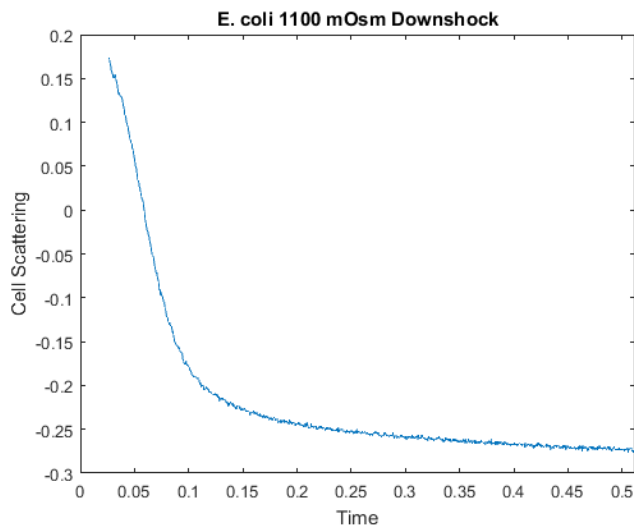


Use of high-pass filters and their effects on noise in light scattering measurements

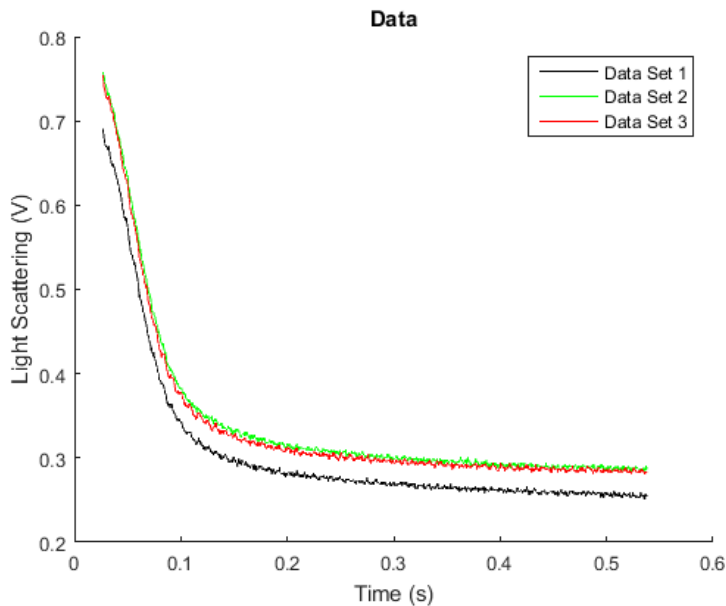
Motivation

When a cell gets placed into a hypo-osmotic media, water rushes into the cell increasing turgor pressure on the inner membrane. Channels in the cell can then open up, expelling both water and other molecules to relieve this pressure and prevent bursting. The opening of the channels and release of material causes a change in refractive index. These physiological changes in the cell can therefore be measured via light scattering. A typical experiment will yield data like this. There exists some variability between data sets that should theoretically be identical. The most apparent source of noise is from the light source, which can in fact be seen as small oscillations in the data. Other mechanical sources of error include changes in positioning or temperature of the light cable or PMT, or the rate of mixing. Other sources include the orientation, sedimentation, and morphological changes of cells. Because these are difficult or impossible to control for, a way to extract these influences from the data would be very helpful in characterizing these channels.



My goal is to use different levels of the Discrete Haar Transform and several high-pass filters to reduce noise in the signal. Level of change was assessed by looking at residuals from filtered data and fitting the data to a basic double exponential fit.

I believe the DHT will be able to reduce some of the light noise found in the data. These oscillations are possibly small enough for the transform to identify and filter out. The slowly-decreasing tail however has clear noise and I believe it is possible to remove. I do not think it will be able to suppress any of the other sources noise because their influence is likely continuous and smooth, unlike the jumpiness of the light noise. As such, I believe DHT will benefit the exponential fit slightly, but it will not be able to consolidate the data at all. I believe they will maintain a similar relationship they do now.



Data:

The data analyzed is the light scattering of a dense solution of *E. coli* being subjected to an 1100 milliosmole downshock. Typically the average of a collection is analyzed, but for this project I will be looking at the 3 sets of data with only the proper control experiments factored in. They are all vectors of length 1024 (2^{10}). All 3 vectors were transformed by the 1st, 5th, and 10th level Discrete Haar Transform, and subsequently filtered with a threshold of 0.001, 0.002, 0.004, 0.1, 0.2, and 1. Effect of the transform was assessed by fitting the data to a basic double exponential fit ($y = a e^{r_1 t} + b e^{r_2 t}$), as well as residuals from unperturbed data. A double exponential fit is used because under the experimental conditions, 2 different membrane channels are expected to open and release water and material from the cell, and while this occurs an exponential decay of light scattering is expected. Therefore it is expected that any pertinent information lost would be manifested in a change in the fit.

3 sets of data have been analyzed. Notice that data set 2 and 3 almost coincide, while data set 1 is initially close, but deviates from the other two after ~0.08 seconds. Because all three are measurements of the same process, the only difference between them (in theory) is sources of noise. These measurements were all taken seconds apart from one another, so it is expected they have faced similar types and amounts of noise.

Because of the extent of analysis and the similarities of the curves, not all graphs will be shown. Graphs were similar across all three curves, and as such are shown as a sort of guide to the trends seen in the analysis.

	Data Set										
Threshold	1				2				3		
	N = 1	N =5	N = 10		N = 1	N =5	N = 10		N = 1	N =5	N = 10
0.001	0.008789	0.011062	0.011062		0.008916	0.011236	0.011236		0.008299	0.010848	0.010848
0.002	0.02057	0.027099	0.027099		0.020867	0.027197	0.027197		0.020902	0.027346	0.027514
0.004	0.035469	0.051880	0.052224		0.036259	0.050868	0.050931		0.037143	0.052313	0.052498
0.1	0.038874	0.204184	0.252002		0.040108	0.209822	0.274574		0.039995	0.201819	0.258614
0.2	0.038874	0.312026	0.405195		0.040108	0.290776	0.381081		0.039995	0.310177	0.412077
1	Above: L^2 Norms – Data filtered with the respective level DHT and threshold was subtracted from filtered data										0.648037

Below: Compression Rates of the above filters

	Data Set										
Threshold	1				2				3		
	N = 1	N =5	N = 10		N = 1	N =5	N = 10		N = 1	N =5	N = 10
0.001	1.295	1.552	1.552		1.275	1.573	1.573		1.250	1.504	1.504
0.002	1.610	2.646	2.646		1.598	2.716	2.716		1.583	2.612	2.632
0.004	1.958	7.262	7.529		1.954	7.161	7.211		1.962	7.585	7.817
0.1	2.000	29.257	85.333		2.000	30.118	102.400		2.000	29.257	85.333
0.2	2.000	32.000	146.286		2.000	32.000	146.286		2.000	32.000	146.286
1	102.400	32.000	204.800			32.000	204.800		85.333	32.000	204.800

Initial Transformation and Filtering

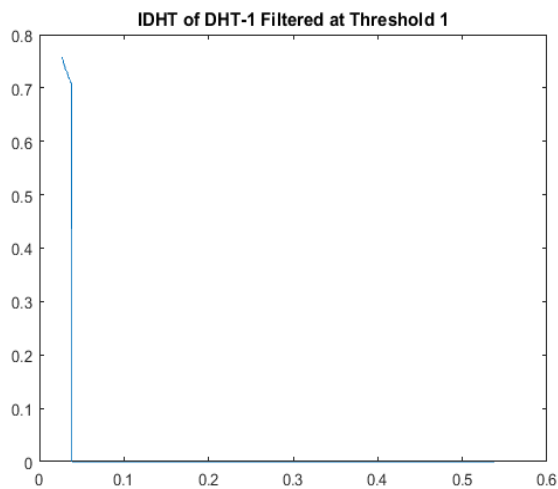
In general, higher N values yield larger residuals. This makes sense, because a small value in a high-level DHT, if not removed by the filter, will add and subtract itself from other entries multiple times over during the inverse process. Interestingly, for both thresholds of 0.001 and 0.002, N = 5 and N = 10 yield the same residual. At these low thresholds, the N = 10 and N = 5 vectors are very similar. These vectors are, strangely, nearly identical, the vector of their difference has all entries whose absolute values are all less than 10^{-12} . (I suspect machine error is responsible for this.) This phenomena is observed for all data at the 0.01 threshold, and two of the data sets at 0.02.

Once again, at a threshold of 0.004, we see an increase in residual, but DHT-5 and DHT-10 are becoming more differentiated. This indicates an upper limit for one of the difference coefficients produced between DHT-5 and DHT-10. Indeed, the compression rate is slightly higher for DHT-10.

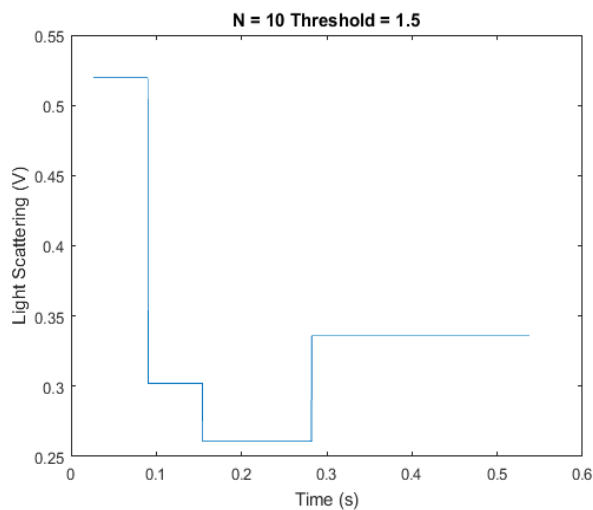
With sufficiently large filters, we can see a sort of saturation of compression. For DHT-1 at threshold 0.1 we get a compression rate of 2 but doubling the threshold does not change this. This is similarly observed for in DHT-5, in which the compression factor reaches 32 and stops. In both of these

cases, all of the difference coefficients have been filtered out, leaving the only the sum coefficients. These are functionally lower bounds for the sum coefficients of the respective transform.

At a threshold of 1, we see a sudden spike in both the residual and compression rate for DHT-1 in all three data sets. At this point, pertinent information has been lost from the sum coefficients and data cannot be properly recovered. Losing the sum coefficients appears to be the only way to prevent a strong recovery after filtering.



The result of over-filtering. We have lost almost all of our data. This only corresponds to a compression rate of 102, higher levels of DHT can undergo more compression and still completely recover.

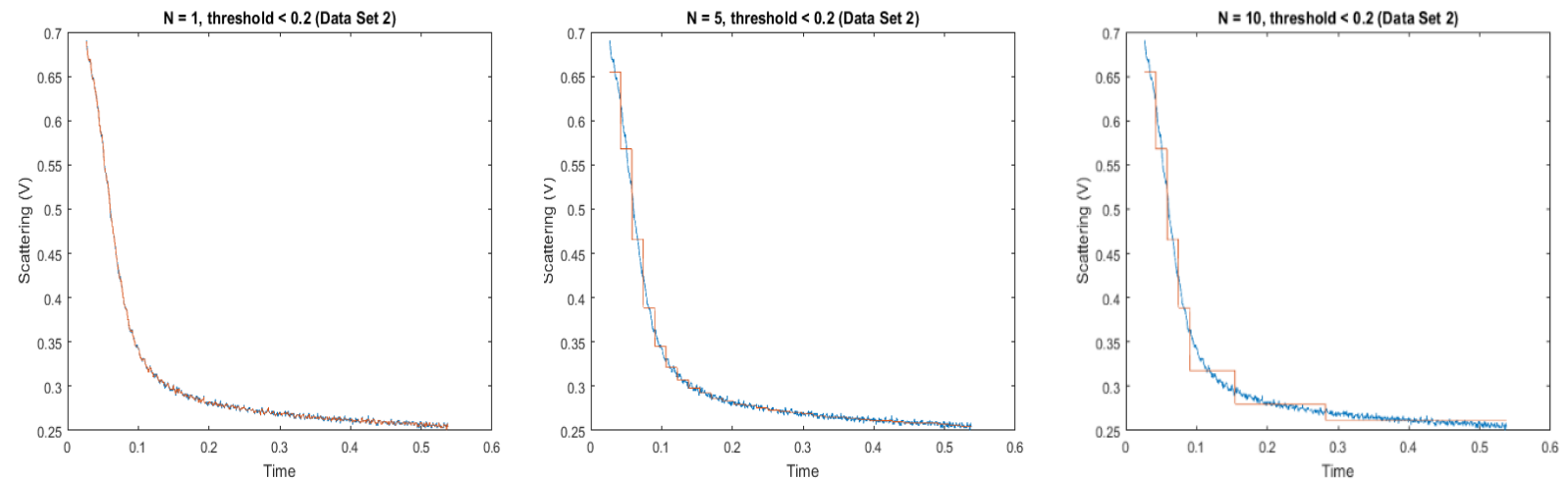


The result of over-filtering DHT-10 (Data set 3). While some of the general shape is conserved, there is a spike where the tail should be and the function is no longer monotonic.

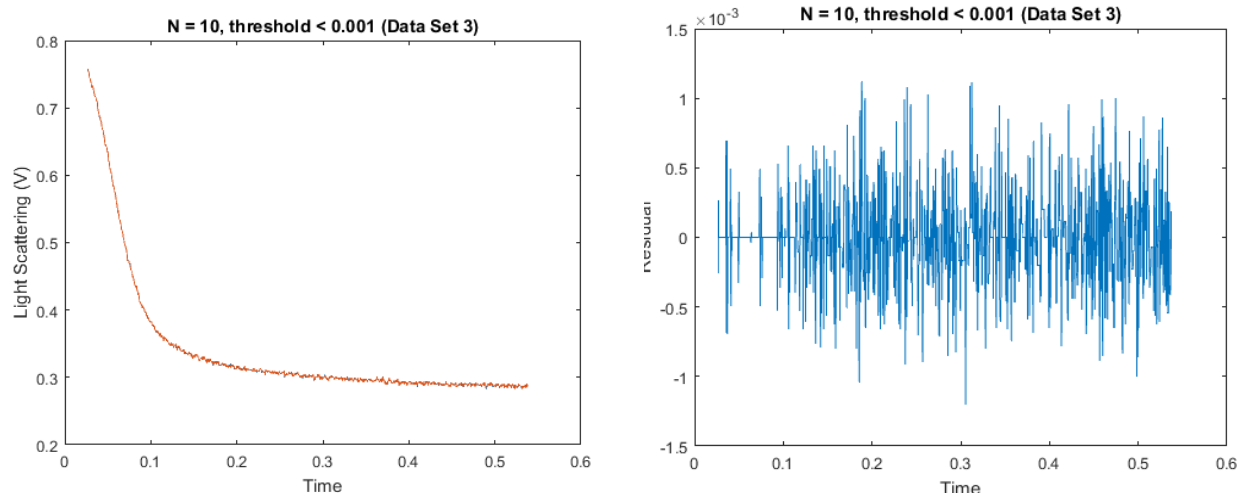
Data Recovery

Amazingly, across all three trials and almost all filters, the data could be recovered. Unfortunately, according to the constants, the filtering did not improve the quality of curve-fitting. The quality was good without filtering though, so expecting improvement would be unreasonable. Filtering

did however make the fits slightly worse when using both a high level of transform and filter. At low thresholds, the level of DHT performed has essentially no effect. This is to be expected, very little compression occurs. What is astonishing is the amount of compression that could be done without compromising the fit. For example, DHT-10 and a threshold of 0.2 could reliably recover the data model with a compression rate of 146. This appears to be the upper bound for acceptable compression rates, as increasing further to 205 did result in a drop in quality across all three data sets. Still, a compression rate of 146 results in 7 coefficients. The fact that these can model our data at all is notable.



Inverse transforms after N-DHT iterations, with a threshold of 0.2. At DHT-1, some of the noise appears lost in the tail, although it is still present. DHT-5 gives the best result at this threshold, the noise in the tail has been reduced greatly. This corresponds to a compression rate of 32. DHT-10 has gone through more compression, with a compression rate of 146. Even with so few coefficients, we keep the same form as the data.



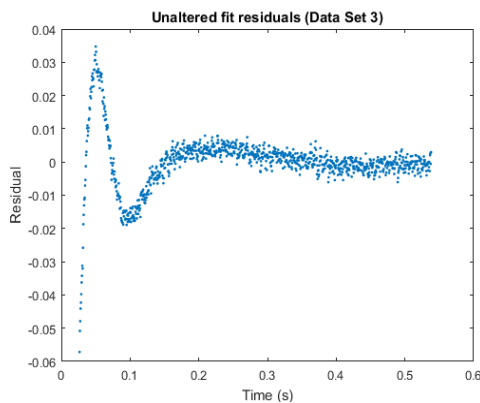
Left: DHT-10 Transform filtered at 0.001, and transformed using IDHT. Both the inverse vector and untransformed vector are plotted, but they are almost perfectly superimposed. Small specks of blue can be seen if zoomed in. These are the places where values change.
Right: Residuals of the left graph.

Applying a high-pass filter generally did not affect the extrapolated rates greatly, but it did increase their uncertainties. (Tables is included at end.) The exception to this was in data set 1, in which DHT-5 at threshold of 0.2 and both DHT-5 and DHT-10 at threshold of 1 yielded rate constants less than the unfiltered data. This is interesting because the same compression rates did not affect the other data sets in this way. This implies a pertinent difference between the data, even though they are measurements of the same process. These differences are the results of errors discussed earlier.

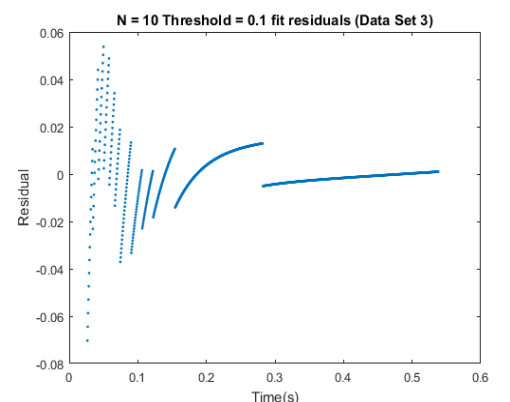
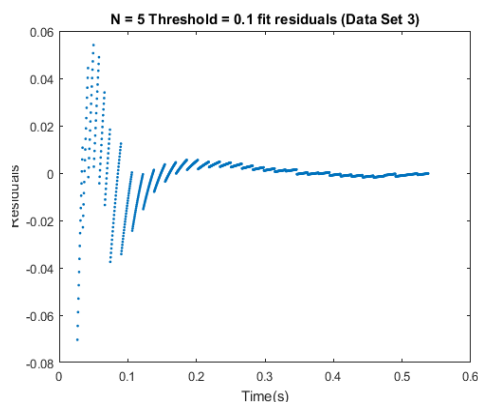
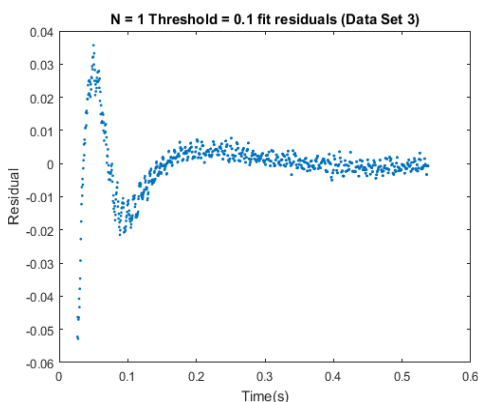
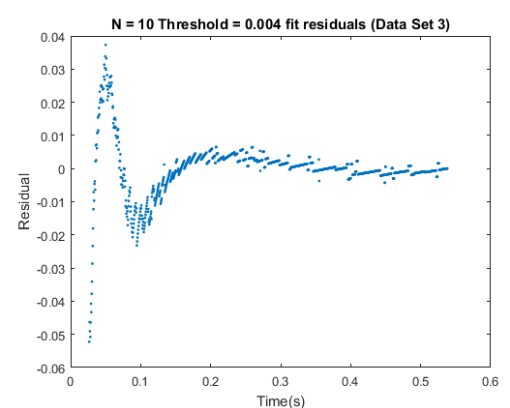
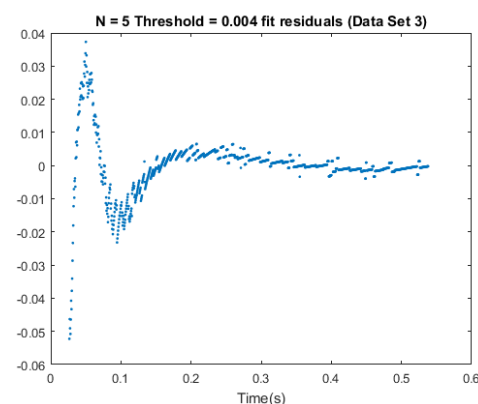
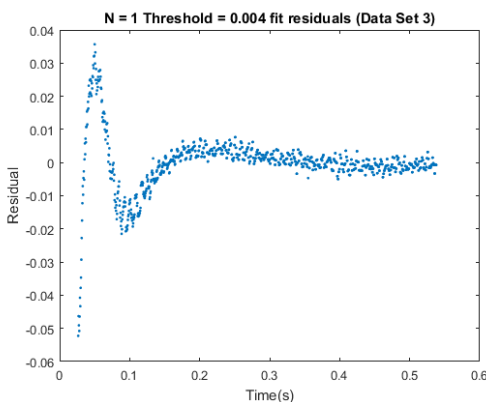
Unsurprisingly, the lower threshold filters had next to no effect on the fits. The 0.001 filter's compression rate is < 1.6 across all 3 trials and all 3 N values. After the inverse transform, the data is for the most part unperturbed, as can be seen in both the superimposed graphs and residuals. The residuals are both small in magnitude and evenly distributed both above and below zero. It is interesting to note that there is a region, corresponding to the steepest region of the curve, in which the residual is 0. This is once again an example of significant difference coefficients not being filtered. This phenomena does however get lost in the higher filters.

While the major constants were for the most part unaltered, this does not give us anything concrete about the filtering procedure. It certainly appears to have no effect, but the higher uncertainties indicates there shouldn't be much confidence placed in this conclusion. We therefore would like to employ a more concrete method to compare fits to the respective data sets. For this, we will once again employ the L^2 norm. We will look at both the residuals and their L^2 norms and compare them to the unaltered data. Interestingly, the double exponential fit only completely failed when the over-filtering mentioned earlier occurred.

The residual graphs themselves do not give any obvious results. The general trend found in the unfiltered residual is present after filtered. At a threshold of 0.004, it does look promising though. In both DHT-5 and DHT-10, this threshold compresses the tail and values around 0. Much like the compression rate and L^2 data shown earlier, there is little difference between the two at this threshold. There is however an increase in residuals at the earlier levels, so additional analysis must be done to determine the extent of benefit. This is also seen at $N = 1$, but to a lesser extent. Increasing the threshold to 0.1 has different effects on all three levels of DHT examined. $N = 1$ appears unaffected. This is logical considering the increase in threshold does not change compression rate much, from 1.58 to 2.



The Residuals from re-fitting the filtered data with a double exponential fit.



There is however a clear change in DHT-5, as the tail loses its noise. This could be counteracted by the notably higher residuals at early times. DHT-10 on the other hand is negatively impacted by this filter, as the tail gets consolidated into larger chunks and approaches zero much more slowly. Considering the transform is 7 Haar functions, this result shouldn't come as much of a shock.

When examining the L^2 norms of the new fits' residuals, it becomes apparent that the 0.004 threshold does benefit the fitting slightly. This effect seems very small, with the norms decreasing by about 0.003 (roughly 1.5%) across the data sets for DHT-10 and DHT-5. DHT-1 is slightly less influenced by this, whose L^2 norm decreased by less than 0.002 across all 3 sets. Most importantly, the L^2 shows that the higher threshold does in fact not improve the fit at all, increasing the norm across all trials. This indicates the squeezing of the tail is offset by the larger residuals at earlier times.

	Data Set 1				Data Set 2				Data Set 3		
Unfiltered	0.2723				0.2666				0.2744		
	N = 1	N = 5	N = 10		N = 1	N = 5	N = 10		N = 1	N = 5	N = 10
0.001	0.2721	0.2721	0.2721		0.2665	0.2664	0.2664		0.2743	0.2742	0.2742
0.002	0.2717	0.2712	0.2712		0.2660	0.2654	0.2654		0.2738	0.2735	0.2735
0.004	0.2707	0.2687	0.2687		0.2651	0.2636	0.2638		0.2728	0.2716	0.2716
0.1	0.2706	0.3241	0.3512		0.2648	0.3565	0.3949		0.2727	0.3287	0.3604
0.2	0.2706	0.4162	0.5122		0.2648	0.3925	0.4809		0.2727	0.4185	0.5316
1		0.4162	0.7313			0.3925	0.6806			0.4185	0.7338

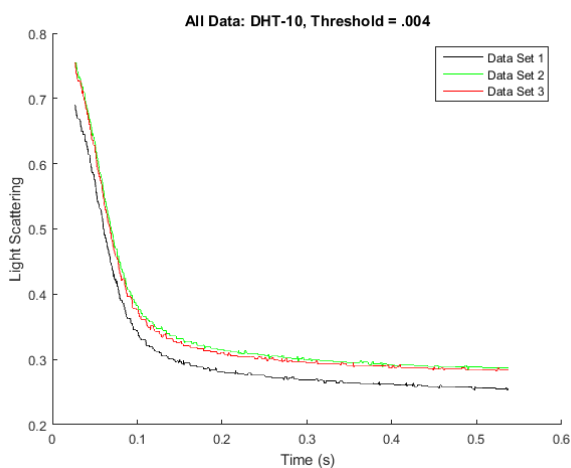
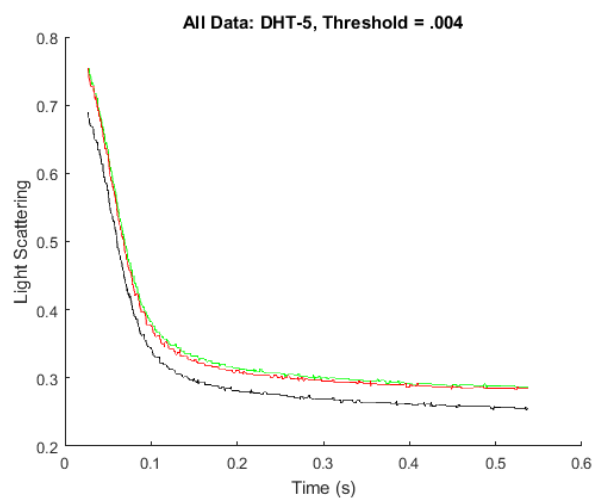
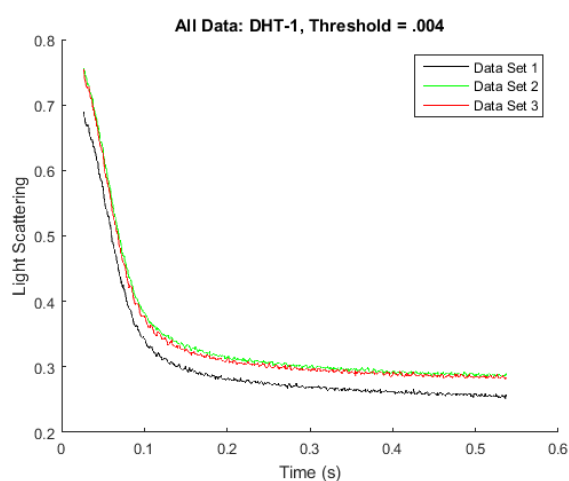
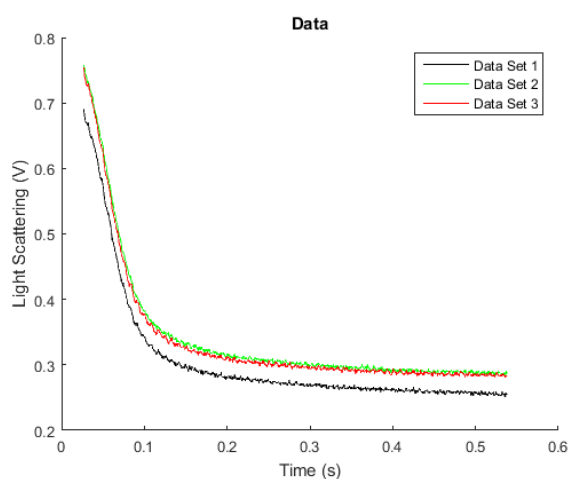
L^2 Norms : Each curve was subjected to the respective level of DHT and filtered using the respective filter. This processed curve was then fit using the same double exponential fit as before, and L^2 Norms of those residuals was taken. 'Unfiltered' gives the residual of the unfiltered data and its double exponential fit.

Because there was a wider range of magnitudes in these residuals, L^1 norm was also calculated. This made the effect of filtering much more apparent than the fits and L^2 norms did. Once again, the 0.004 threshold decreased the norm the most across all 3 sets. It was much more influential, decreasing the 1-norm by about 7% across all trials. This result tells us that all residuals were in general decreased. This was likely masked in the L^2 norms by the higher magnitude residuals that may not have changed much.

	Data Set 1				Data Set 2				Data Set 3		
Unfiltered	4.9847				4.8874				5.1119		
	N = 1	N = 5	N = 10		N = 1	N = 5	N = 10		N = 1	N = 5	N = 10
0.001	4.9765	4.9679	4.9679		4.8782	4.8684	4.8684		5.1052	5.1033	5.1033
0.002	4.9339	4.8904	4.8904		4.8329	4.7830	4.7830		5.0712	5.0140	5.0152
0.004	4.8708	4.6621	4.6747		4.7530	4.5668	4.5652		4.9793	4.7955	4.7981
0.1	4.8646	55.0173	6.7418		4.7450	5.4024	7.2591		4.9783	5.1854	7.1157
0.2	4.8646	6.2316	9.2077		4.7450	5.8859	8.7964		4.9783	6.3760	9.5776
1		6.2316	11.7110			5.8859	11.0241			6.3760	11.9876

L^1 Norms of the residuals discussed in the table above.

Our final goal was to test whether the filtering process could in any way consolidate the curves. Because it gave the best results, this was only tested using a threshold 0.004 filter. As the graphs show, the higher level transforms clearly reduce the amount of noise in the tail. There is, again, very little difference between the DHT-5 and DHT-10 curves. With this noise eliminated, the distance between



data sets 2 and 3 is more apparent, while data set 1 appears to be in a similar position to where it was before.

However, both the L^1 and L^2 norms of the difference vectors indicate that the distance between the data is for the most part unchanged. Interestingly, the L^1 norm between data set 1 and the other sets didn't change. The L^2 norm did however change for all three possible comparisons. (The L^2 norm changing also means that this phenomena isn't an issue with the code, the L^1 norm actually just didn't change. We also see a change in L^1 on the difference vector between sets 2 and 3. I'm perplexed.) The changes in the L^2 norms across all transforms is very small. We saw a roughly 1.5% change in the L^2 norms before and after filtering, and the changes in the L^2 between data sets 2 and 3 is similar if not slightly smaller. The change in L^1 is much smaller than examined earlier, with only a 2% change. We can therefore not conclude whether or not filtering brings these data closer together, as the difference is smaller than just removing the noise. The change between set 1 and the others is however much smaller, at around 0.13%. These small L^2 changes are also not accompanied by a change in L^1 . L^2 seems unshaken in its position.

Vectors	Norm	Unfiltered	N = 1	N = 5	N = 10
1,2	1	31.01692	31.01692	31.01692	31.01692
	2	0.99804	0.99725	0.99673	0.99677
1,3	1	36.40261	36.40261	36.40261	36.40261
	2	1.17412	1.17335	1.17261	1.17263
2,3	1	5.47911	5.42877	5.40050	5.44287
	2	0.20122	0.19693	0.19290	0.19299

Conclusion

The Discrete Haar Transform was able to filter out some of the noise present in the data. Much of the oscillation in the data was removed, resulting in a smoother curve. Whatever other sources contributed to the differences in the data were also not removed, as filtering did not change the exponential fitting of the data significantly, nor did decrease the space between the sets.

$$y = a e^{r_1 t} + b e^{r_2 t}$$

Data Set 1		N = 1			N = 5			N = 10	
		Value	Uncertainty		Value	Uncertainty		Value	Uncertainty
Threshold	a	0.941993	0.012891		0.941992	0.012888		0.941992	0.012888
0.001	r1	-23.7012	0.374713		-23.7011	0.374601		-23.7011	0.374601
	b	0.304995	0.002515		0.304993	0.002514		0.304993	0.002514
	r2	-0.12971	0.02183		-0.1297	0.021824		-0.1297	0.021824
	R-squared	0.992456			0.99246			0.99246	
Threshold	a	0.941969	0.01287		0.941962	0.012843		0.941962	0.012843
0.002	r1	-23.7007	0.374105		-23.6999	0.373328		-23.6999	0.373328
	b	0.304993	0.002511		0.304985	0.002506		0.304985	0.002506
	r2	-0.1297	0.021795		-0.12962	0.021751		-0.12962	0.021751
	R-squared	0.99248			0.99251			0.99251	
Threshold	a	0.941925	0.012821		0.941914	0.012725		0.941858	0.012724
0.004	r1	-23.7	0.372689		-23.6991	0.369894		-23.6951	0.369889
	b	0.304993	0.002502		0.30498	0.002483		0.304939	0.002484
	r2	-0.1297	0.021713		-0.12957	0.021552		-0.12919	0.021559
	R-squared	0.992536			0.992646			0.992643	
Threshold	a	0.941869	0.012814		0.931645	0.015181		0.941246	0.012792
0.1	r1	-23.6986	0.372523		-23.54	0.44642		-23.6672	0.372058
	b	0.304989	0.002501		0.305017	0.003014		0.304672	0.002502
	r2	-0.12967	0.021705		-0.12997	0.026121		-0.1267	0.021733
	R-squared	0.992542			0.989255			0.992532	
Threshold	a	0.941869	0.012814		0.900576	0.01851		0.93687	0.015091
0.2	r1	-23.6986	0.372523		-22.802	0.563895		-23.5499	0.440744
	b	0.304989	0.002501		0.302885	0.003962		0.303577	0.002979
	r2	-0.12967	0.021705		-0.11374	0.034342		-0.11555	0.025926
	R-squared	0.992542			0.982176			0.989406	
Unfiltered									
A	0.941998	0.012897	Threshold		0.900576	0.01851		0.801444	0.02714
r1	-23.7013	0.37486	1		-22.802	0.563895		-20.3403	0.935482
B	0.304995	0.002516			0.302885	0.003962		0.291952	0.007528
r2	-0.12971	0.021838			-0.11374	0.034342		-0.01603	0.065836
R-squared	0.99245				0.982176			0.943079	

Anthony Schams
MATH416
December 23, 2016

Data Set 2		N = 1			N = 5			N = 10	
		Value	Uncertainty		Value	Uncertainty		Value	Uncertainty
Threshold	a	0.872456	0.01278		0.87246	0.012777		0.87246	0.012777
0.001	r1	-23.8786	0.401998		-23.879	0.401892		-23.879	0.401892
	b	0.279895	0.002473		0.279897	0.002472		0.279897	0.002472
	r2	-0.16369	0.023464		-0.16372	0.023458		-0.16372	0.023458
	R-squared	0.99161			0.991615			0.991615	
Threshold	a	0.872435	0.012756		0.872438	0.012729		0.872438	0.012729
0.002	r1	-23.8782	0.401265		-23.8779	0.400402		-23.8779	0.400402
	b	0.279894	0.002469		0.279889	0.002463		0.279889	0.002463
	r2	-0.16368	0.023422		-0.16363	0.023373		-0.16363	0.023373
	R-squared	0.99164			0.991675			0.991675	
Threshold	a	0.87233	0.012709		0.872263	0.012637		0.872251	0.012637
0.004	r1	-23.8756	0.399827		-23.8729	0.397592		-23.8721	0.397594
	b	0.279887	0.00246		0.279866	0.002447		0.279858	0.002447
	r2	-0.16363	0.023341		-0.16341	0.023215		-0.16333	0.023217
	R-squared	0.991698			0.991788			0.991787	
Threshold	a	0.872304	0.012695		0.849283	0.016463		0.871806	0.012709
0.1	r1	-23.8751	0.399396		-23.3358	0.532535		-23.8491	0.400009
	b	0.279887	0.002458		0.278704	0.003366		0.279637	0.002465
	r2	-0.16363	0.023317		-0.15371	0.031915		-0.16104	0.0234
	R-squared	0.991716			0.984904			0.991664	
Threshold	a	0.872304	0.012695		0.834787	0.017713		0.865663	0.014987
0.2	r1	-23.8751	0.399396		-23.0083	0.58343		-23.648	0.474667
	b	0.279887	0.002458		0.278118	0.003748		0.278023	0.002954
	r2	-0.16363	0.023317		-0.14903	0.035505		-0.14372	0.02813
	R-squared	0.991716			0.981613			0.988037	
Unfiltered									
a	0.941998	0.012897	Threshold	a	0.834787	0.017713		0.743359	0.025477
r1	-23.7013	0.37486	1	r1	-23.0083	0.58343		-20.4543	0.947871
b	0.304995	0.002516		b	0.278118	0.003748		0.26671	0.007016
r2	-0.12971	0.021838		r2	-0.14903	0.035505		-0.03661	0.067299
R-squared	0.99245			R-squared	0.981613			0.942834	

Data Set 3		N = 1			N = 5			N = 10	
		Value	Uncertainty		Value	Uncertainty		Value	Uncertainty
Threshold	a	0.930093	0.012516		0.930095	0.012513		0.930095	0.012513
0.001	r1	-23.1667	0.369695		-23.1667	0.369594		-23.1667	0.369594
	b	0.310865	0.002601		0.310864	0.0026		0.310864	0.0026
	r2	-0.14669	0.022069		-0.14669	0.022063		-0.14669	0.022063
	R-squared	0.99258	0.992559		0.992584	0.992563		0.992584	0.992563
Threshold	a	0.930079	0.012492		0.93007	0.012478		0.930062	0.012478
0.002	r1	-23.1664	0.368994		-23.166	0.368578		-23.1653	0.368574
	b	0.310864	0.002596		0.310859	0.002593		0.310852	0.002593
	r2	-0.14669	0.022027		-0.14664	0.022003		-0.14657	0.022004
	R-squared	0.992608	0.992586		0.992624	0.992603		0.992624	0.992602
Threshold	a	0.930009	0.012445		0.929933	0.012389		0.929911	0.012389
0.004	r1	-23.165	0.36764		-23.1622	0.366017		-23.1605	0.366008
	b	0.310861	0.002587		0.310834	0.002576		0.310817	0.002576
	r2	-0.14666	0.021948		-0.14641	0.021856		-0.14626	0.021858
	R-squared	0.992661	0.99264		0.992723	0.992702		0.992722	0.992701
Threshold	a	0.929968	0.012441		0.919588	0.014814		0.929366	0.012501
0.1	r1	-23.1641	0.36752		-22.9919	0.442864		-23.1345	0.369497
	b	0.310858	0.002586		0.310774	0.003137		0.310552	0.002605
	r2	-0.14664	0.021942		-0.14606	0.026589		-0.14383	0.022113
	R-squared	0.992666	0.992644		0.989303	0.989272		0.99256	0.992538
Threshold	a	0.929968	0.012441		0.890057	0.017928		0.922086	0.014331
0.2	r1	-23.1641	0.36752		-22.279	0.55486		-22.889	0.426732
	b	0.310858	0.002586		0.308563	0.004092		0.308647	0.003051
	r2	-0.14664	0.021942		-0.12959	0.034678		-0.12657	0.025985
	R-squared	0.992666	0.992644		0.982557	0.982506		0.989854	0.989824
Unfiltered					0.890057	0.017928		0.79832	0.026394
a	0.930097	0.012521	Threshold	a	-22.279	0.55486		-19.9337	0.918015
r1	-23.1668	0.369832	1	r1	0.308563	0.004092		0.296865	0.007742
b	0.310865	0.002602		b	-0.12959	0.034678		-0.02643	0.066319
r2	-0.14669	0.022077		r2	0.982557	0.982506		0.944598	0.944435
R-squared	0.992575	0.992553		R-squared	0.930095	0.012513		0.930095	0.012513