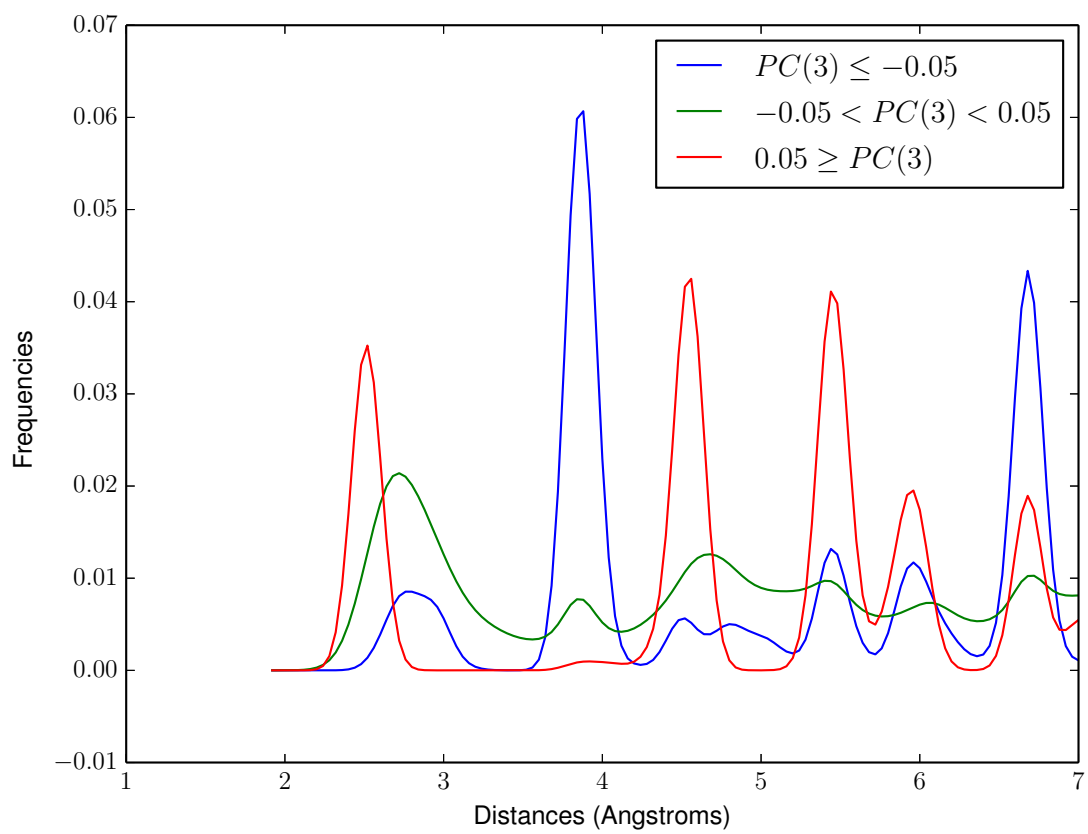


Figure 21: Third Principal Component Outliers

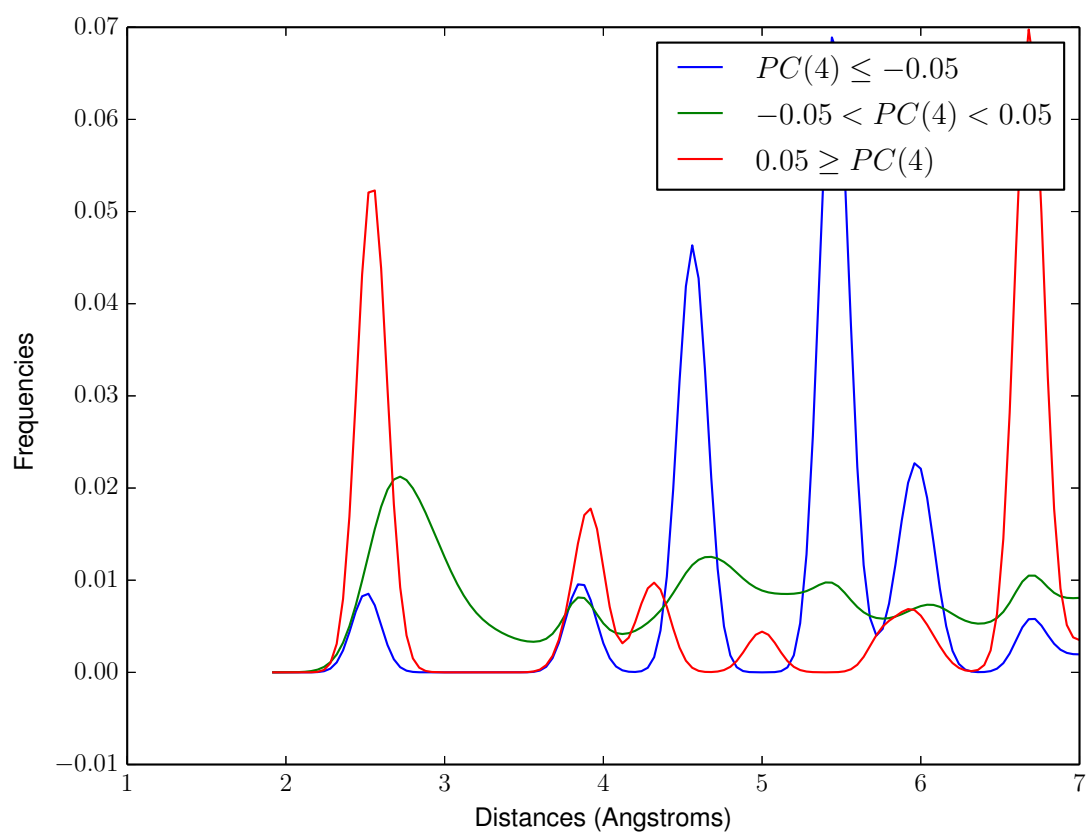


Figure 22: Fourth Principal Component Outliers

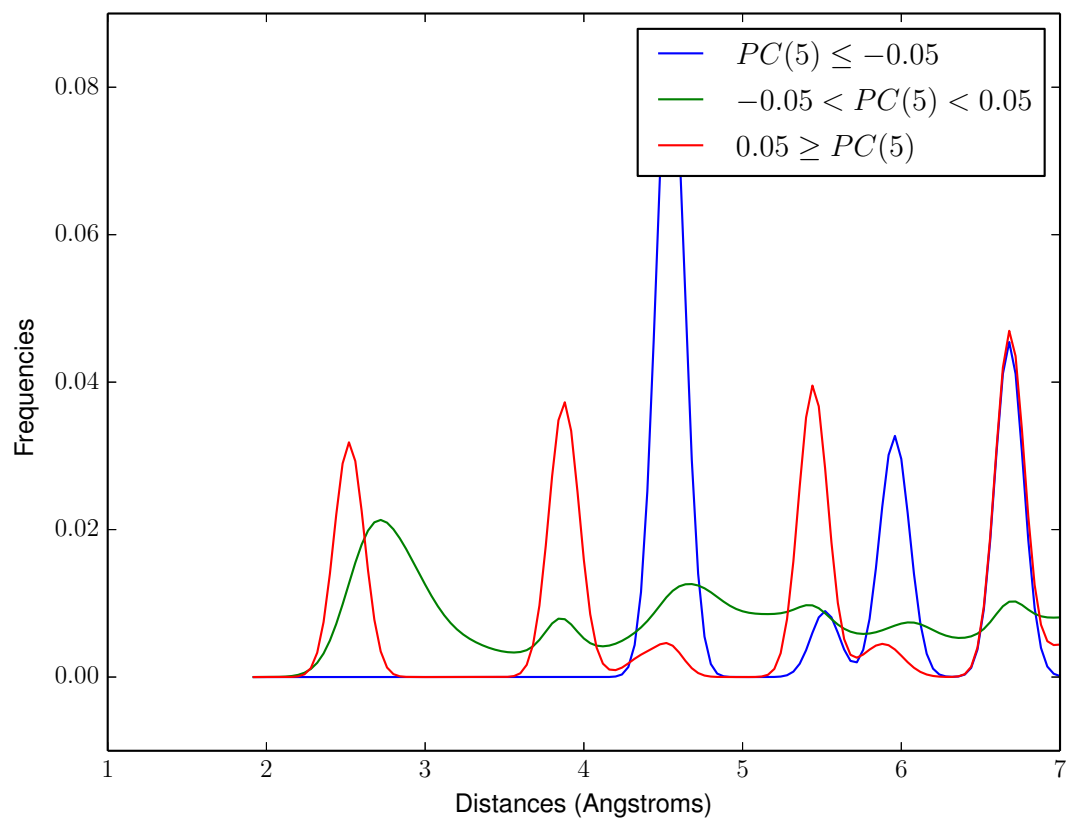


Figure 23: Fifth Principal Component Outliers

## 4.5 Experimental Image Recognition

### 4.5.1 3 Principal Components

<b>Image</b>	<b>Best Match</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>ExptGaAs</b>	SiLiCalc11436	SiLiCalc11634	SiLiCalc11967	SiLiCalc12738	SiLiCalc10225
<b>ExptInAs</b>	SiLiCalc10643	SiLiCalc10560	SiLiCalc10693	SiLiCalc10617	SiLiCalc10621
<b>SiLiExpt1</b>	SiLiCalc10208	SiLiCalc10315	SiLiCalc10317	SiLiCalc10188	SiLiCalc10187
<b>SiLiExpt2</b>	SiLiCalc10317	SiLiCalc10287	SiLiCalc10320	SiLiCalc10283	SiLiCalc10273
<b>SiLiExpt3</b>	SiLiCalc10287	SiLiCalc10239	SiLiCalc10259	SiLiCalc10317	SiLiCalc10232
<b>SiLiExpt4</b>	SiLiCalc10229	SiLiCalc10225	SiLiCalc10232	SiLiCalc10239	SiLiCalc10259
<b>SiLiExpt5</b>	SiLiCalc10225	SiLiCalc10256	SiLiCalc10232	SiLiCalc10229	SiLiCalc10231
<b>SiLiExpt6</b>	SiLiCalc10322	SiLiCalc10225	SiLiCalc10247	SiLiCalc10229	SiLiCalc10256
<b>SiLiExpt7</b>	SiLiCalc10225	SiLiCalc10322	SiLiCalc10256	SiLiCalc10247	SiLiCalc10229
<b>SiLiExpt8</b>	SiLiCalc10225	SiLiCalc10322	SiLiCalc10247	SiLiCalc10337	SiLiCalc10256

Table 3: Recognition with 3 Principal Components

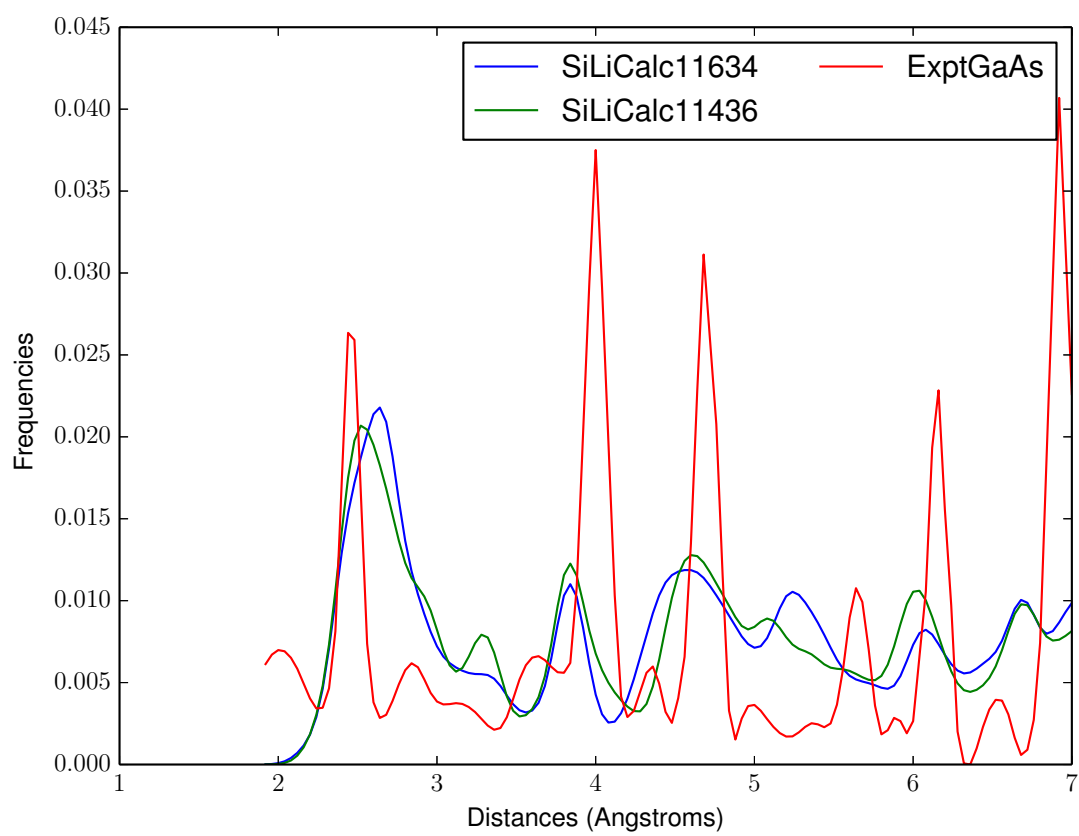


Figure 24: PCA Matches: ExptGaAs, SiLiCalc11436, SiLiCalc11634

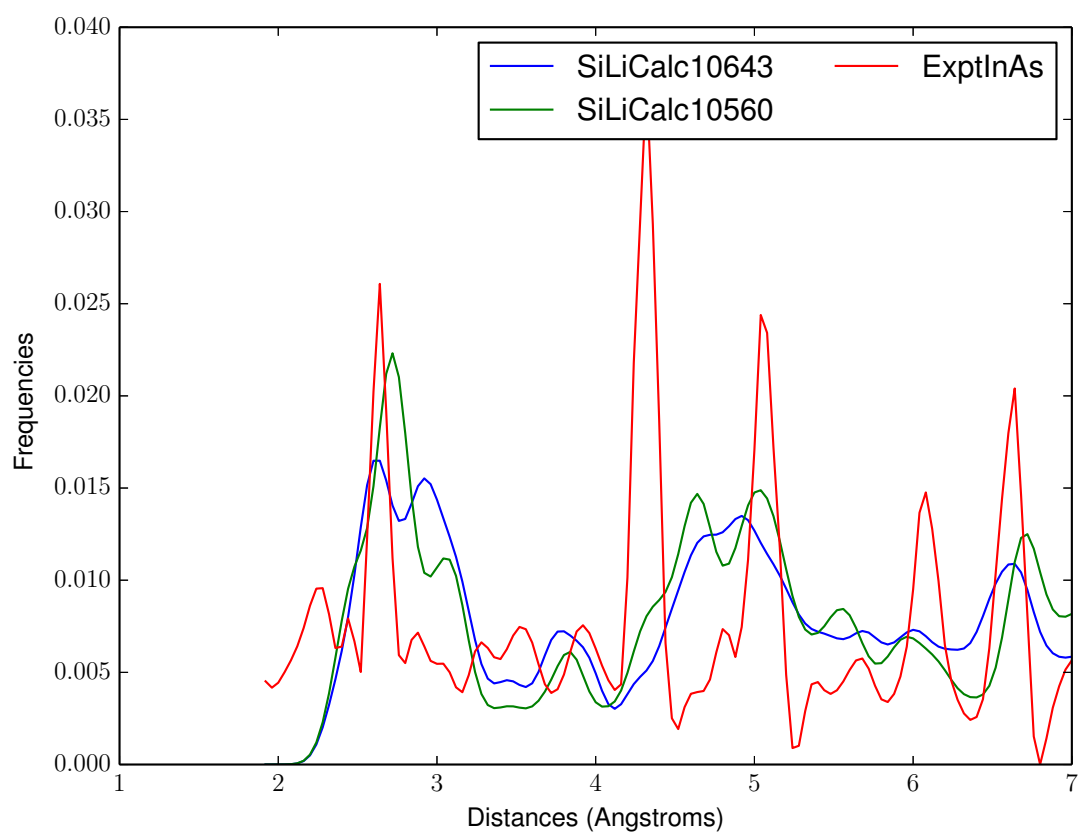


Figure 25: PCA Matches: ExptInAs, SiLiCalc10643, SiLiCalc10560

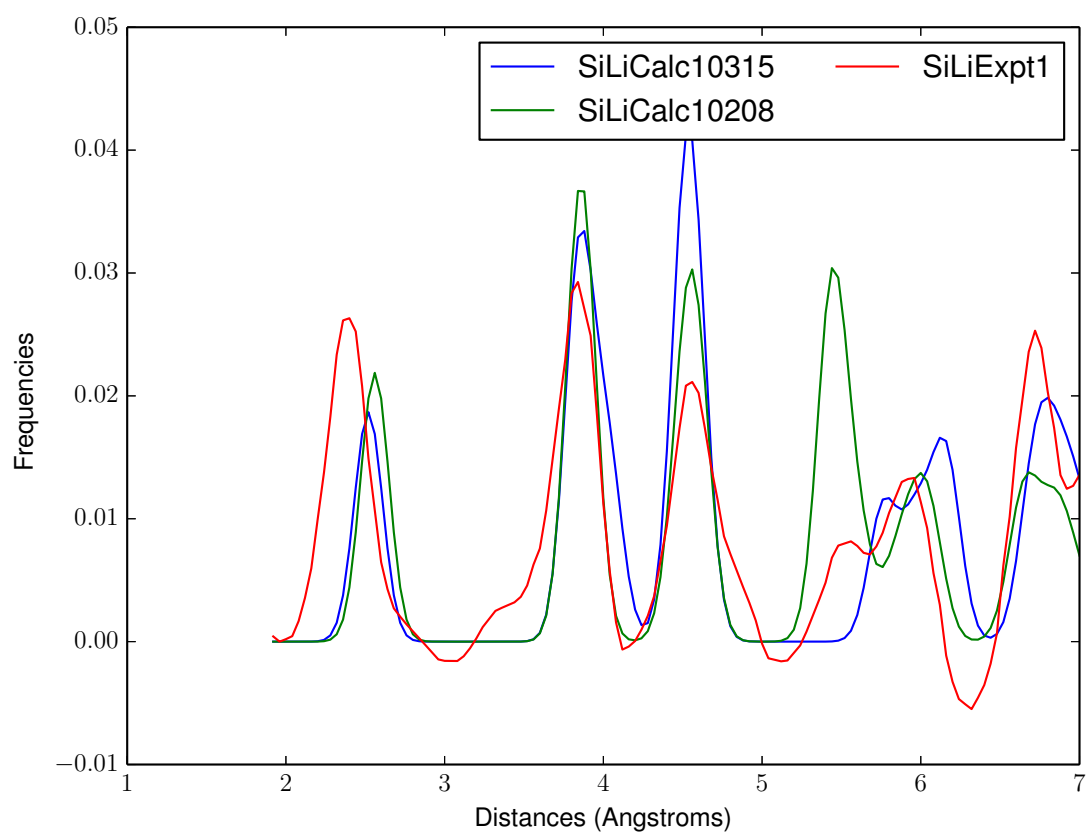


Figure 26: PCA Matches: SiLiExpt1, SiLiCalc10208, SiLiCalc10315

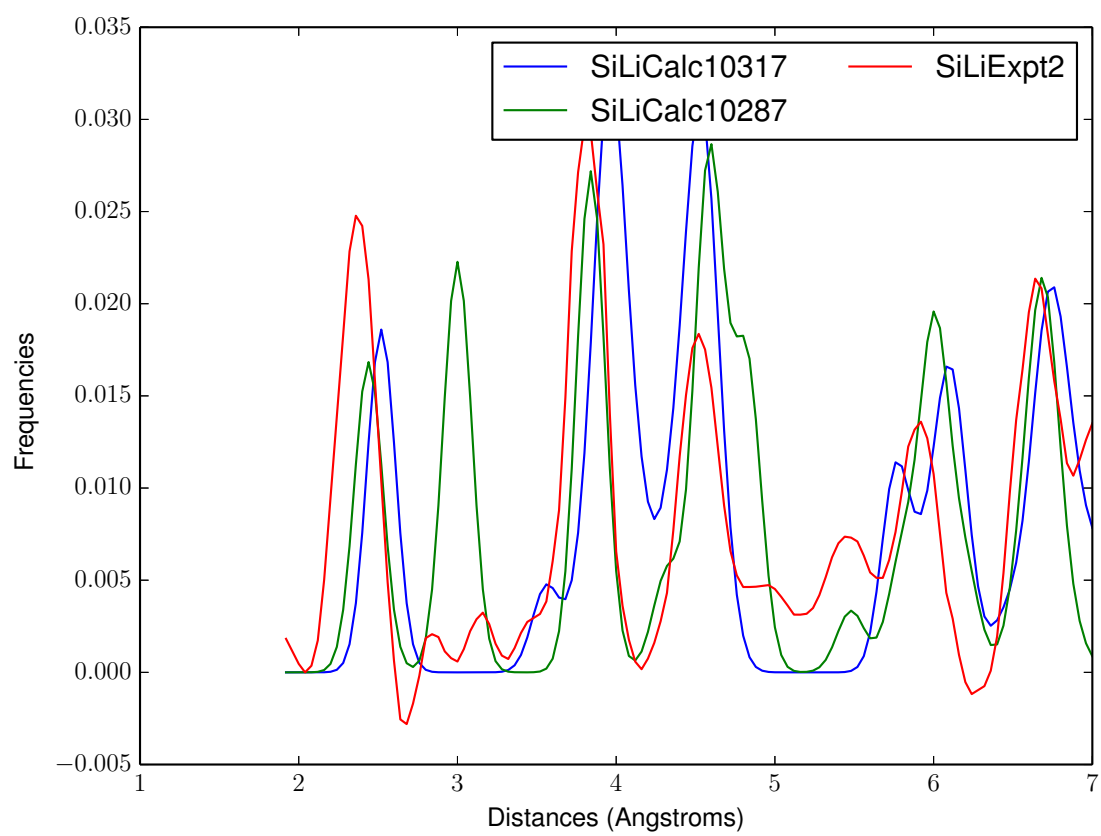


Figure 27: PCA Matches: SiLiExpt2, SiLiCalc10317, SiLiCalc10287



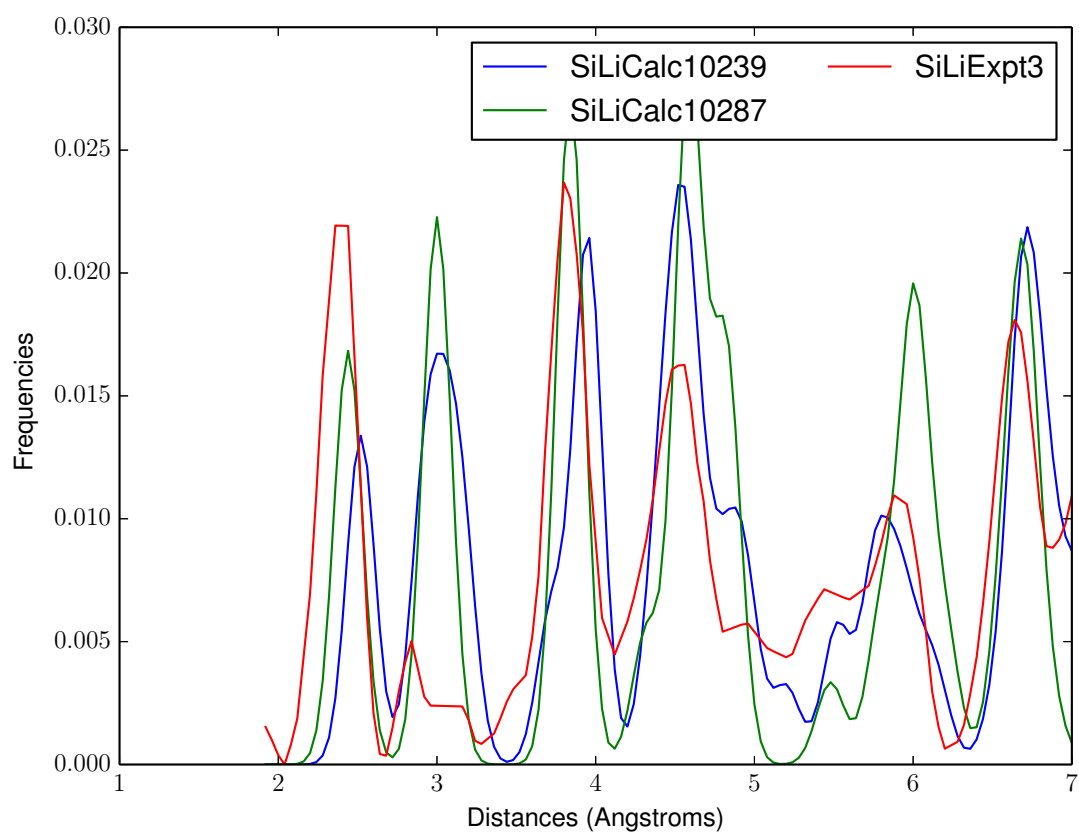


Figure 28: PCA Matches: SiLiExpt3, SiLiCalc10287, SiLiCalc10239

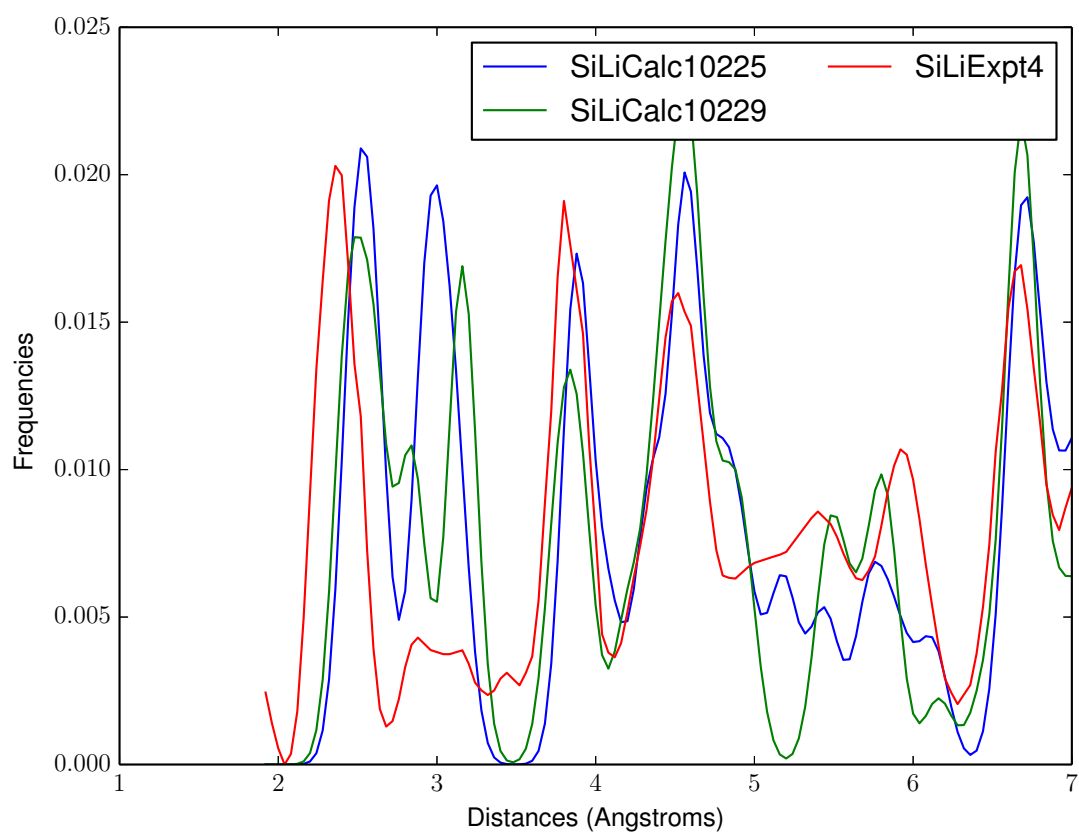


Figure 29: PCA Matches: SiLiExpt4, SiLiCalc10229, SiLiCalc10225

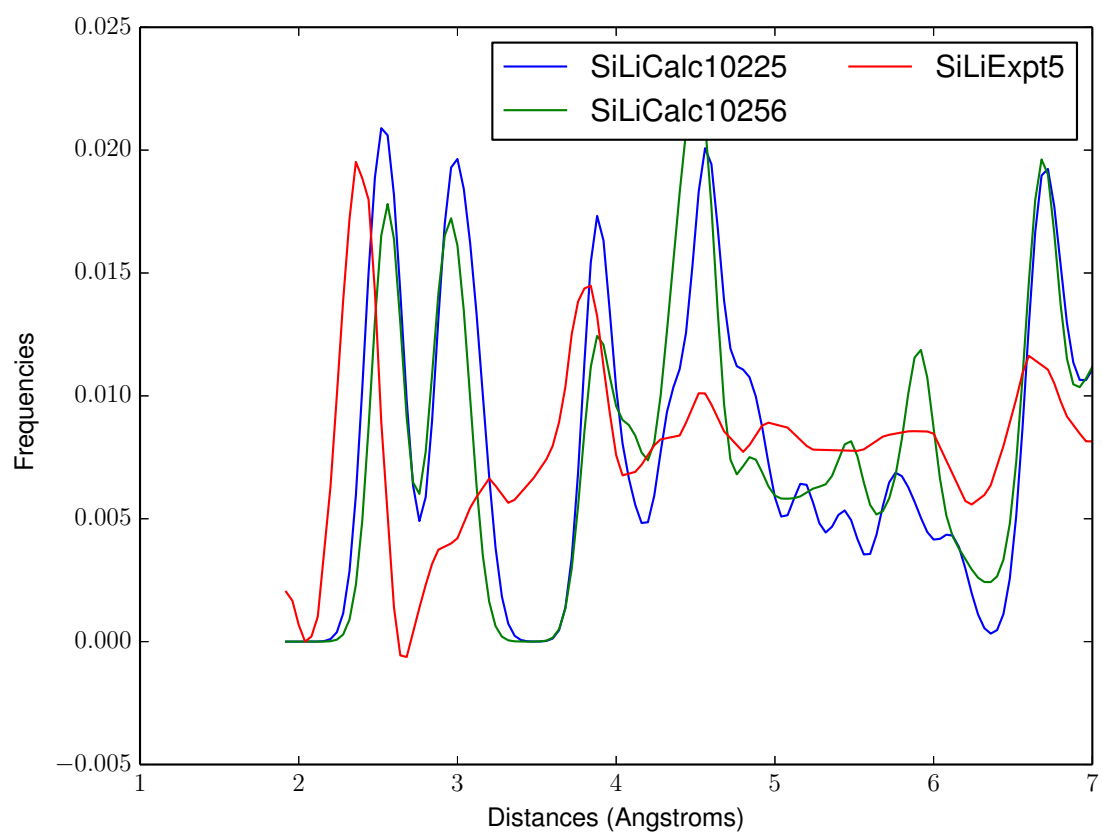


Figure 30: PCA Matches: SiLiExpt5, SiLiCalc10225, SiLiCalc10256

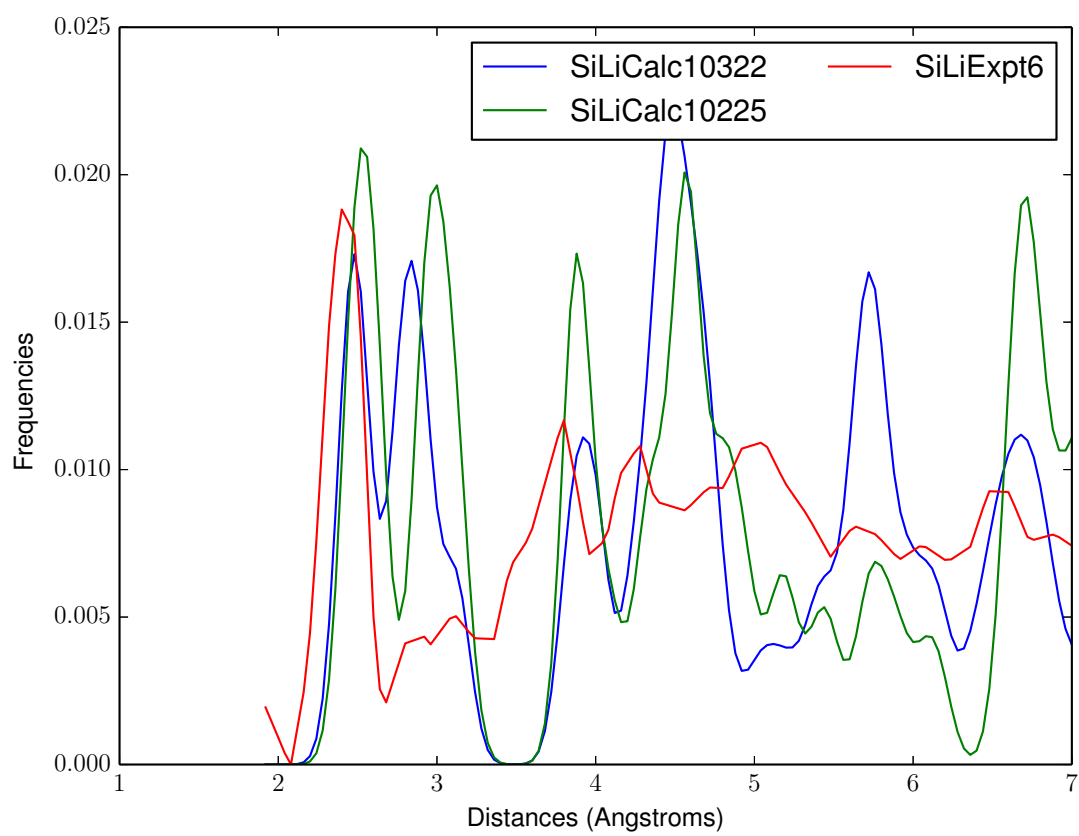


Figure 31: PCA Matches: SiLiExpt6, SiLiCalc10322, SiLiCalc10225

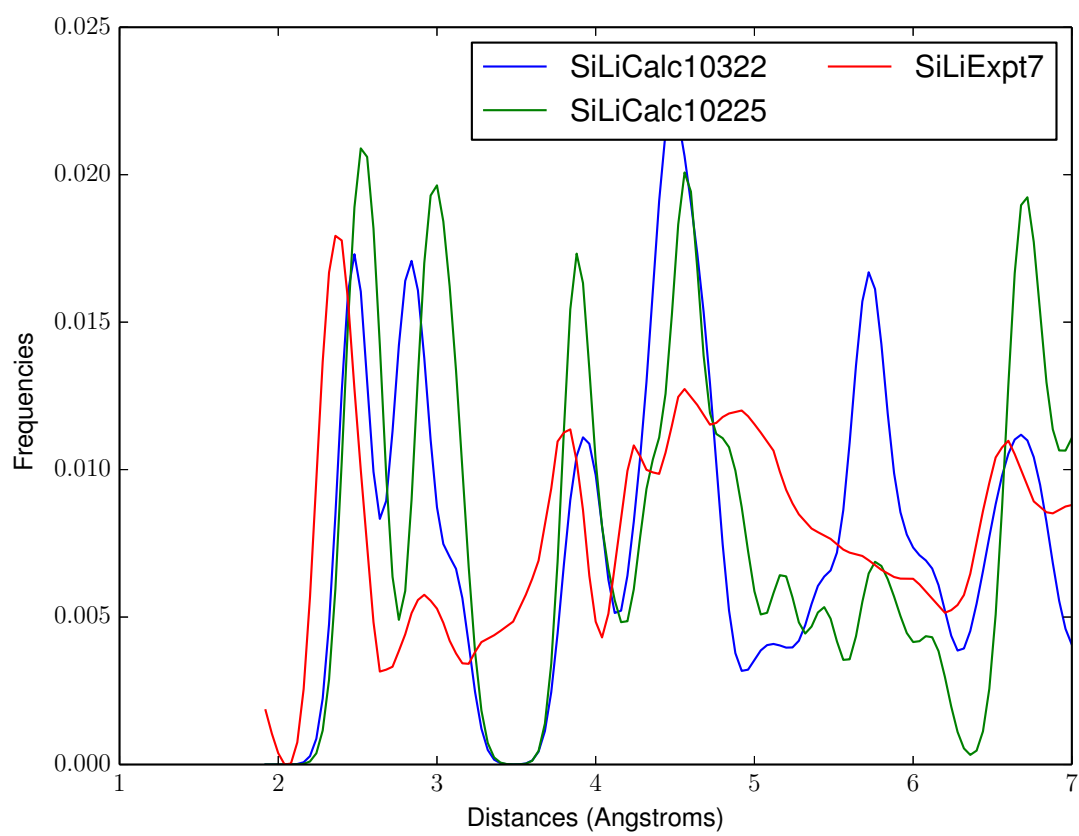


Figure 32: PCA Matches: SiLiExpt7, SiLiCalc10225, SiLiCalc10322

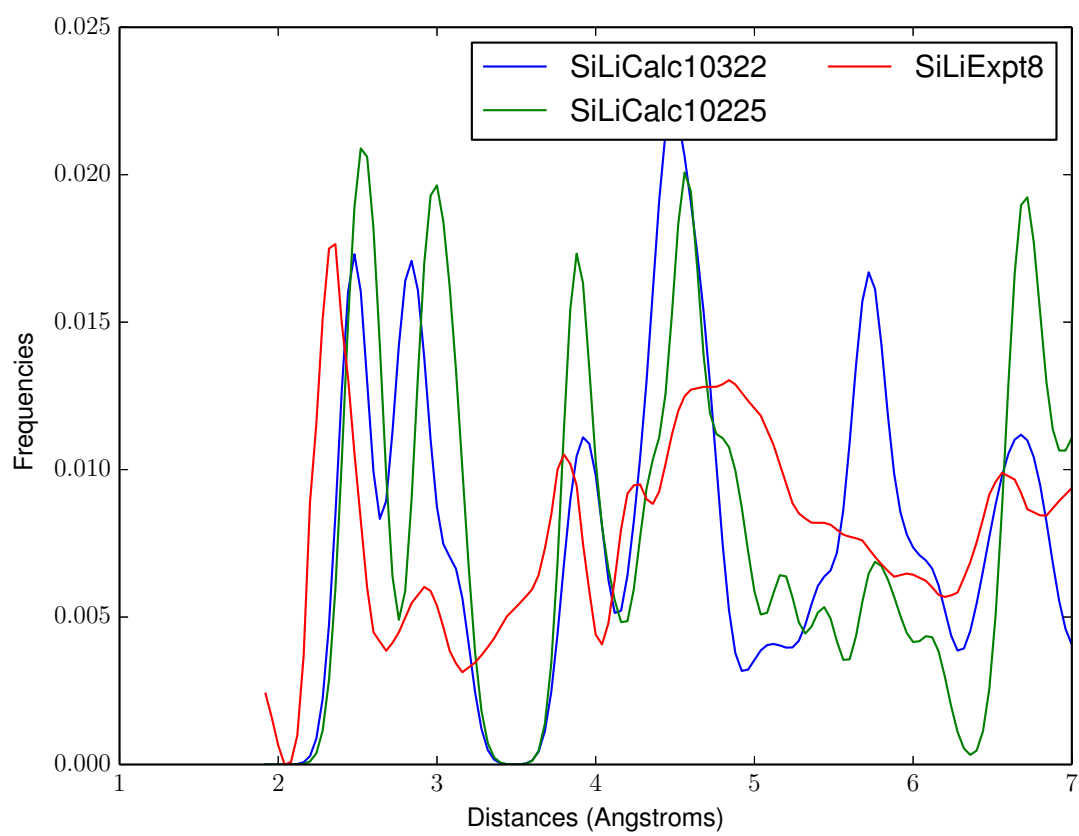


Figure 33: PCA Matches: SiLiExpt8, SiLiCalc10225, SiLiCalc10322

#### 4.5.2 10 Principal Components

<b>Image</b>	<b>Best Match</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>ExptGaAs</b>	<b>CalcGaAs</b>	SiLiCalc10329	SiLiCalc11337	SiLiCalc11436	SiLiCalc10571
<b>ExptInAs</b>	SiLiCalc10646	SiLiCalc10805	SiLiCalc10792	SiLiCalc10836	SiLiCalc10767
<b>SiLiExpt1</b>	SiLiCalc10213	SiLiCalc10215	<b>SiLiCalc10001</b>	SiLiCalc10003	SiLiCalc10313
<b>SiLiExpt2</b>	SiLiCalc10001	SiLiCalc10003	SiLiCalc10209	SiLiCalc10317	SiLiCalc10313
<b>SiLiExpt3</b>	SiLiCalc10257	SiLiCalc10317	SiLiCalc10259	SiLiCalc10258	SiLiCalc10256
<b>SiLiExpt4</b>	SiLiCalc10257	SiLiCalc10258	SiLiCalc10256	SiLiCalc10229	SiLiCalc10232
<b>SiLiExpt5</b>	SiLiCalc10445	SiLiCalc10616	SiLiCalc11436	SiLiCalc10329	SiLiCalc11337
<b>SiLiExpt6</b>	SiLiCalc10445	SiLiCalc10616	SiLiCalc11436	SiLiCalc10693	SiLiCalc11337
<b>SiLiExpt7</b>	SiLiCalc10445	SiLiCalc10693	SiLiCalc11337	SiLiCalc10616	SiLiCalc10482
<b>SiLiExpt8</b>	SiLiCalc10445	SiLiCalc10693	SiLiCalc10329	SiLiCalc11337	SiLiCalc10482

Table 4: Recognition with 10 Principal Components

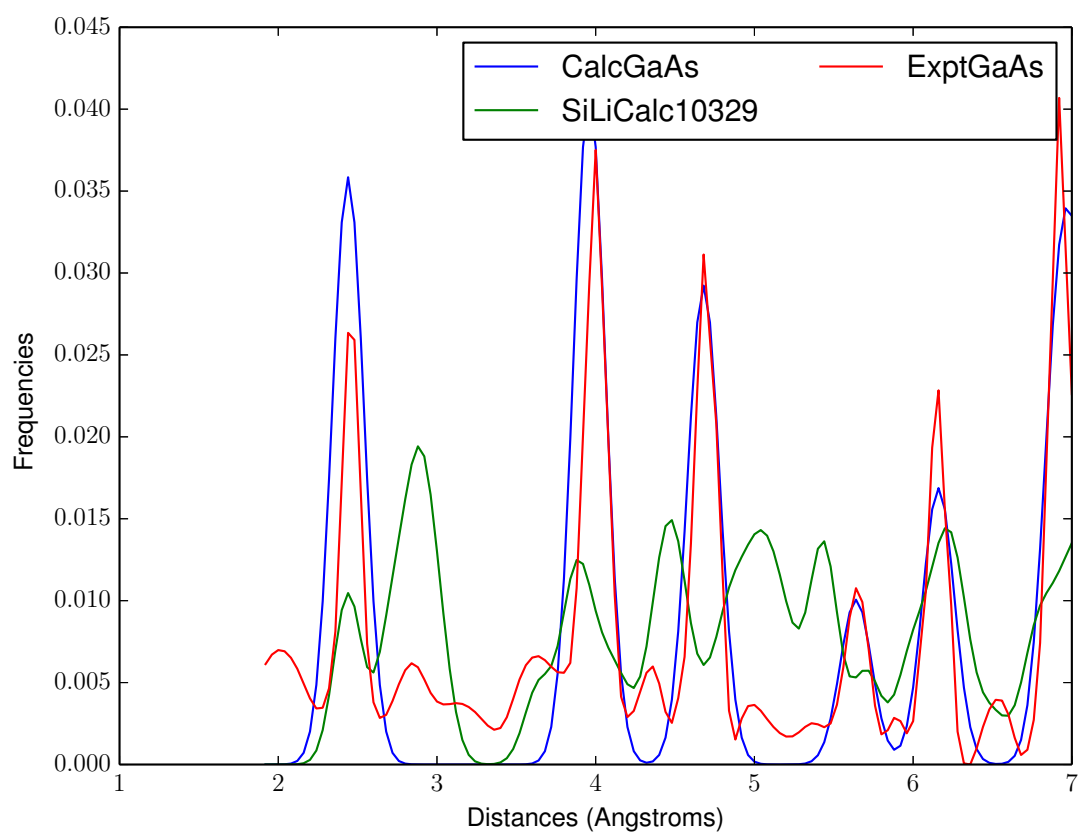


Figure 34: PCA Matches: ExptGaAs, CalcGaAs, SiLiCalc10329



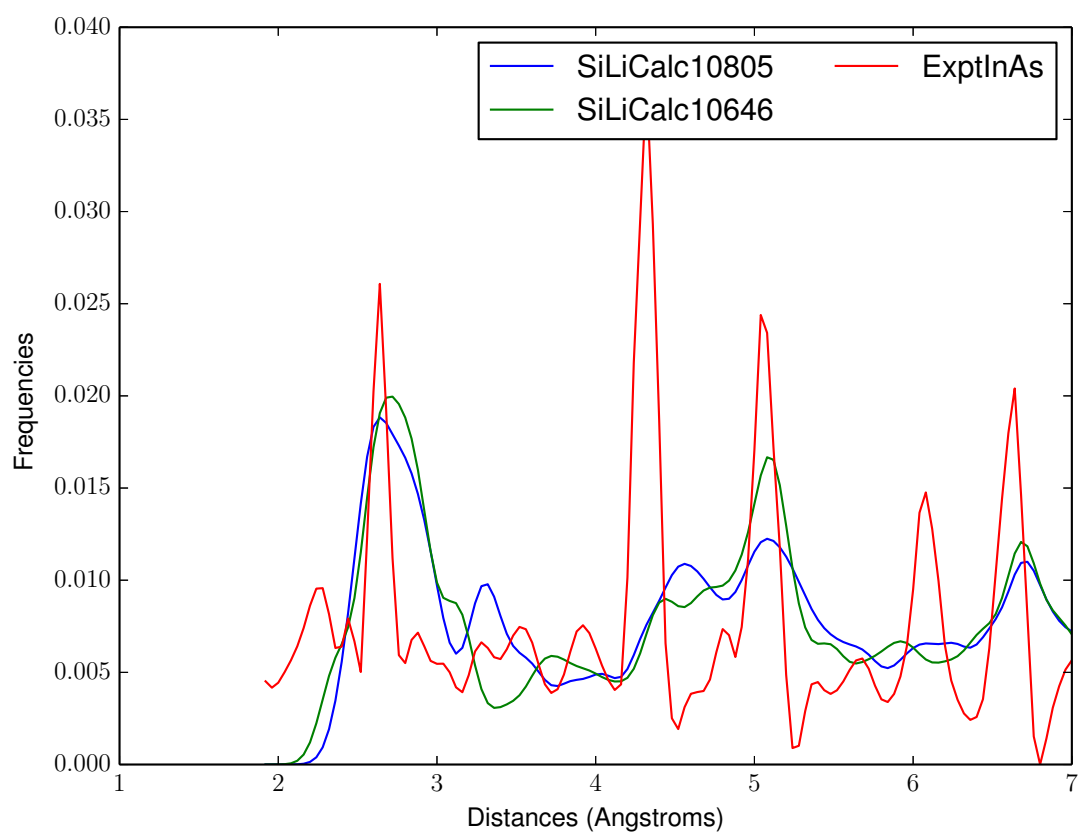


Figure 35: PCA Matches: ExptInAs, SiLiCalc10646, SiLiCalc10805

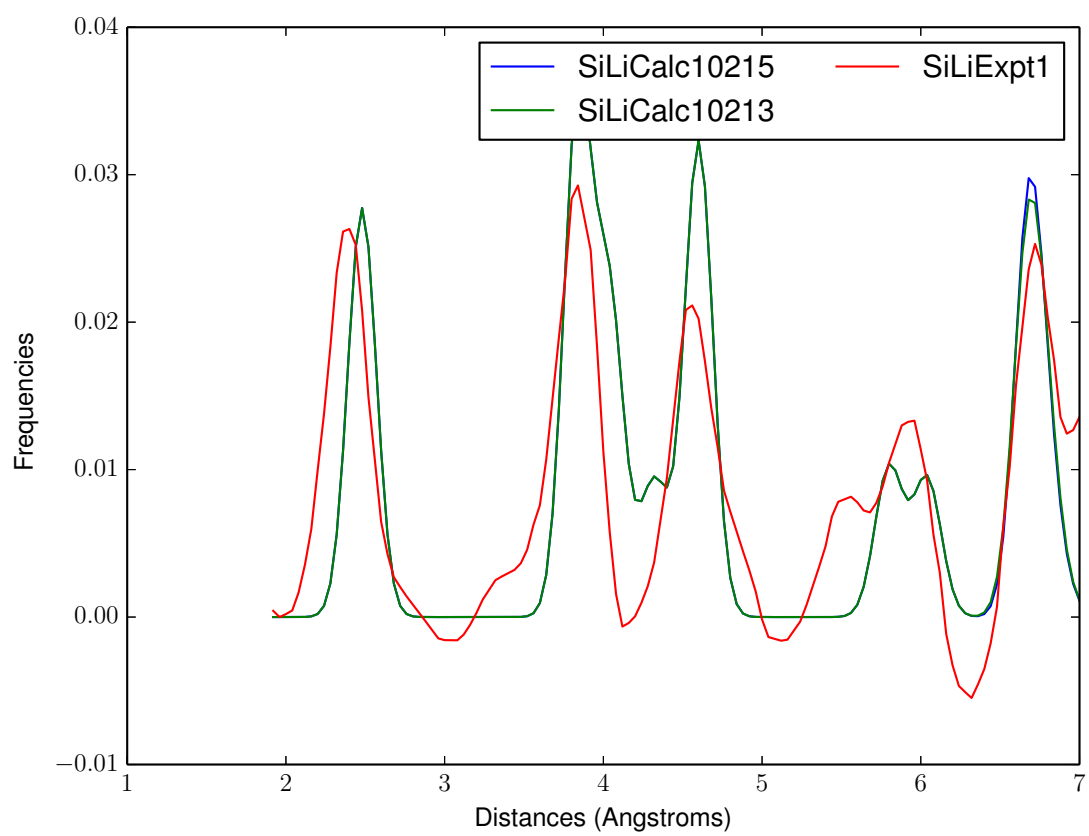


Figure 36: PCA Matches: SiLiExpt1, SiLiCalc10213, SiLiCalc10215

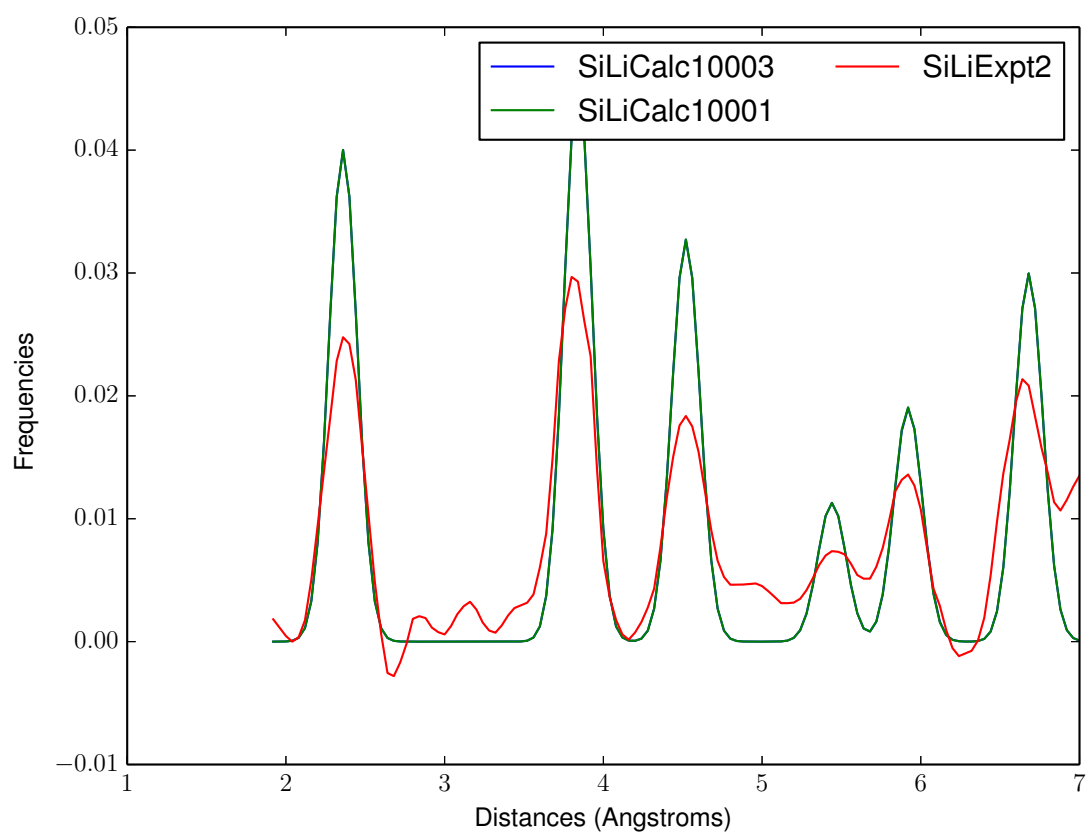


Figure 37: PCA Matches: SiLiExpt2, SiLiCalc10001, SiLiCalc10003

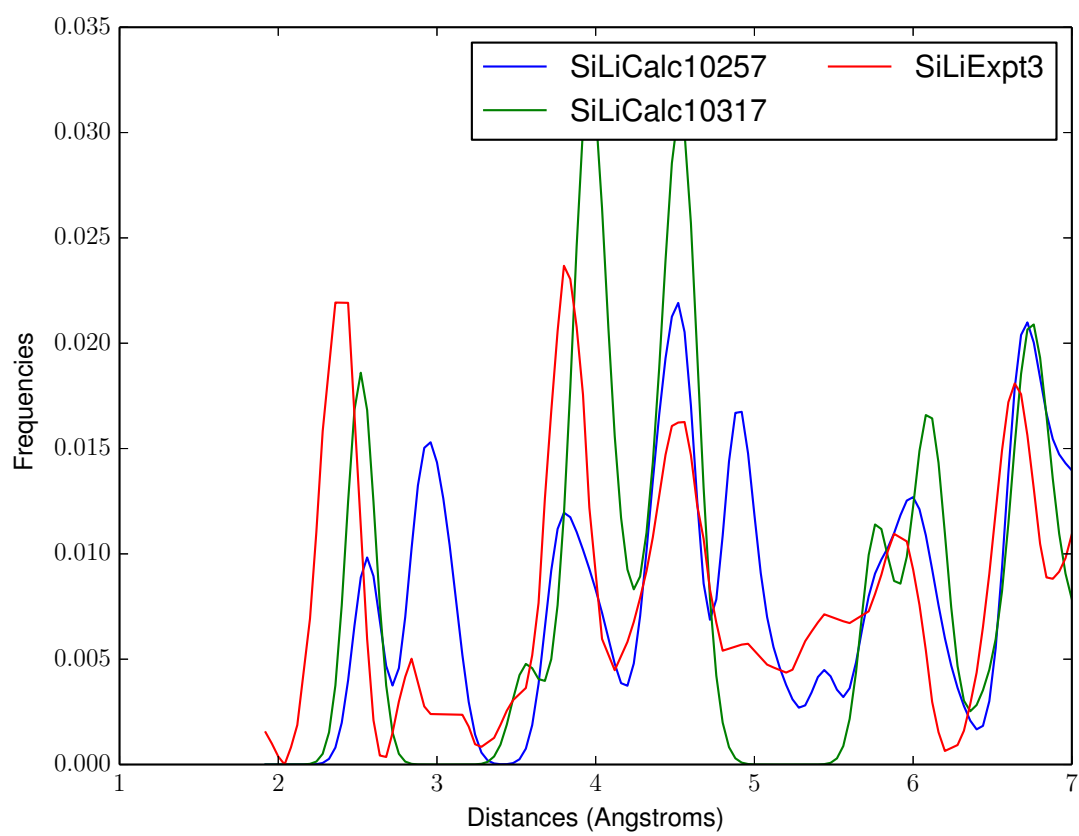


Figure 38: PCA Matches: SiLiExpt3, SiLiCalc10257, SiLiCalc10317

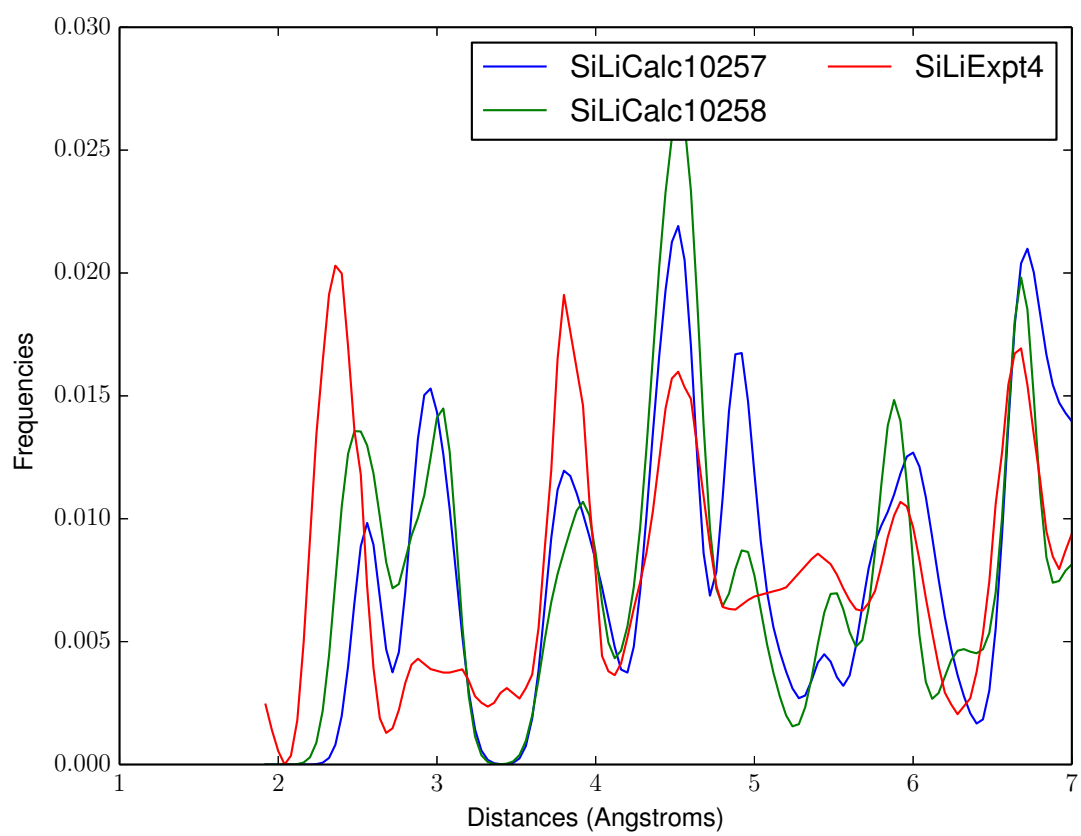


Figure 39: PCA Matches: SiLiExpt4, SiLiCalc10257, SiLiCalc10258

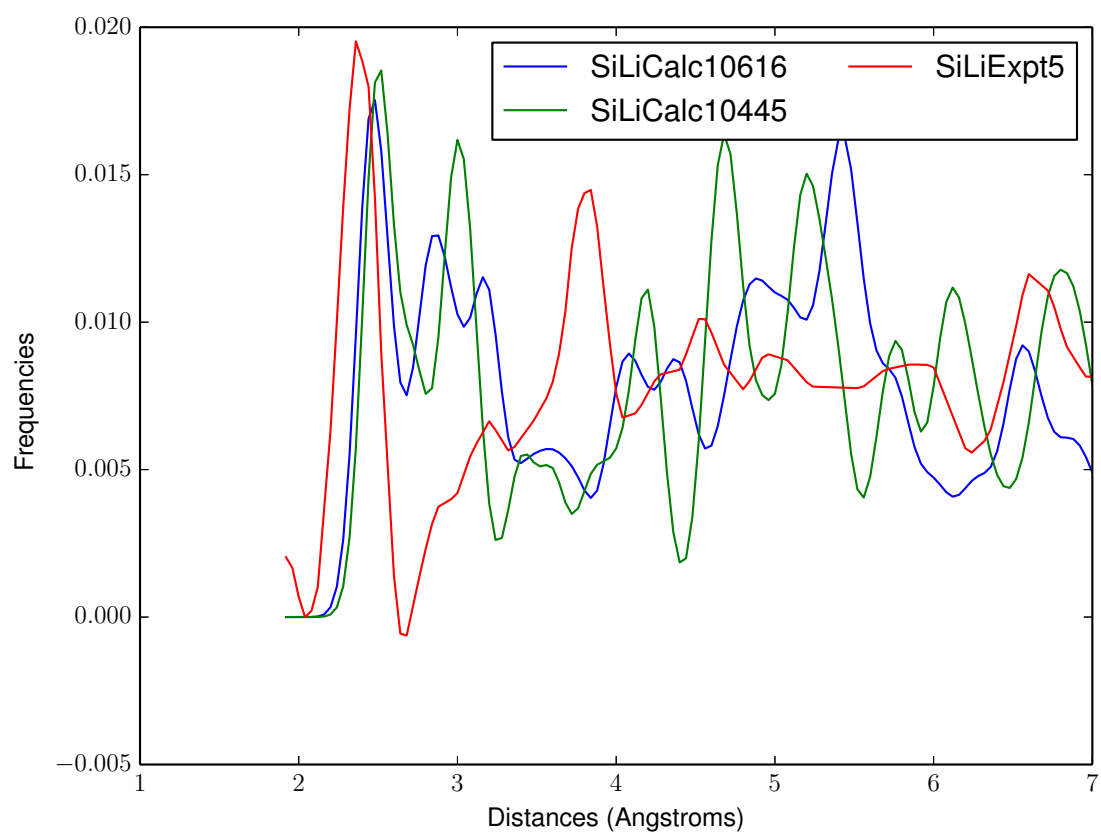


Figure 40: PCA Matches: SiLiExpt5, SiLiCalc10445, SiLiCalc10616

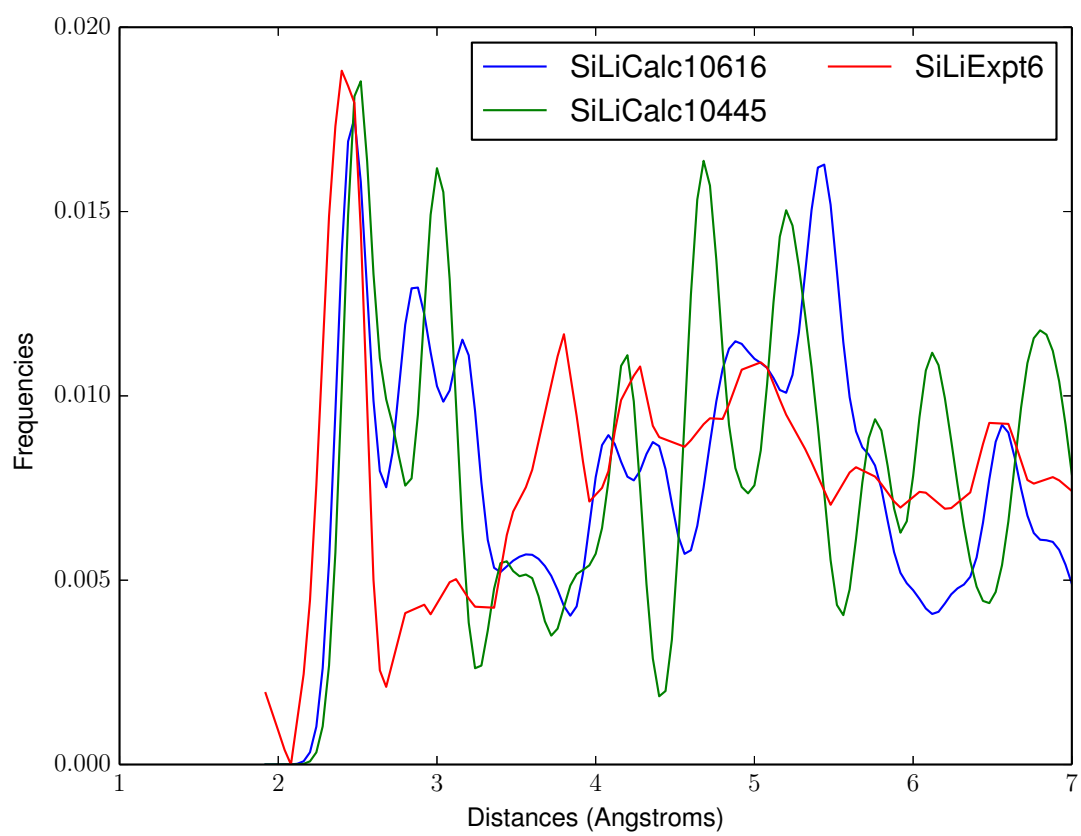


Figure 41: PCA Matches: SiLiExpt6, SiLiCalc10445, SiLiCalc10616

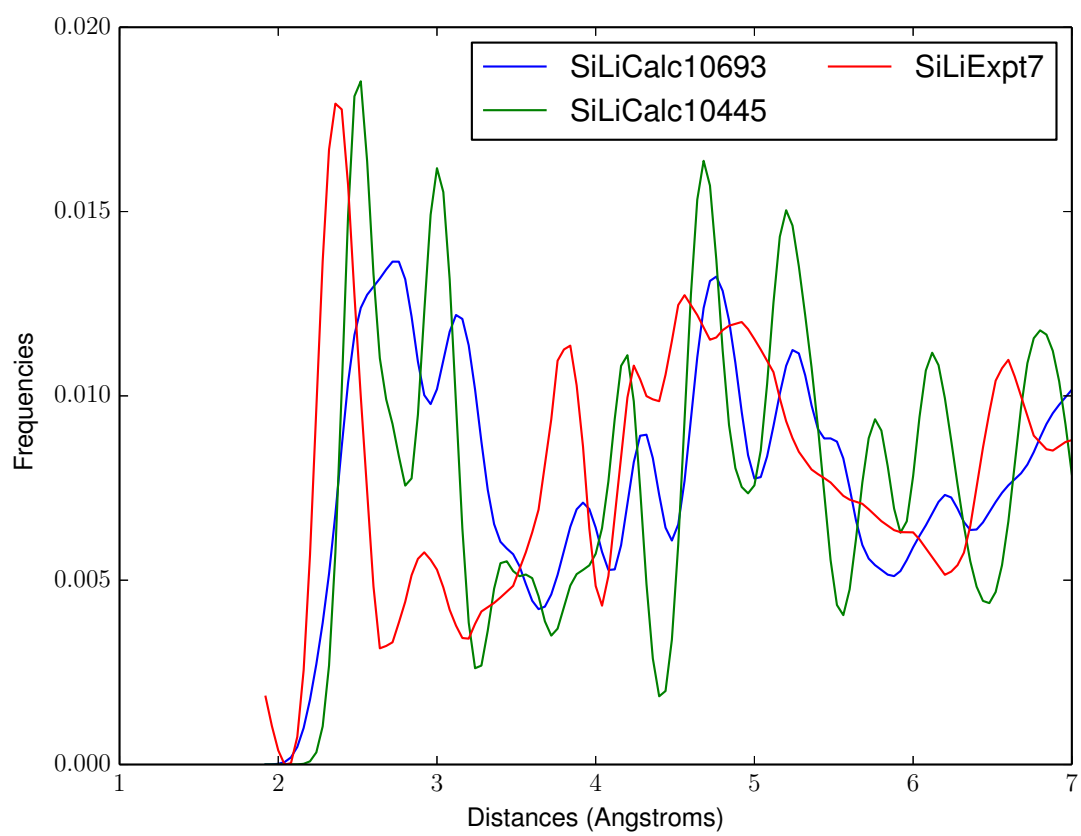


Figure 42: PCA Matches: SiLiExpt7, SiLiCalc10445, SiLiCalc10693



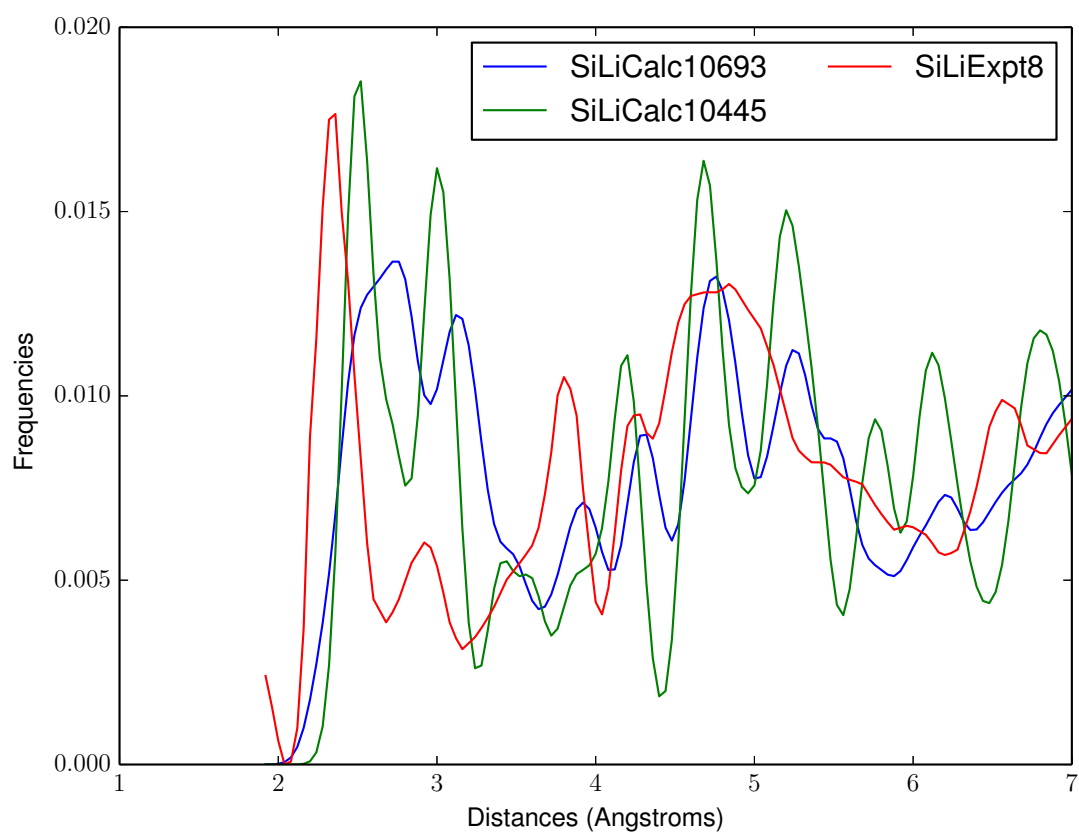


Figure 43: PCA Matches: SiLiExpt8, SiLiCalc10445, SiLiCalc10693

### 4.5.3 128 Principal Components

<b>Image</b>	<b>Best Match</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
<b>ExptGaAs</b>	<b>CalcGaAs</b>	SiLiCalc10445	SiLiCalc11436	SiLiCalc10693	SiLiCalc11337
<b>ExptInAs</b>	SiLiCalc10429	SiLiCalc10602	SiLiCalc10838	SiLiCalc10901	SiLiCalc10607
<b>SiLiExpt1</b>	SiLiCalc10194	<b>SiLiCalc10001</b>	SiLiCalc10003	SiLiCalc10136	SiLiCalc10147
<b>SiLiExpt2</b>	SiLiCalc10001	SiLiCalc10003	SiLiCalc10194	SiLiCalc10136	SiLiCalc10147
<b>SiLiExpt3</b>	SiLiCalc10258	SiLiCalc10229	SiLiCalc10245	SiLiCalc11436	SiLiCalc10259
<b>SiLiExpt4</b>	SiLiCalc10258	SiLiCalc11436	SiLiCalc10229	SiLiCalc11337	SiLiCalc11634
<b>SiLiExpt5</b>	SiLiCalc10616	SiLiCalc11337	SiLiCalc10693	SiLiCalc11436	SiLiCalc11336
<b>SiLiExpt6</b>	SiLiCalc10616	SiLiCalc10693	SiLiCalc11337	SiLiCalc11436	SiLiCalc11336
<b>SiLiExpt7</b>	SiLiCalc10693	SiLiCalc11337	SiLiCalc10482	SiLiCalc10616	SiLiCalc10651
<b>SiLiExpt8</b>	SiLiCalc10693	SiLiCalc10651	SiLiCalc11337	SiLiCalc10482	SiLiCalc10616

Table 5: Recognition with 128 Principal Components

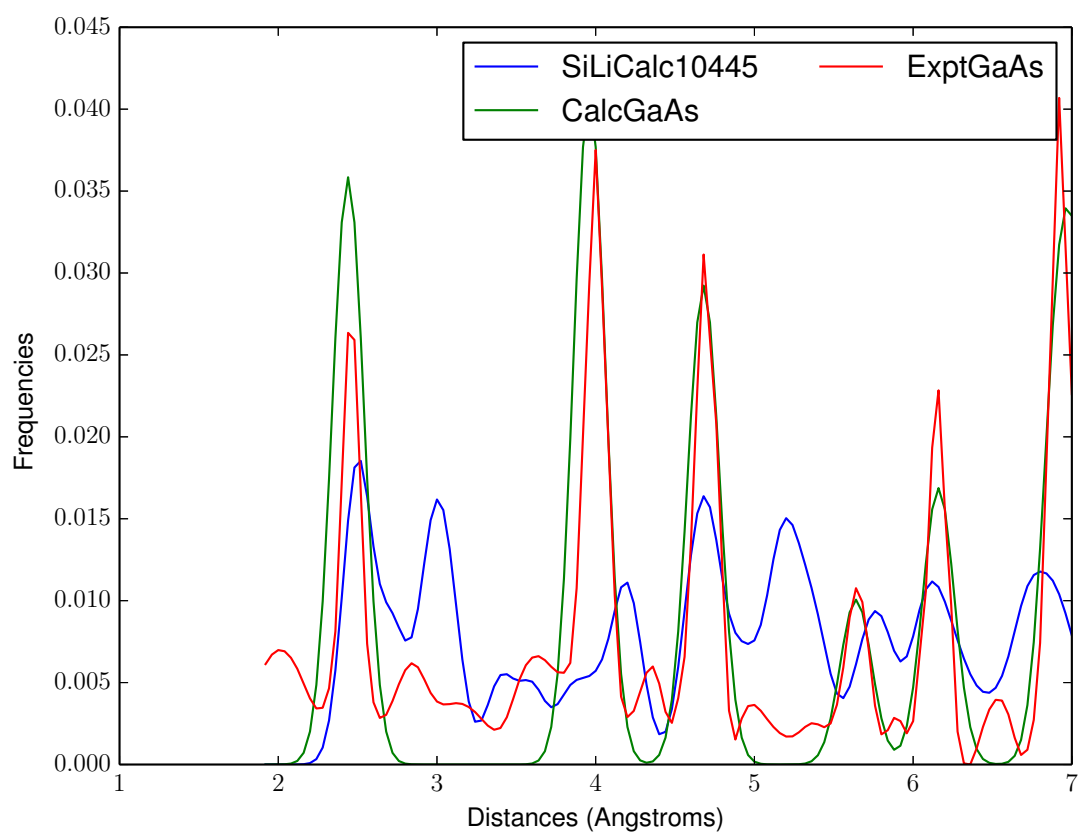


Figure 44: PCA Matches: ExptGaAs, CalcGaAs, SiLiCalc10445

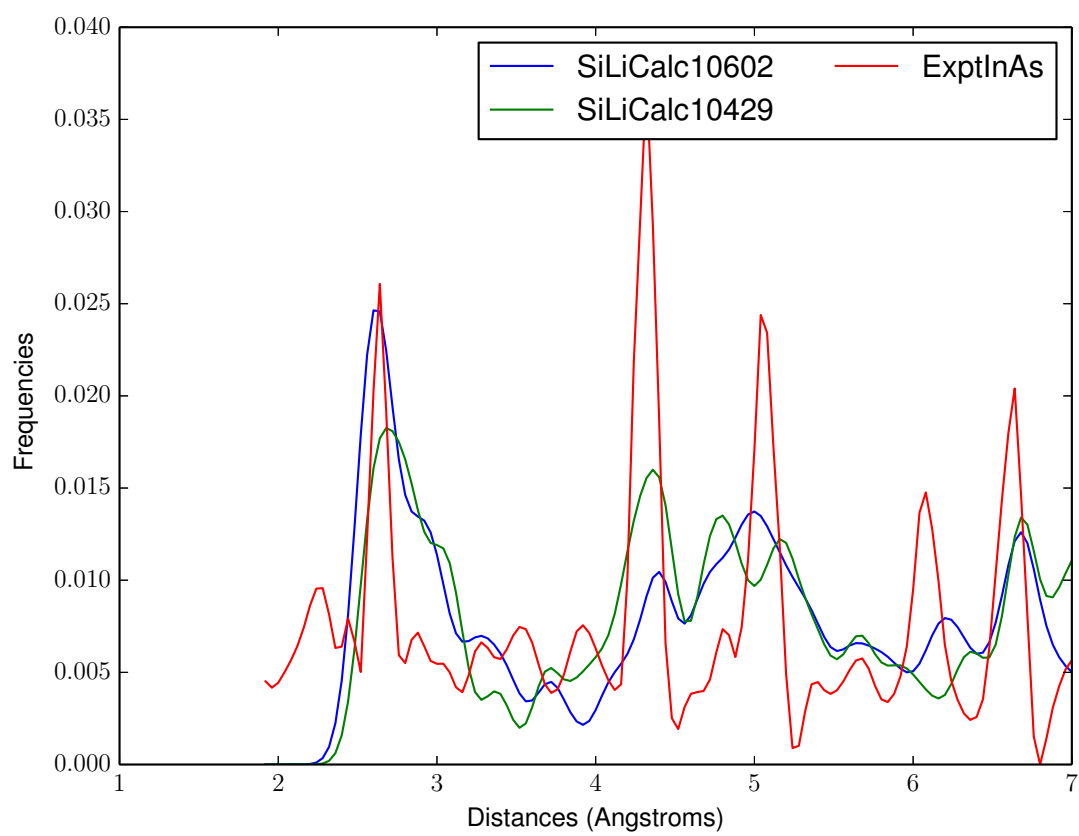


Figure 45: PCA Matches: ExptInAs, SiLiCalc10429, SiLiCalc10602

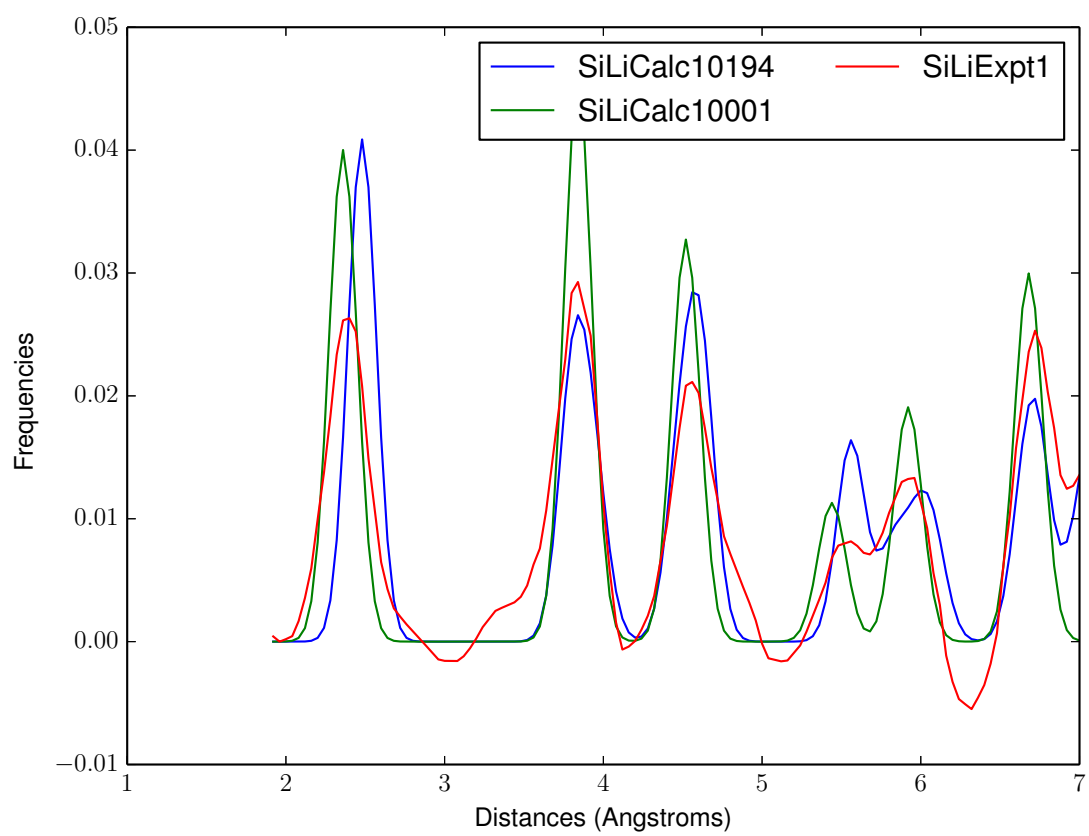


Figure 46: PCA Matches: SiLiExpt1, SiLiCalc10194, SiLiCalc10001

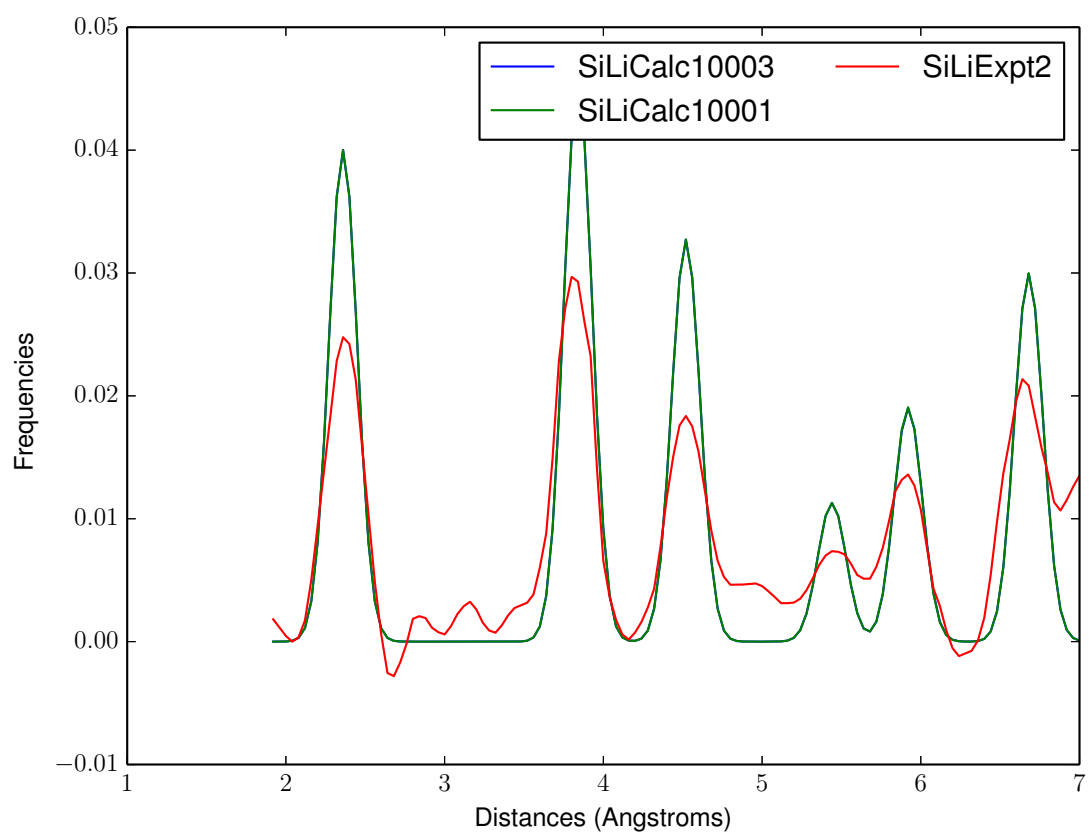


Figure 47: PCA Matches: SiLiExpt2, SiLiCalc10001, SiLiCalc10003

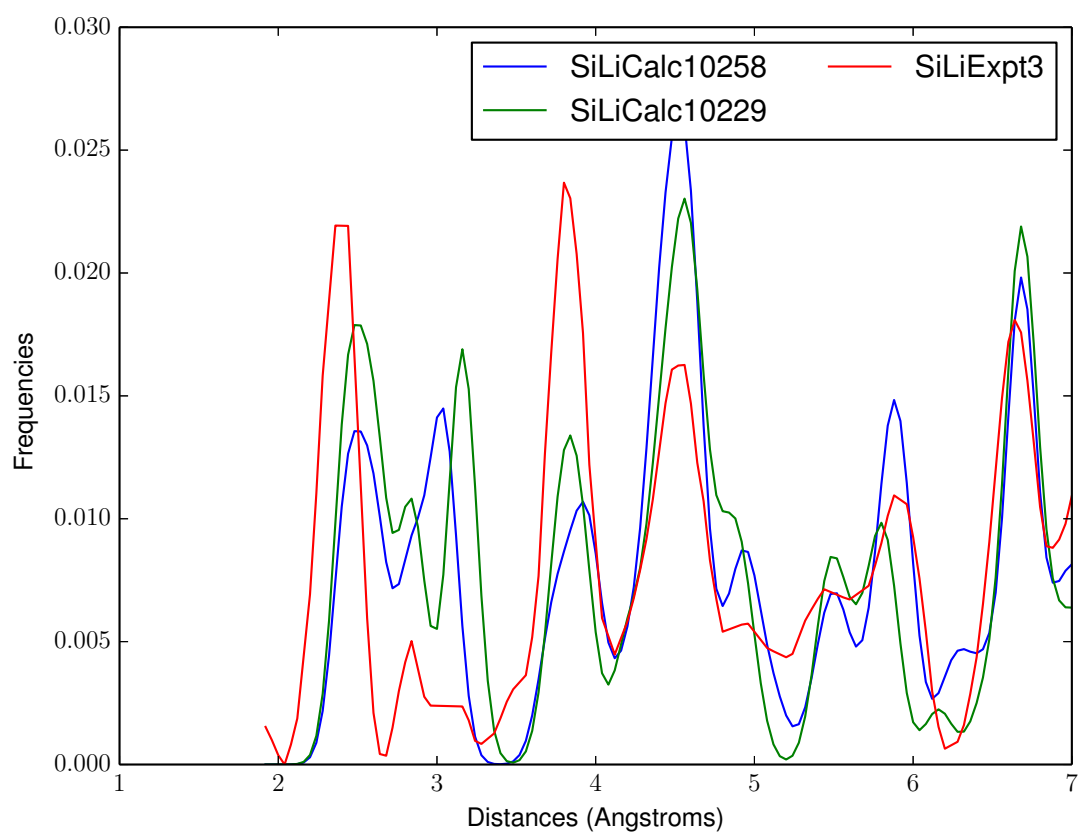


Figure 48: PCA Matches: SiLiExpt3, SiLiCalc10258, SiLiCalc10229

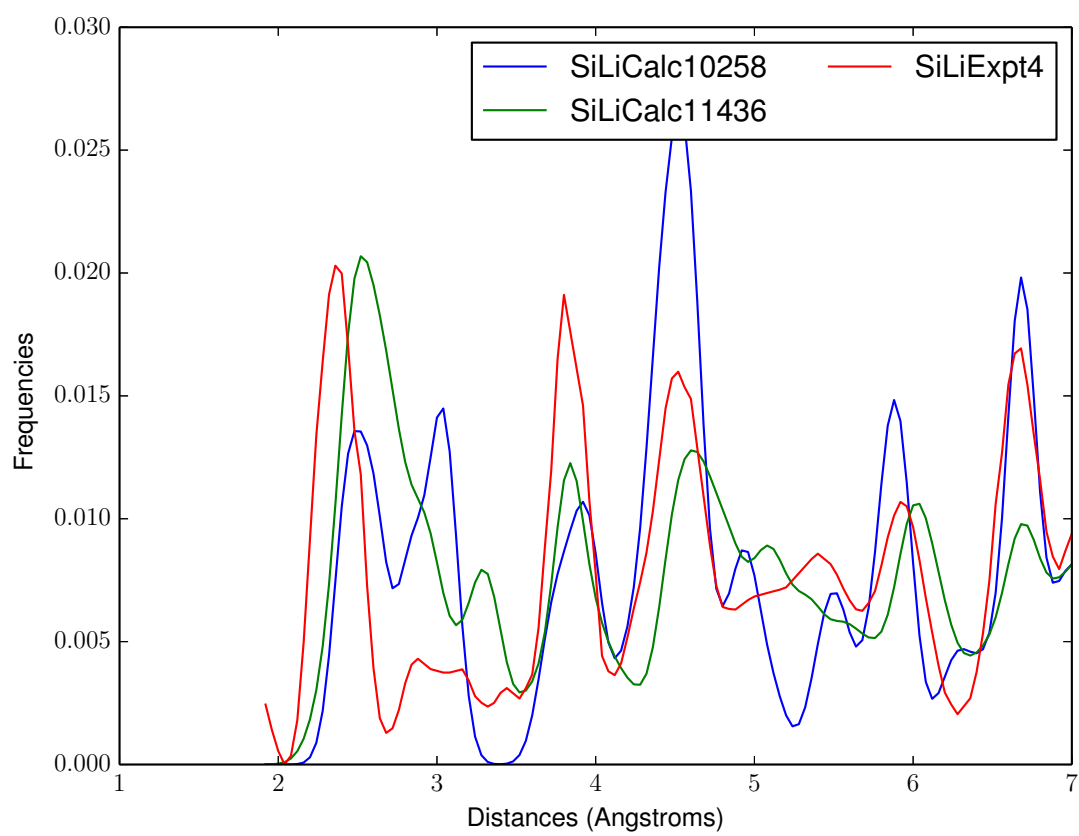


Figure 49: PCA Matches: SiLiExpt4, SiLiCalc10258, SiLiCalc11436



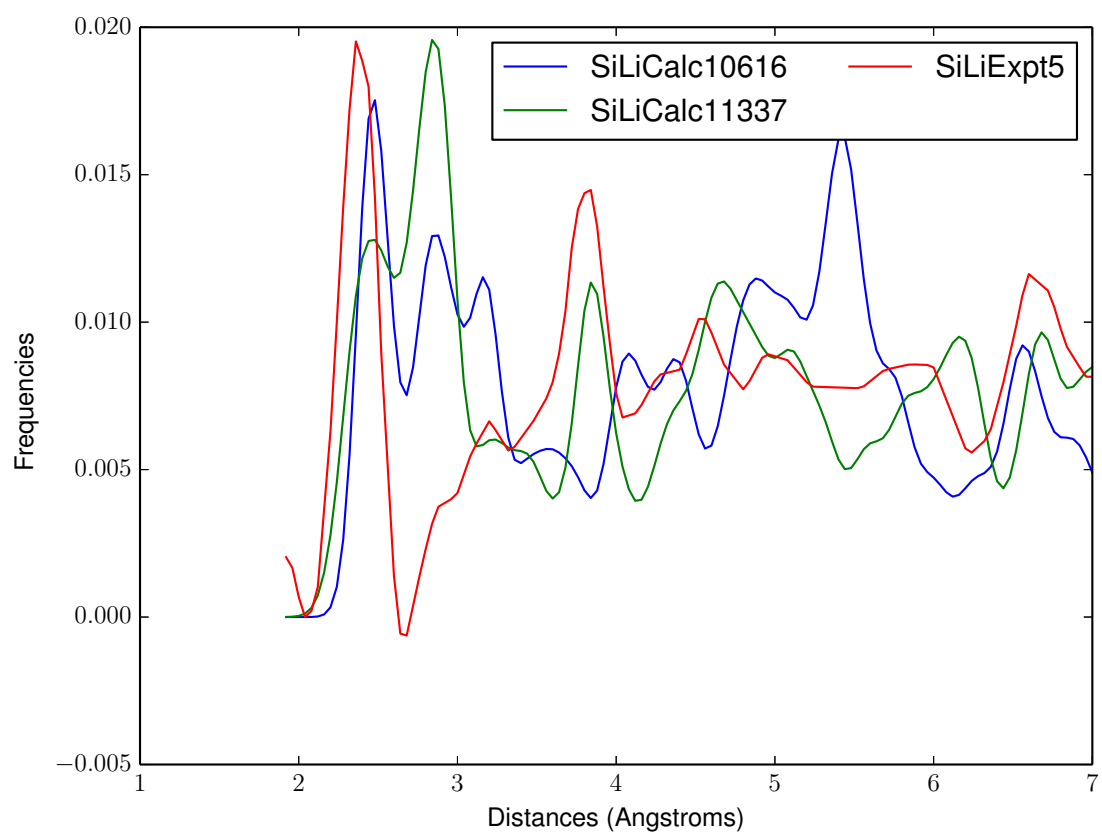


Figure 50: PCA Matches: SiLiExpt5, SiLiCalc10616, SiLiCalc11337

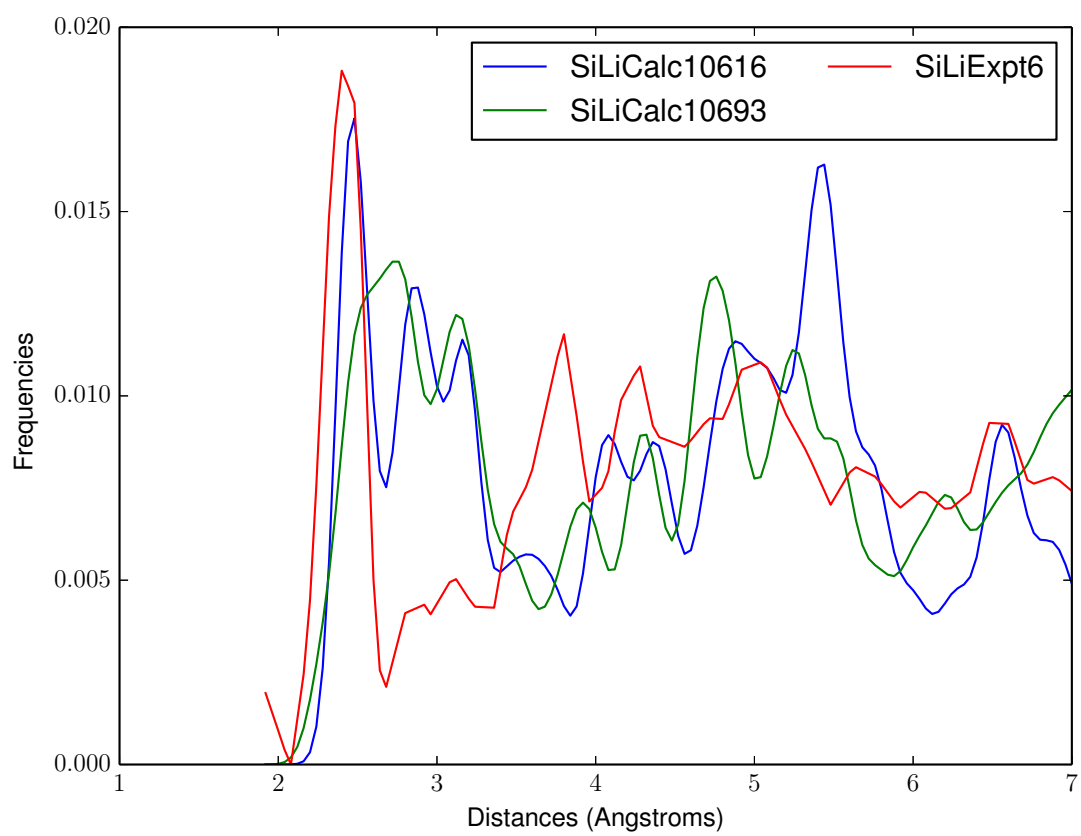


Figure 51: PCA Matches: SiLiExpt6, SiLiCalc10616, SiLiCalc10693

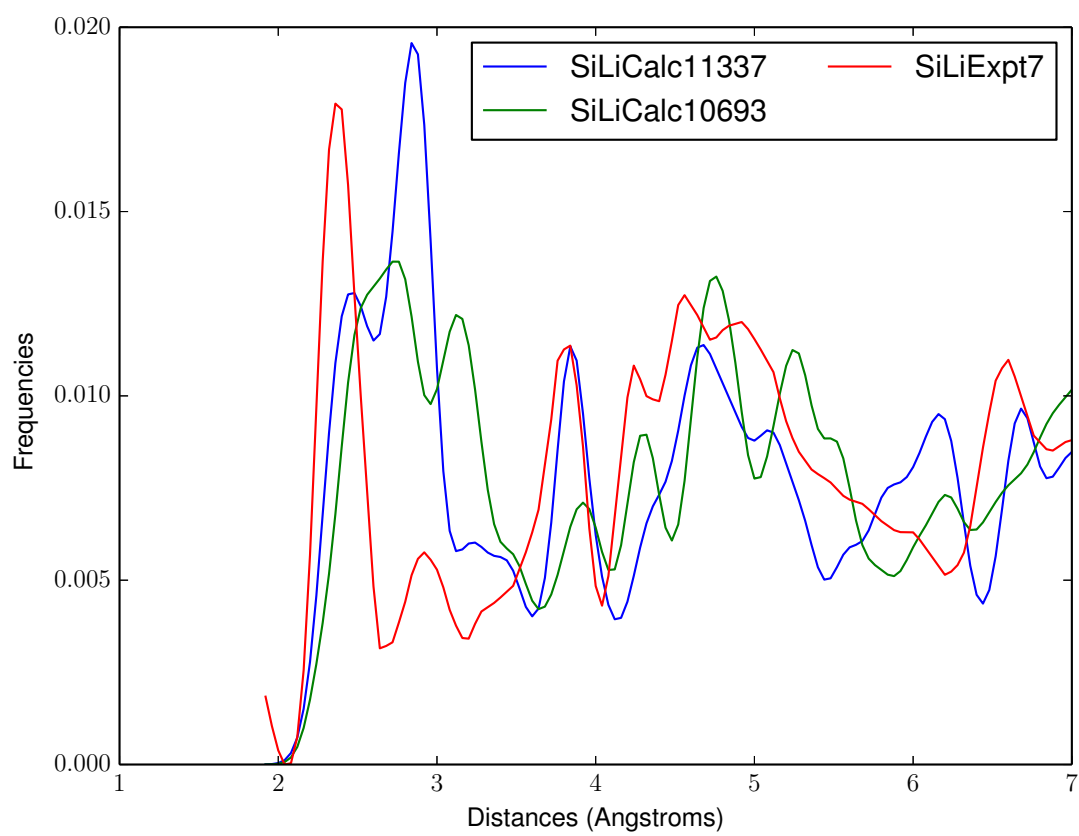


Figure 52: PCA Matches: SiLiExpt7, SiLiCalc10693, SiLiCalc11337

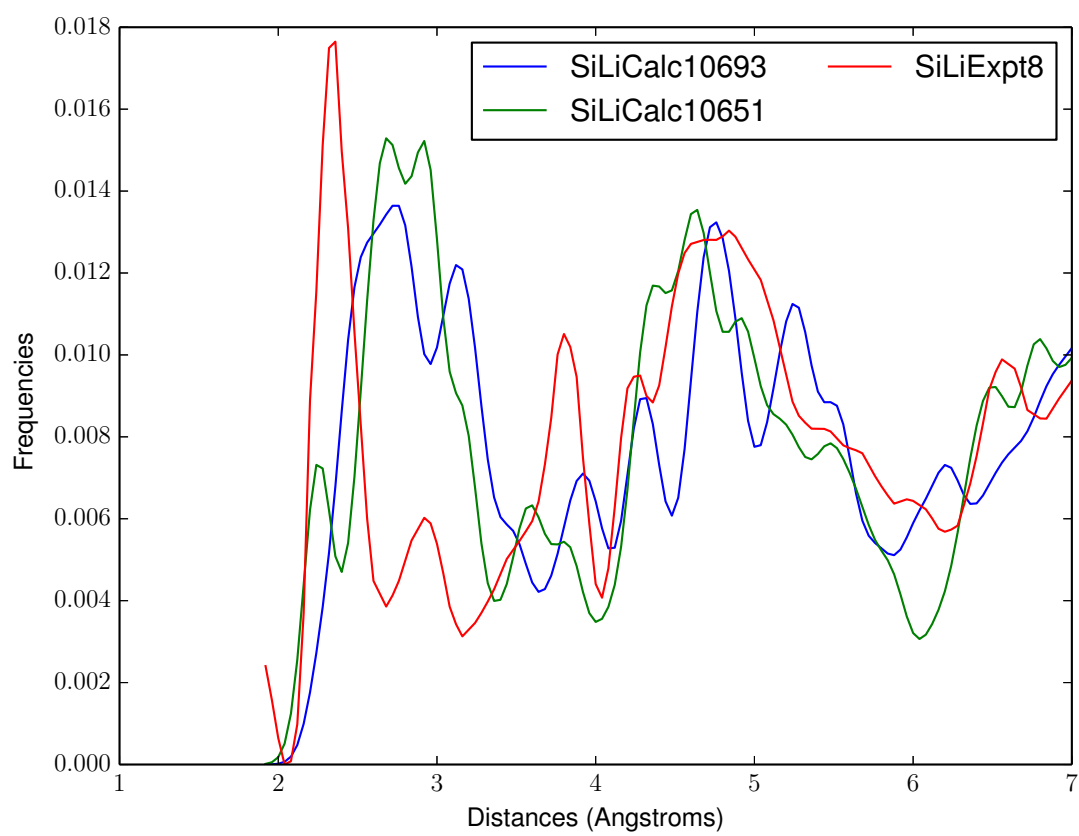


Figure 53: PCA Matches: SiLiExpt8, SiLiCalc10693, SiLiCalc10651

## 4.6 Synthetic Experimental Image Recognition

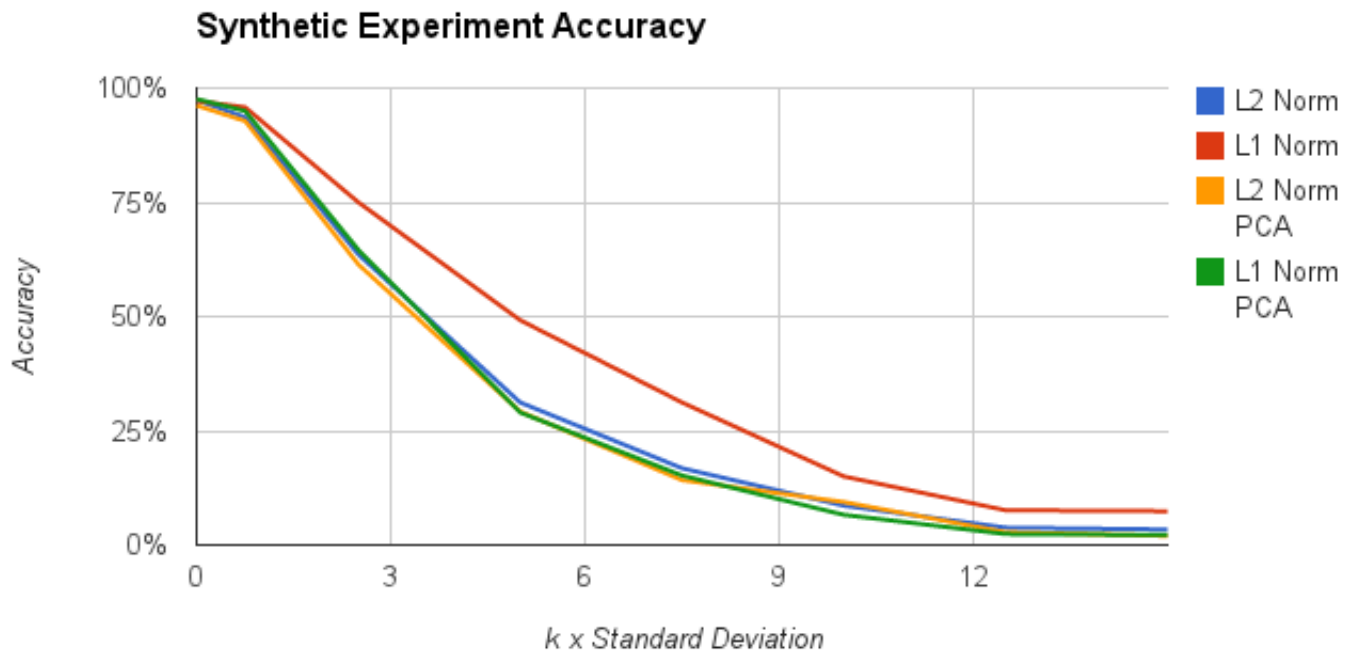


Figure 54: Synthetic Experimental Images Accuracy

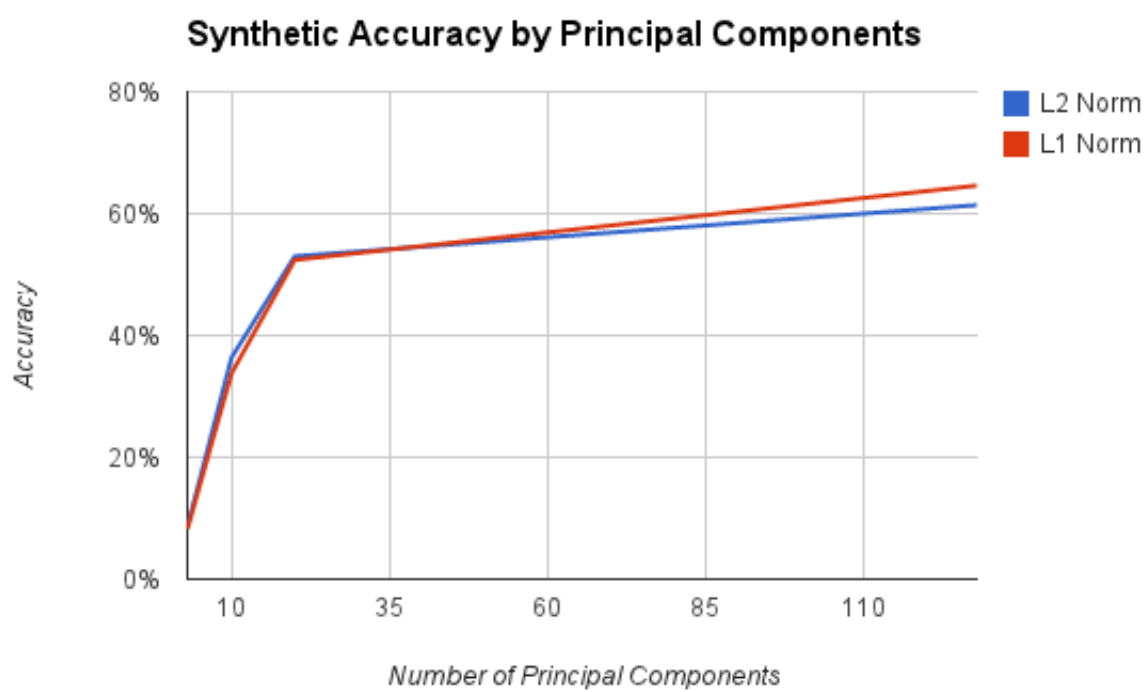


Figure 55: Accuracy vs Number of Principal Components

## 5 Recognition Using Sparse Representations

### 5.1 Experimental Image Recognition

Experiment	Match
ExptGaAs	CalcGaAs
ExptInAs	CalcInAs
SiLiExpt1	SiLiCalc10001
SiLiExpt2	SiLiCalc10001
SiLiExpt3	SiLiCalc10001
SiLiExpt4	SiLiCalc10003
SiLiExpt5	SiLiCalc10003
SiLiExpt6	SiLiCalc10616
SiLiExpt7	SiLiCalc10382
SiLiExpt8	SiLiCalc10382

Table 6: Experimental Image Recognition

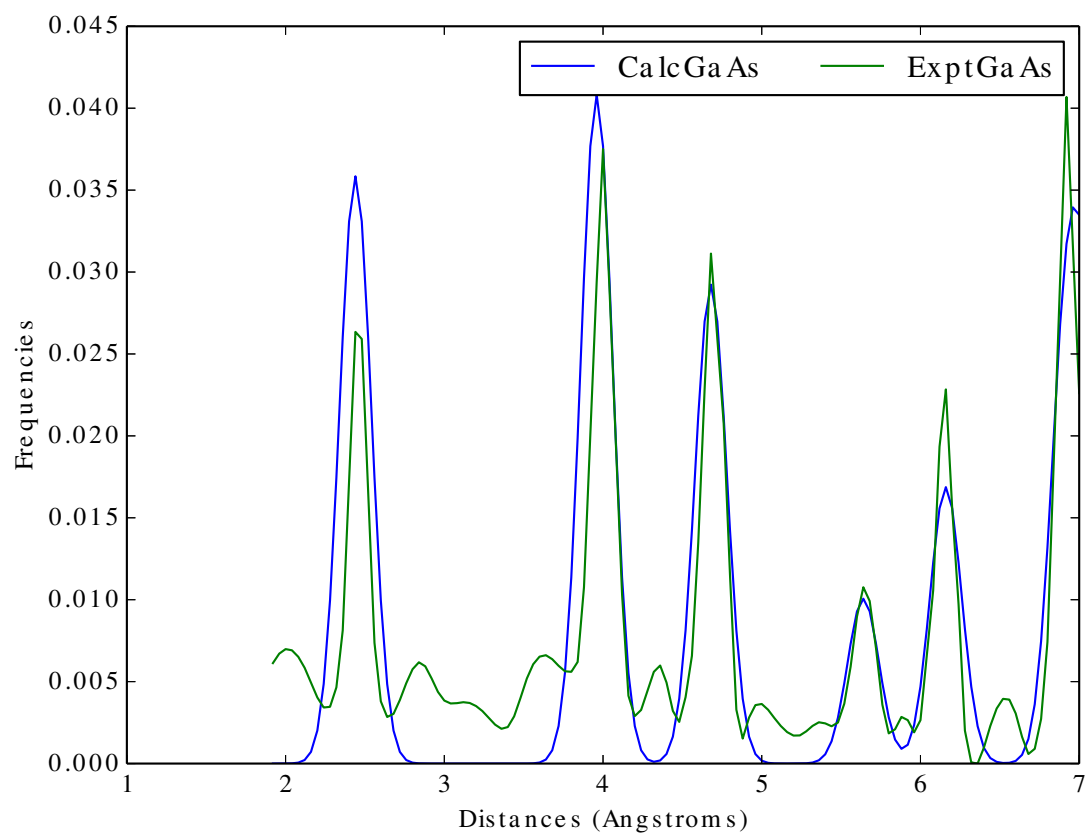


Figure 56: ExptGaAs, CalcGaAs



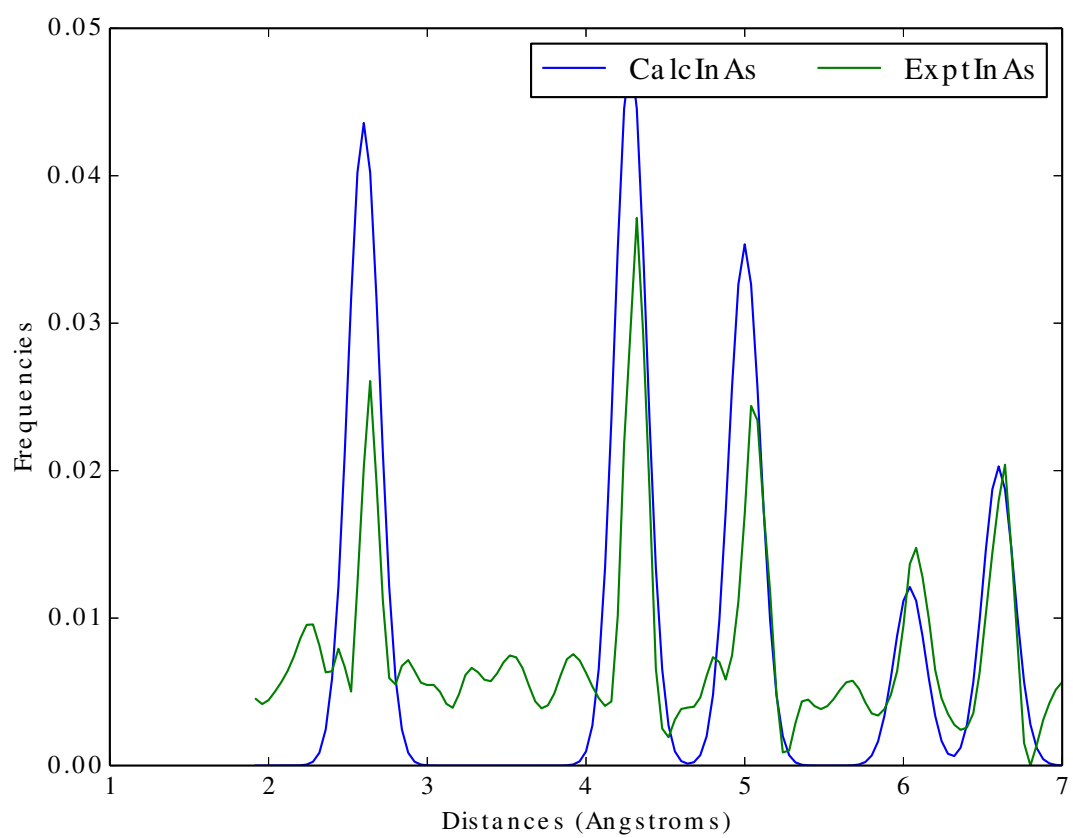


Figure 57: ExptInAs, CalcInAs

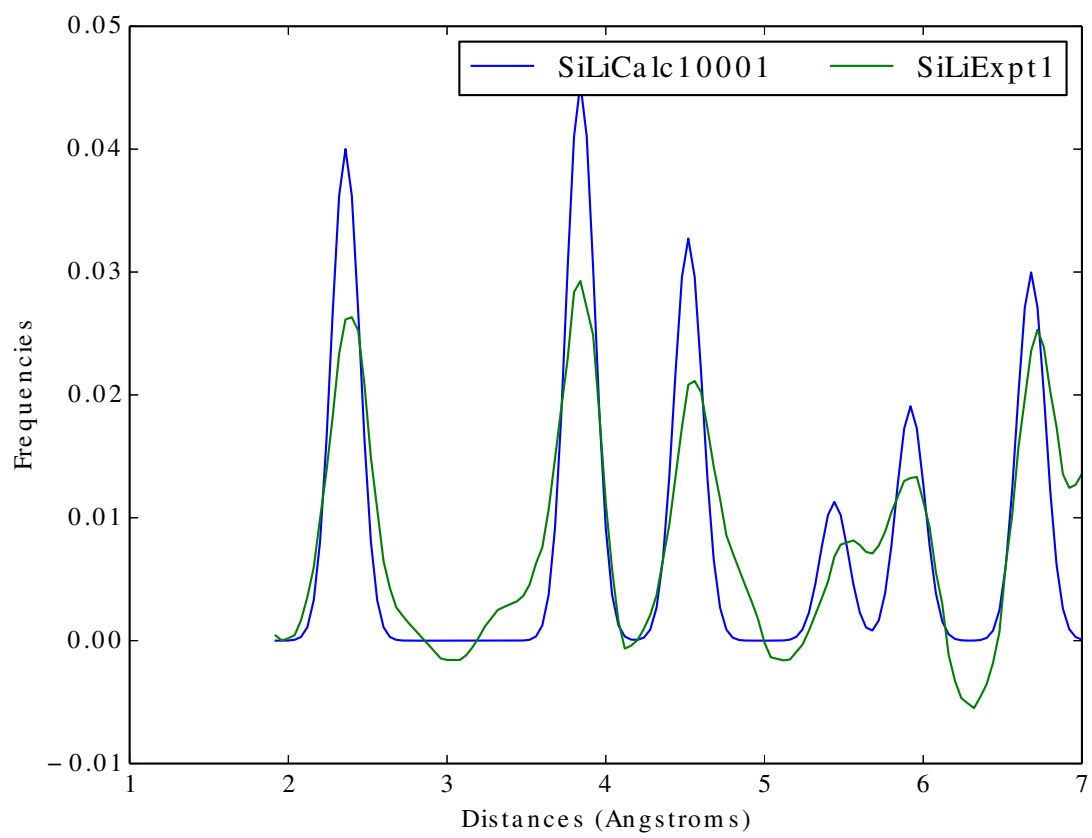


Figure 58: SiLiExpt1, SiLiCalc10001

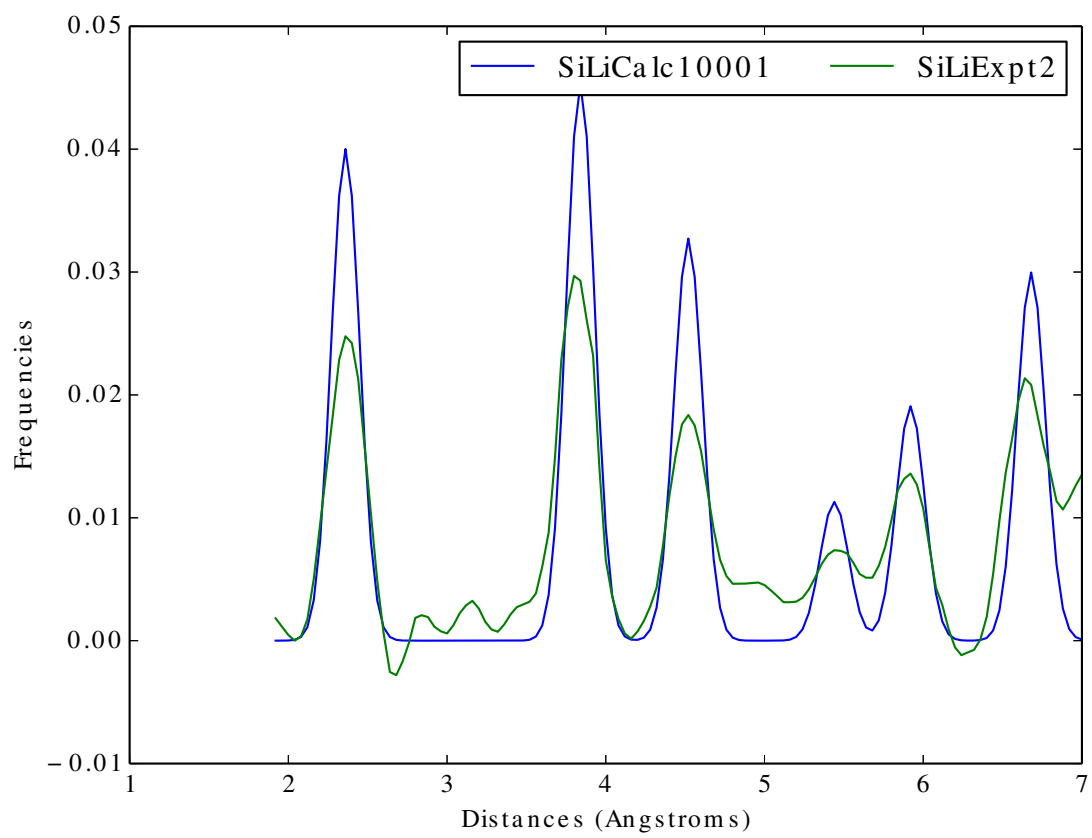


Figure 59: SiLiExpt2, SiLiCalc10001

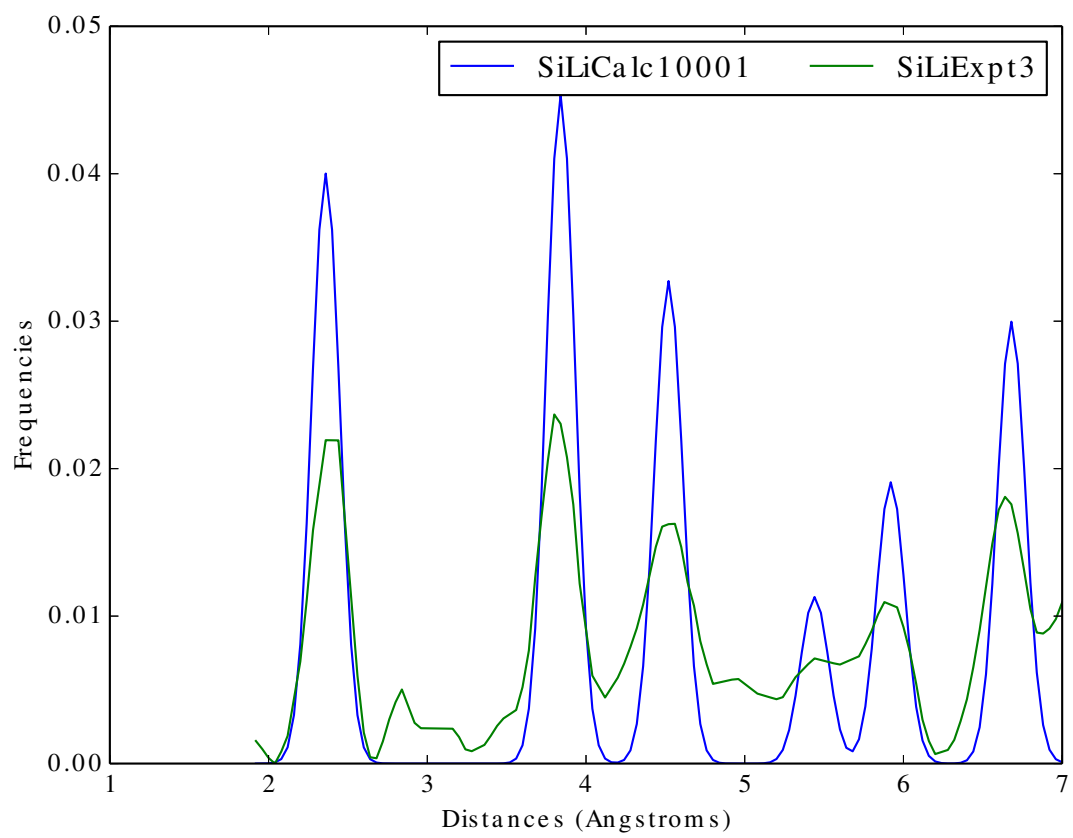


Figure 60: SiLiExpt3, SiLiCalc10001

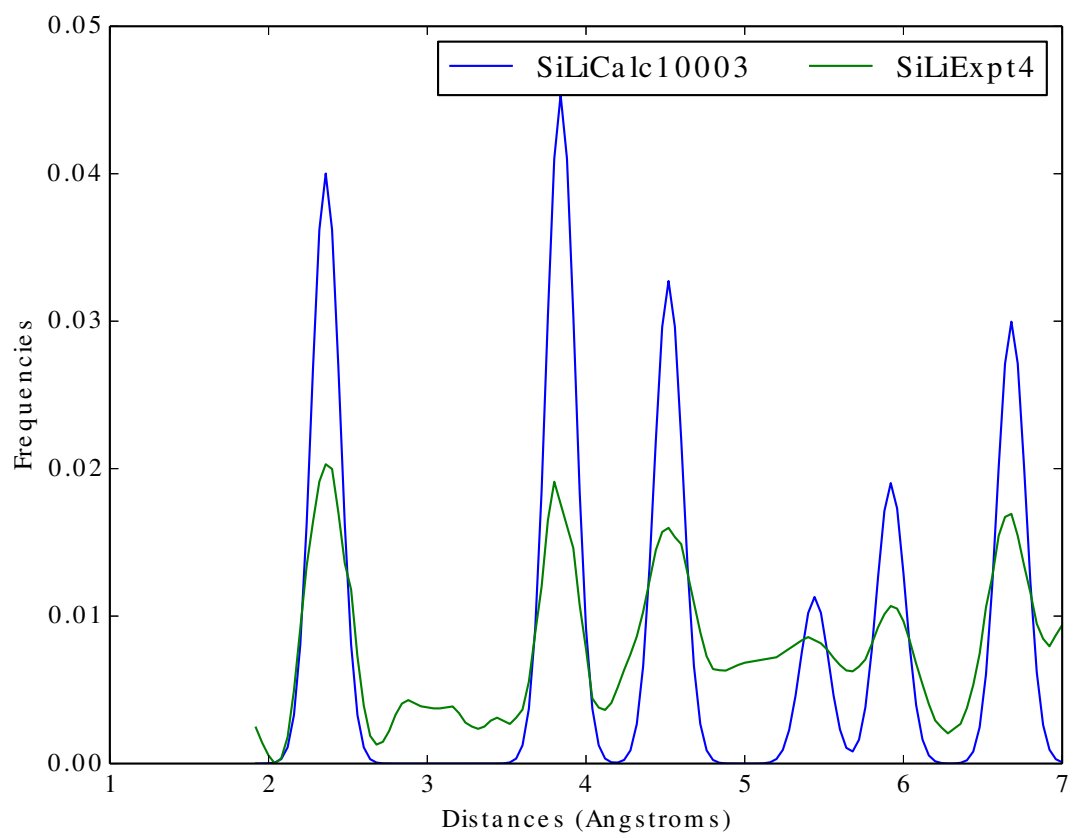


Figure 61: SiLiExpt4, SiLiCalc10003

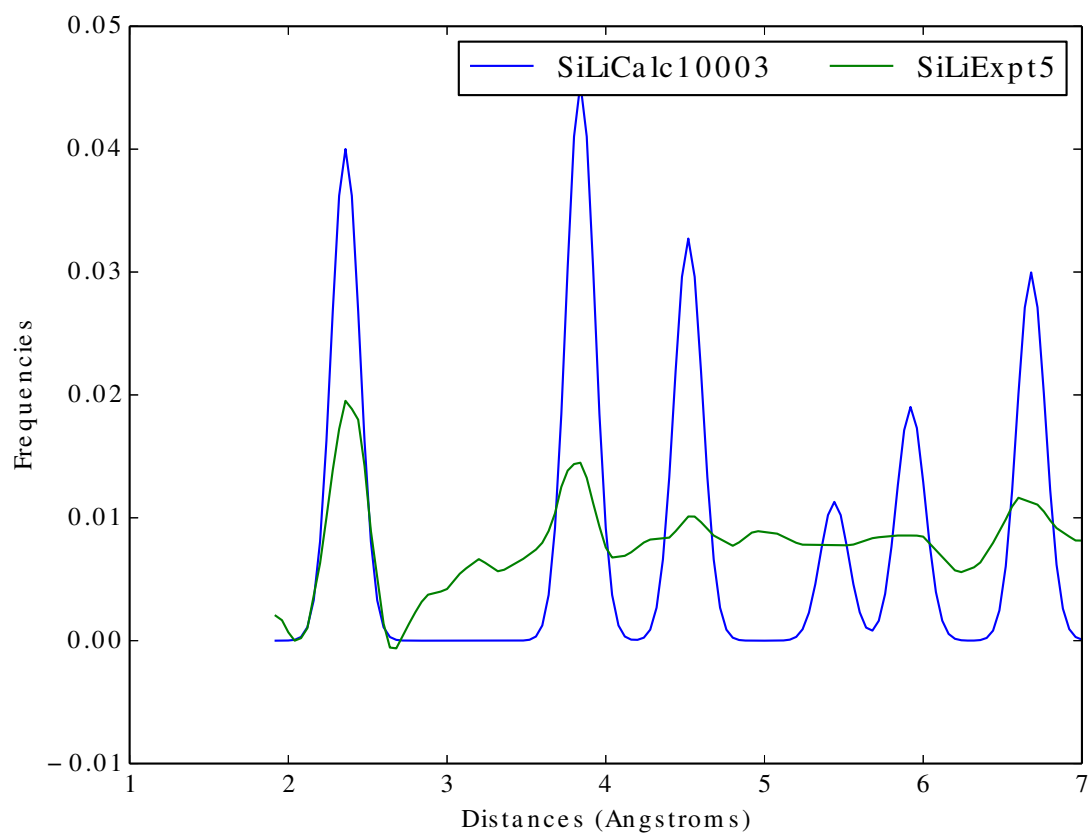


Figure 62: SiLiExpt5, SiLiCalc10003

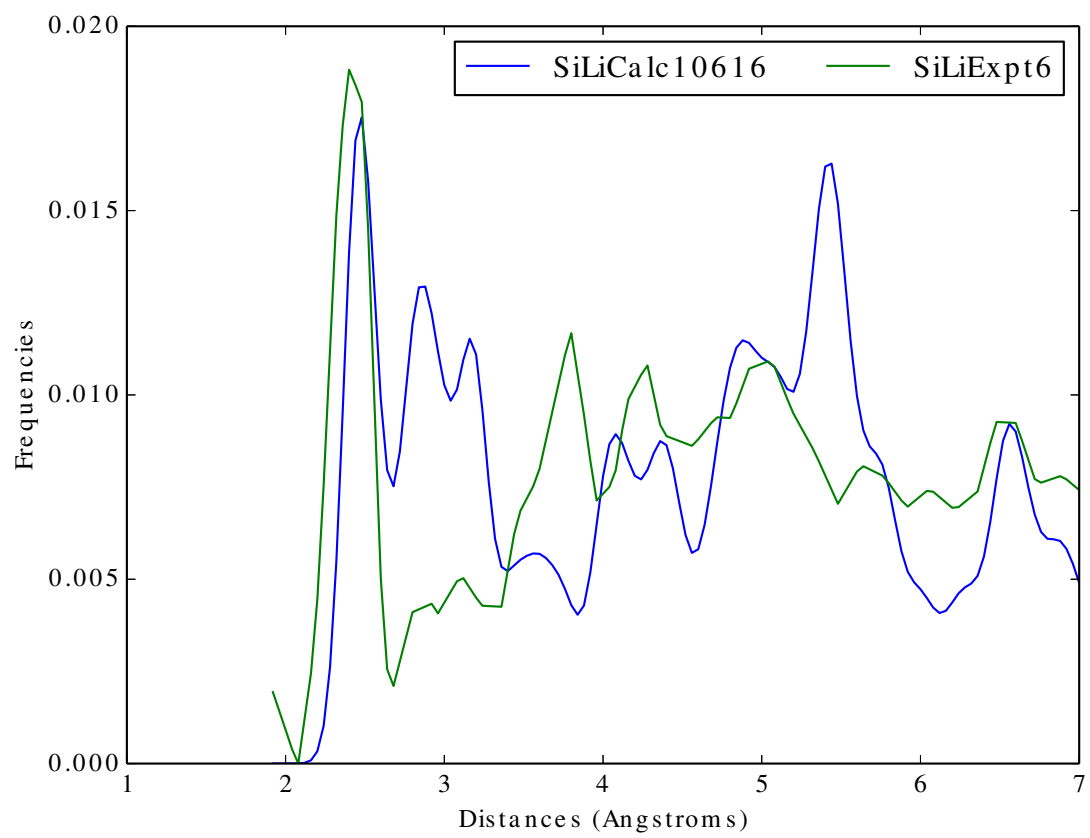


Figure 63: SiLiExpt6, SiLiCalc10616

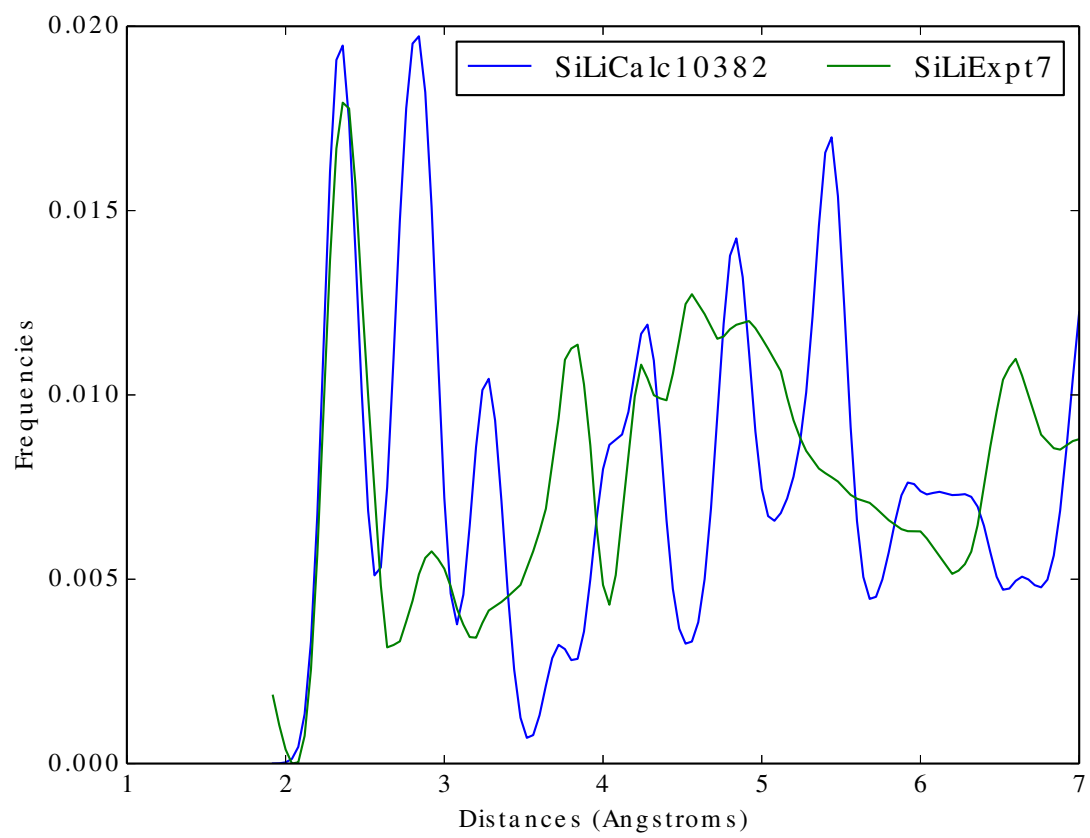


Figure 64: SiLiExpt7, SiLiCalc10382



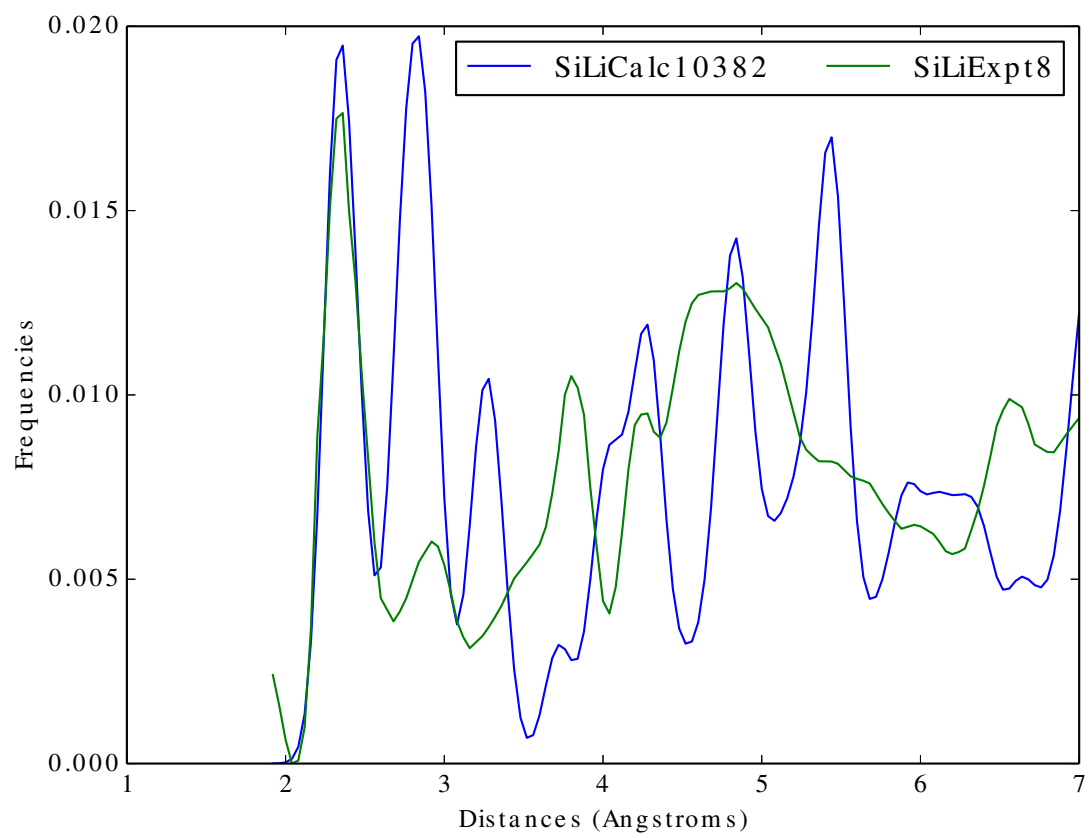


Figure 65: SiLiExpt8, SiLiCalc10382

## 5.2 Synthetic Experimental Image Recognition

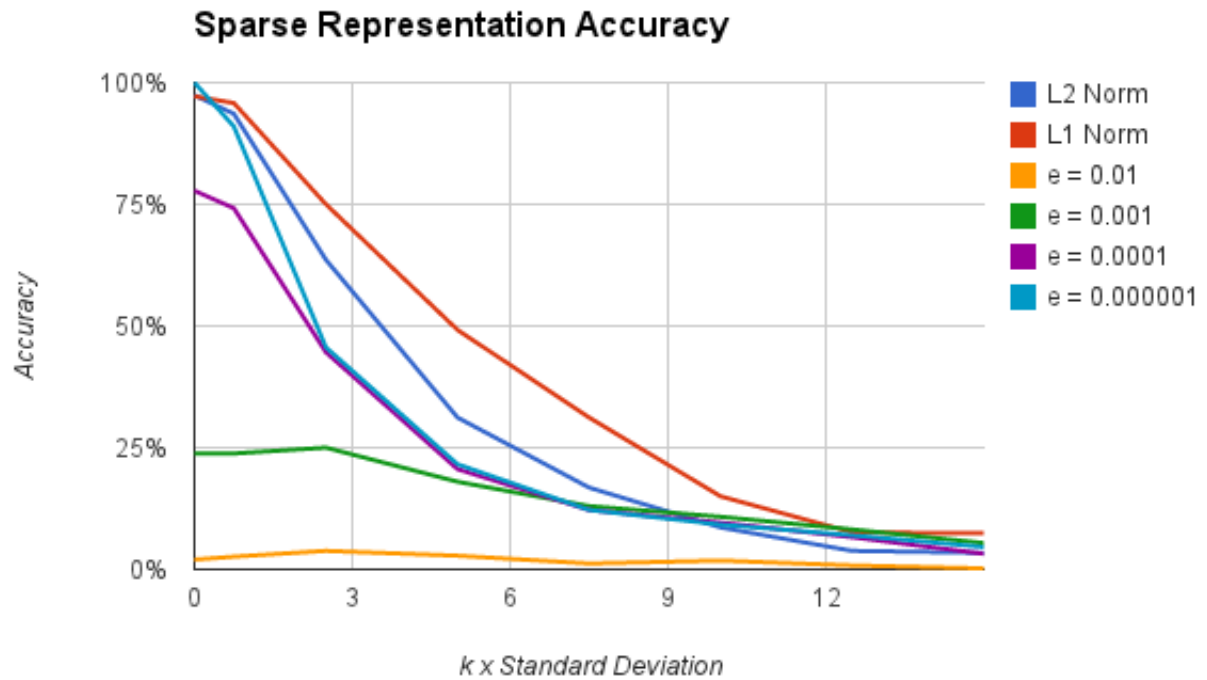


Figure 66: Synthetic Experimental Image Recognition Accuracy

## 6 Sources

[http://en.wikipedia.org/wiki/Atom\\_vibrations](http://en.wikipedia.org/wiki/Atom_vibrations)

[http://en.wikipedia.org/wiki/Radial\\_distribution\\_function](http://en.wikipedia.org/wiki/Radial_distribution_function)

[http://en.wikipedia.org/wiki/Weierstrass\\_transform](http://en.wikipedia.org/wiki/Weierstrass_transform)

[http://matplotlib.org/api/mlab\\_api.html](http://matplotlib.org/api/mlab_api.html)

[http://en.wikipedia.org/wiki/Principal\\_components\\_analysis](http://en.wikipedia.org/wiki/Principal_components_analysis)