Assessing the usefulness of an atomic force microscope in the study of surface roughness

Alfie Renn
L3 Advanced Laboratory, Optics Lab Team

Submitted: April 15, 2021, Date of Experiment: October 27th - December 3rd 2020

The Atomic Force Microscope (AFM) is a relatively novel tool in the field of surface science. This report investigates the usefulness of the AFM in refractive optics by imaging a glass plate sanded to two different standards by two sizes of microparticle. Post-processing software is used to retrieve roughness parameters for each surface which are analysed to find a metric to differentiate the two surfaces. Overall, the AFM ends up less versatile than a tool like the electron microscope, but far more affordable and easier to use.

1. INTRODUCTION

The Atomic Force Microscope (AFM) is a tool able to image surfaces from the micro- to the nano-scale, far surpassing the resolution of traditional optical microscopes, and is a simpler alternative to the more expensive electron microscopes [1]. Since its invention [2], it has been incredibly useful to fields such as surface science and nanotechnology, and provides an affordable method of imaging surfaces on the nano-scale. Surface science has many applications, for example it provides insights into properties of catalysts, hydrophobicity, as well as many quantum phenomena [3].

The setup for atomic force microscopy is shown in figure 1. It is conducted by shining a laser onto a cantilever arm which is free to flex a little. The laser is reflected into a photodiode which can detect the deflection of the cantilever arm via the position of the laser dot on the photodiode. The arm ends in a very sharp tip which contacts a sample of material, which is translated both parallel and perpendicular to the arm. In this translation, the shape of the surface will cause the arm to deflect upwards or downwards in peaks or valleys, respectively, and this deflection can be measured by the photodiode and the shape of the sample can be inferred.

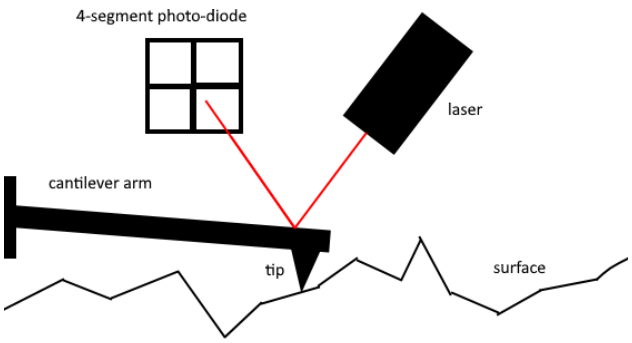


FIG. 1: A simplified diagram of an Atomic Force Microscope (AFM). Shown is the laser reflecting on the cantilever arm and into the photodiode, while the arm is resting on an example surface.

Another field of physics adjacent to surface science is modern optics, where it is often important to know the properties of surfaces to determine what will happen to light incident on material boundaries [4]. In this report, we conduct an experiment to measure the surface roughness of a glass lens when sanded with different grades of microparticles, using an AFM for measurements. We will analyse various roughness parameters [5, 6] and surface features of the sanded glass to determine if this technique could be used to differentiate two glass plates having been sanded with two different grades of microparticles. If this is possible, then there will be a measurable difference between the roughness parameters of our two surfaces.

2. METHODS

This section will first discuss how the samples scanned with the AFM were prepared, followed by an explanation of how images were taken using the AFM, building on information given above. Finally, the mathematical descriptions of several roughness parameters will be explored, which will be used to analyse our surfaces.

The initial state of the surface was a smooth glass plate, often used as a lens in modern optics. Then, two different sanding pastes were prepared by combining silicon carbide polishing powders and silicone grease. These two pastes were used to sand the glass surface continually for four minutes, after which the paste was removed with a degreaser. The two different microparticle sizes used were G120 $\approx$ 115 µm and F400 $\approx$ 17 µm. The sanding process is modelled as many small, hard spheres moving across the surface with high friction, which then carve grooves and channels into the glass, but are not themselves damaged as they are hard. This will change the lay of the surface of the glass. Next, the samples were imaged using the AFM as described below.

A. The Atomic Force Microscope

The AFM used to gather the data for this report was the EDU-AFM1 from optical equipment company Thorlabs [7]. A schematic of this AFM is shown in figure 2, and as seen, this particular AFM operates by moving a stage with a sample secured on it rather than moving the cantilever and laser setup. This means that calibration is a lot simpler and makes changing the sample easier. For this AFM, the size of the tip is 7-10 nm, meaning that it can resolve features down to a

resolution of 14-20 nm.

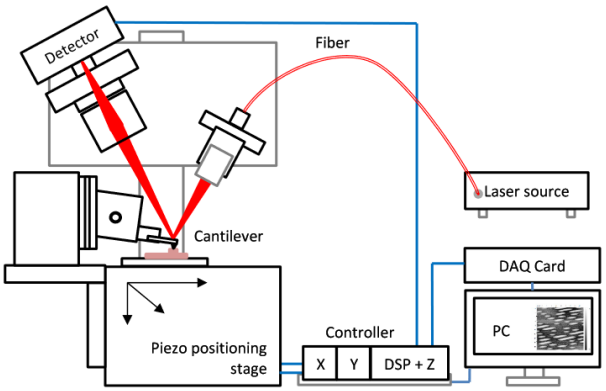


FIG. 2: A diagram of the AFM used in this report: the ThorLabs EDU-AFM1. The sample is moved via the controller connected to the PC and the same components are seen as in figure 1. Image taken from [7].

The sample stage can move in all three cardinal directions using piezo nanopositioners, or by using hand-turned adjusters, allowing manual movement down to 1 µm precision.

To take an image, the AFM must be initialised. First, the laser is be centred on the photodiode. Then, the tip is positioned above the area of the sample to be scanned, and then moved vertically until the tip meets the sample. At this point, there is a feedback loop between the photodiode and z-piezo controller (as shown as the blue line in figure 2) which alters the z-piezo voltage (thus changing the height of the sample) in order to keep the laser dot at the centre of the photodiode. This is known as 'constant force' mode, as there is always a constant force on the cantilever arm. The range of measurable heights is then clearly the range of the z-piezo - about 10 µm.

Next, to take the actual image, the x- and y-piezos move the sample stage in the x- and y-directions, scanning across the sample. The lateral range of this AFM is 20 µm in both the x- and y-directions. The software moves the stage in discrete jumps, scanning a pixel, then moving across the horizontal direction in lines, and then builds an image by scanning line-by-line in the vertical direction (where 'horizontal' is perpendicular to the cantilever arm). At each point, the voltage of the z-piezo is recorded, which can later be turned into a height via calibration.

For calibration, a sample HS100-MG microstructure [8] is used, the shape of which is shown in figure 3. This is imaged using the AFM and then the pitch (in x- and y-directions) and depth (z-direction) of the features are measured using the postprocessing software Gwyddion [9]. These are compared to the real values of 5 µm and 100 nm respectively to determine calibration factors. For the lateral directions, a scaling factor as a percentage is created from this data. This was input into the software to influence the x- and y-piezo feedback loops so that 1 µm in the software would accurately represent 1 µm in reality. Finally, the z value is reported as a voltage from 0 to 50 V. So, there is a calibration factor of 176 nm V^{-1} (found via the sample microstructure) used in this work to translate from voltage to height, resulting in a final vertical maximum range of 9 µm. As seen later, this was not a large enough range for some of our samples.

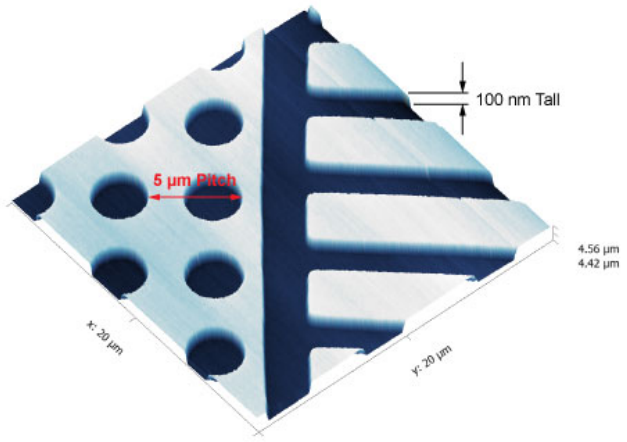


FIG. 3: A scan of the HS100-MG microstructure with sizes labelled, to be used for calibration [7, 8].

B. Image Analysis

The output data from the AFM is represented as a large matrix of greyscale pixel values, from a minimum to a maximum value (V_{min} to V_{max}). The first operation to perform on this data is to convert each pixel's value from volts (V) to nanometres (nm), via the calibration factor discussed above. Then the position of each pixel can be converted to a distance by knowing the resolution of the image (20 µm for all images in this report). Next, the images are post-processed via the techniques described in sections 2.B.1-4. Finally, the statistical roughness parameters described in sections 2.B.5-7 can be computed.

For the following sections, it is useful to define some common variables. They are shown on figure 4, which is an example profile in one-dimension, but all variables have a simple extension to two-dimensions. z_i is the height of the i$^{\text{th}}$ point as measured from an arbitrary zero-point. $\bar{z}$ is the average (mean) of all the points in the profile, and $r_i = z_i - \bar{z}$ is the height of the i$^{\text{th}}$ point as measured from this average, which is the measure that will mainly be used. Lastly, N is the total number of pixels in the profile.

1. *Scar detection and removal*

Due to how the AFM operates, scanning across the image in lines, sometimes there is a jolt of the machine, or interference causing many pixels to be incorrect. Examples of this are shown in figure 6a. To detect these artifacts, we compare

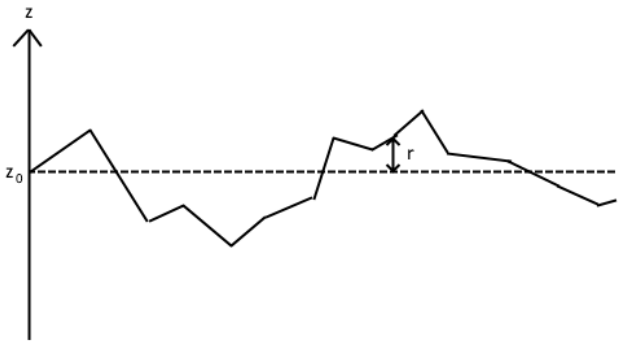


FIG. 4: An example profile in one-dimension, with statistical measures shown.

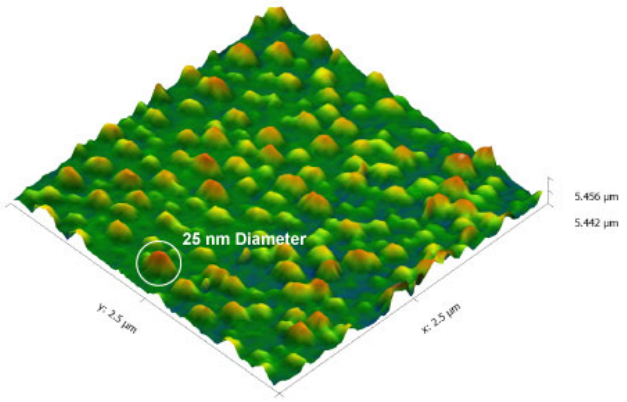


FIG. 5: An image of Co nanoparticles created with an AFM [7]. The horizontal size of the image is 2.5 µm and the vertical size is 15 nm.

pixel values with the values of pixels around them, and if their values differ by more than 30%, and there is a line of a minimum length of 16 pixels, that line is marked as a scar. Then, to remove each scar its pixels are replaced by a linear interpolation of the values of the pixels around the scar (above and below since the scanning direction is horizontal).

2. *Levelling layers by median of differences*

Next, since the AFM takes many minutes to make a scan, the equilibrium position can change in this time, so two lines can be offset by a false height. To fix this we use the 'median of differences' method [9], where each line is shifted so that the median of the height difference between all of the vertical pixels in the row and the row above it becomes zero.

3. *Minimum/mean value subtraction*

Value subtraction means subtracting a value from each pixel in the image. This value could be any scalar, but the minimum height is often picked for easier comparison between images, as in figures 9 & 10. The mean height can also be used, which is necessary for the autocorrelation outlined in section 2.B.7.

4. *Plane-levelling (unused)*

One final post-processing technique is to fit the entire image to a plane. This is useful when imaging surfaces such as figure 5, where the individual atoms on a flat surface are the focus. To plane-level, linear regression is used to fit a flat plane (3 degrees of freedom) to the image. In this work, this technique is not used as we are interested in the underlying shape of the surface, which will by nature not be a plane. It is included here as useful context for discussion.

5. *Statistic 1: Arithmetic average height (S_a)*

The average height is a measure of the average deviation of pixels' heights from the mean value. It is given by

$$S_a = \frac{1}{N}\sum_i^N |r_i|, \tag{1}$$

where S_a is the average height and other variables have the meanings given above. This is the simplest statistic and is used as a very simple description of roughness, but it does not capture any information about the shape of the roughness.

6. *Statistic 2: Root mean square roughness (S_q)*

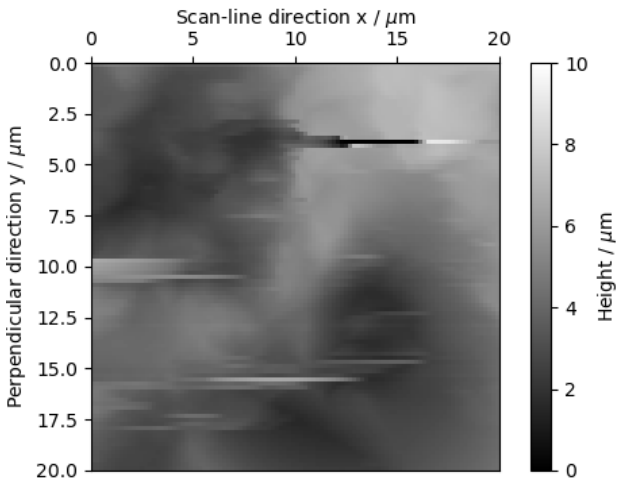
The Root-Mean Square (RMS) roughness is a measure of the standard deviation of the distribution of surface height. It is given by

$$S_q^2 = \frac{1}{N}\sum_i^N r_i^2, \tag{2}$$

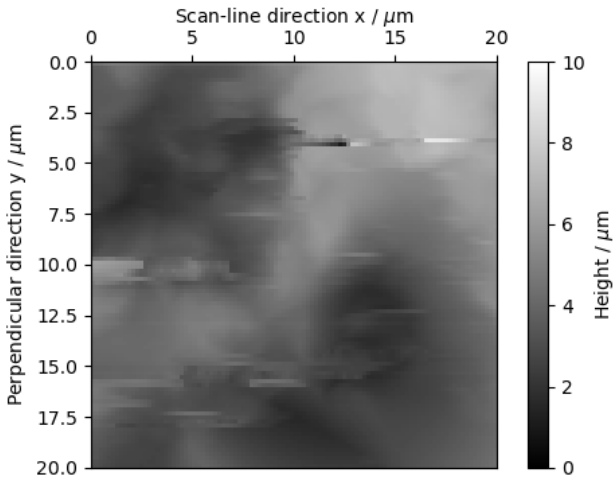
where S_q is the average height and other variables have the same meanings as above. This is a more specific statistic than the arithmetic average height and it is more sensitive to a large deviation from the mean line. It will change more depending on how many large peaks and valleys are in the profile.

7. *Statistic 3: Autocorrelation (S_{al} & S_{tr})*

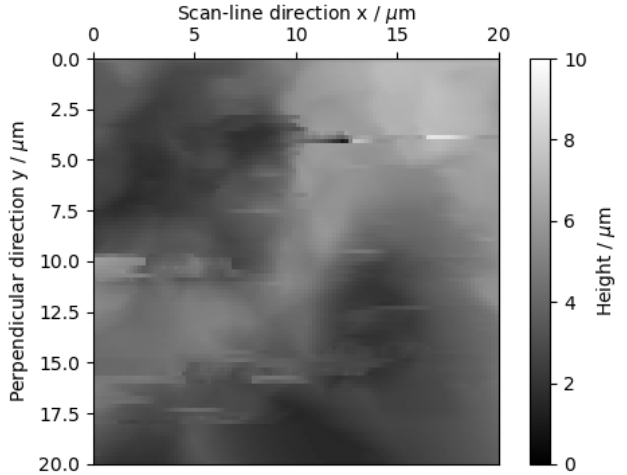
The Autocorrelation Function (ACF) is a quantitative measure of the similarity of a profile with itself and itself at a laterally shifted position. It is therefore useful to examine the large-scale structure of a surface scan. For a continuous, idealised



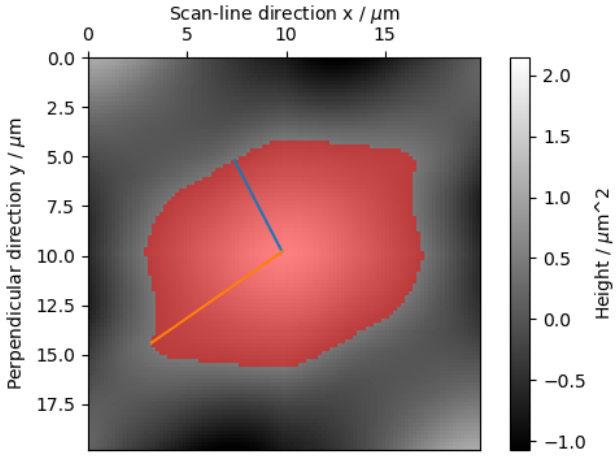
(a) The raw image data, zero-levelled.



(b) The data after scar-removal.



(c) The data after median line-levelling. This is a subtle effect, made more obvious by comparing the dark and light regions.



(d) This is the autocorrelation function of the data. It is one pixel smaller in x and y than the data, due to the nature of the ACF. Masked in red is the region where the ACF is above 20% of its maximum value. The shortest autocorrelation length is shown in blue and the longest in orange. These are respectively 5.19 μm and 8.04μm.

FIG. 6: An example process for one of the scans, including steps of postprocessing and statistical analysis. This scan is the F10 scan, also seen in figure 10h, but with a different colour scale.

1-dimensional profile it is given by

$$\mathcal{ACF}(\delta x) = \frac{1}{L}\int_0^L r(x)r(x+\delta x)dx, \tag{3}$$

where L is the length of the profile, and $r(x)$ is the height at a distance x along it. The ACF is used in the same domain as the original function, so the shift distance δx will be in the range $0 < \delta x < L$. Note that the function $r(x)$ is treated as periodic so $r(x + L) = r(x)$. The ACF is then normalised to have a value of 1 at a shift distance of 0, to allow better comparison between the ACFs of different profiles.

For a discrete set of data, as we have here, and extending the ACF to two dimensions, one can express the discrete ACF as

$$\mathcal{ACF}(m,n) = \frac{1}{(N-n)(M-m)}\sum_{l}^{N-n}\sum_{k}^{M-m} r_{k+m,l+n}r_{k,l} \tag{4}$$

where n and m are the row and column indices respectively, and similarly N and M are the total number of rows and columns. Finally, $r_{k,l}$ is now the value of the pixel at row k and column l. This ACF is also usually plotted in the same 2D area as the original image, as seen in figure 6d.

The ACF does not tell much without some further analysis. In the following, two statistics derived from the ACF are introduced: the autocorrelation length S_{al} and the texture ratio S_{tr}.

First, an area on the ACF is masked where it is larger than 20% of its largest value (larger than 0.2 if normalised to 1) - this is shown in red in figure 6d. Then, the direction in which the ACF reduces the fastest is found (i.e. the angle at which the ACF drops below 0.2 in the shortest distance). This is the blue line in figure 6d. This is called the autocorrelation length S_{al}, and represents the distance between distinct and unrelated features on the surface. Points on a surface which are larger than this distance away from each other can be described as being created by different surface events. Specifically for this investigation, this can be thought of as features which were created by different microparticles carving channels through the glass surface. S_{al} can range from 0 to infinity, from a completely random surface to a completely periodic surface, respectively.

In a similar manner, the direction in which the ACF reduces the slowest is found (the orange line in figure 6d), and the ratio of the autocorrelation length with this length is taken. This is called the texture ratio S_{tr}, whose value ranges from $0 < S_{tr} < 1$. If S_{tr} is close to 1 then the surface does not have a directionality, but if it is closer to 0 it shows a directional preference of the surface. White noise would have a texture ratio of 1, whilst a pattern sinusoidal in x and constant in y would have a texture ratio close to 0.

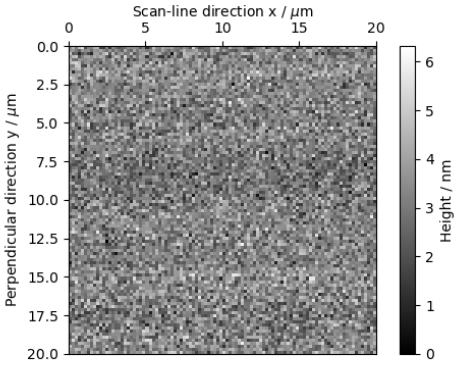
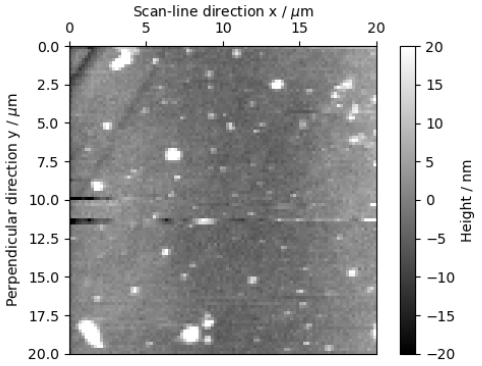


FIG. 7: An image taken with no sample under the tip of the AFM, to determine the background noise in the imaging process.

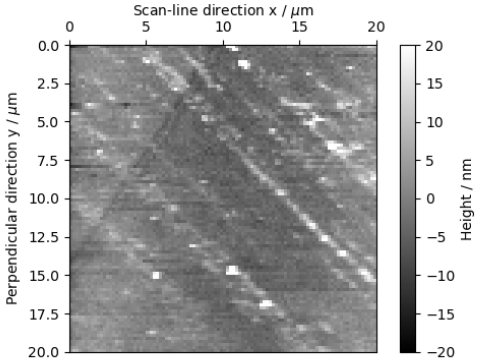
3. RESULTS

First, a scan of the smooth glass was taken, shown in figure 8. Figure 7 shows an image taken with no surface under the needle of the AFM to investigate how much background noise was evident in the investigation.

The surfaces were then sanded and imaged in several locations. These images were then post-processed according to sections 2.B.1-3, the results of which are shown in figures 9 and 10. Part of this process included creating a criterion for discarding some of the scans, as lots had large swathes of dead



(a) An image of the smooth, unsanded glass.



(b) A second image of the smooth, unsanded glass.

FIG. 8: The smooth glass. Note the smaller scale, on the order of nm.

area which was outside the vertical range of the AFM. In order to discard erroneous scans while not discarding all of the data it was decided that if a scan had more than a threshold of 5% of its pixels at its minimum or maximum value, it was erroneous. This can be seen mostly in figure 9, as for example image G12 is clearly erroneous but it is not so certain for images such as G23 or G18.

Then, the statistics as described in section 2.B.5-7 were computed for each image and combined into the values seen in table I, with the error on each statistic being the standard error. These statistics are also plotted in figure 11. Most of the removed images have roughness statistics larger than the unremoved case, which is reasonable since they have a larger height-variation. The outlier here is image G12, which has very low roughness parameters. This makes sense as it is nearly a flat plane at the height-range limit.

Also, most of the removed images have a larger autocorrelation length, which is reconcilable as one must look further on a flat plane to find new features. Most also have a larger texture ratio, as planes of noise are non-directional, so overall the erroneous images will have less directionality than those with more texture.

FIG. 9: The G120 scans. Each was taken with the microscope. They have been post-processed. Note the amount of 'dead' area. Scans marked with an * have more than 5% of their area at the minimum or maximum values.

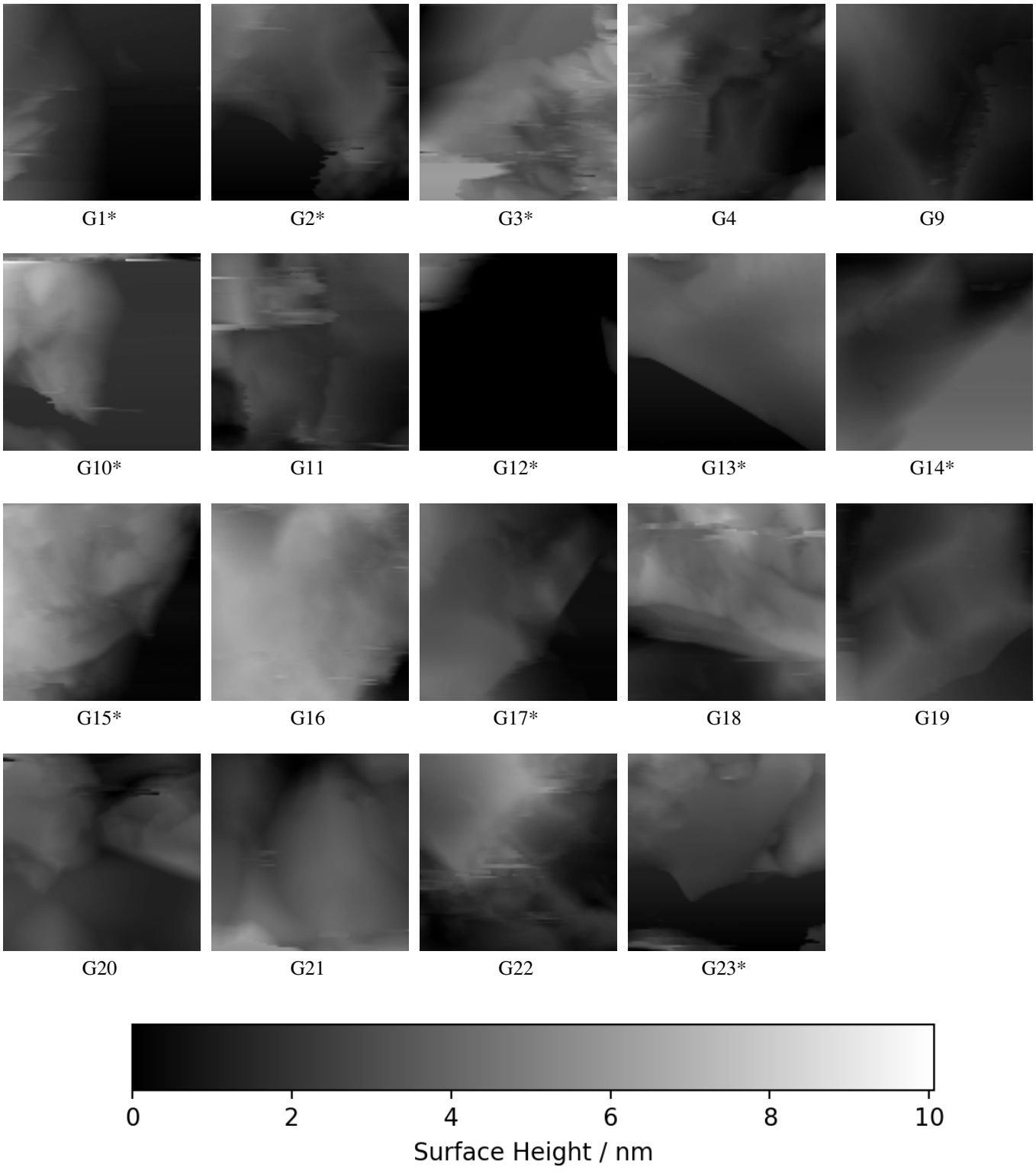
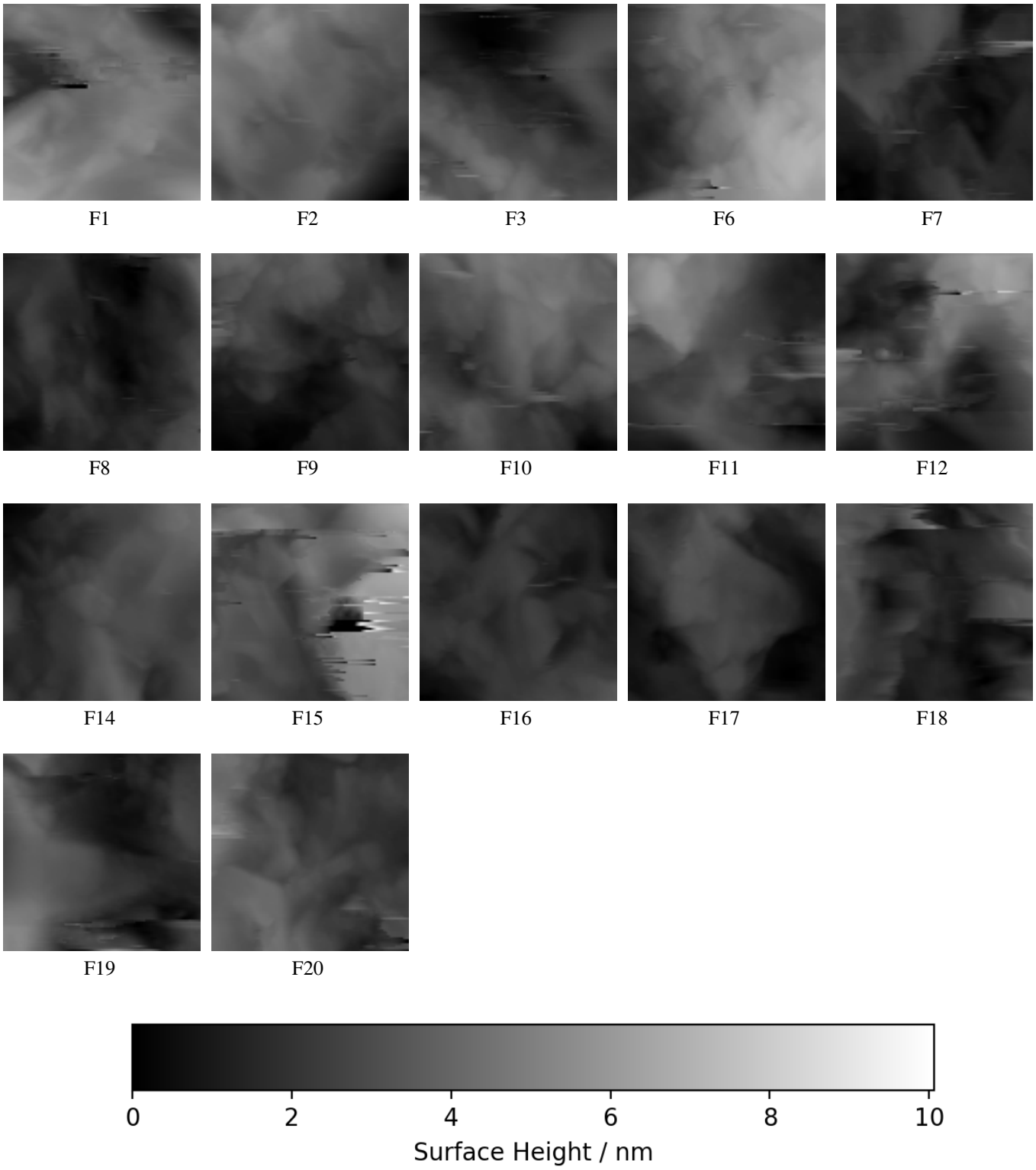
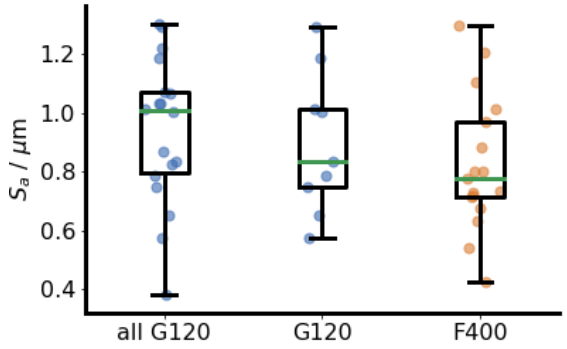
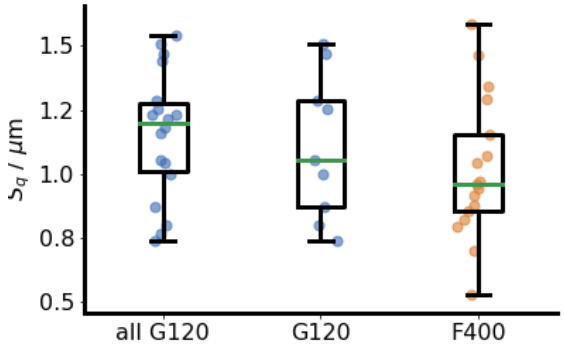


FIG. 10: The F400 scans. Each was taken with the microscope. They have been post-processed. Note the lesser amount of 'dead' area.

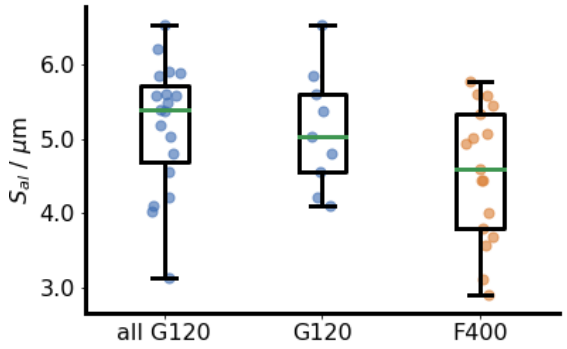




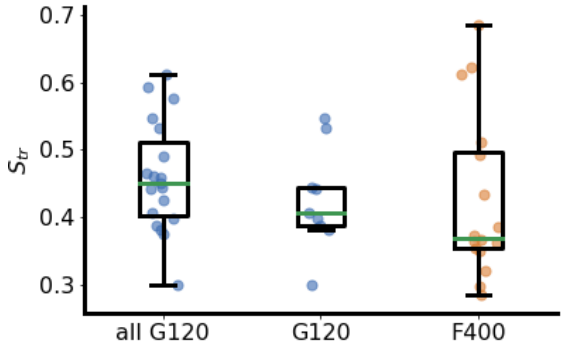
(a) Average surface roughness S_a. Note that outlier at 1.85 µm (from image G15*) is not shown in the 'all G120' plot for visual clarity, but still used in the statistics. The lowest value of 'all G120' is from image G12*.



(b) RMS roughness S_q. Note that outlier at 2.13 µm (from image G15*) has is not shown in the 'all G120' plot for visual clarity, but still used in the statistics.



(c) Autocorrelation length S_{al}. The lowest value of 'all G120' is from image G12*



(d) Texture ratio S_{tr}. Note that outlier at 0.88 (from image F16) is not shown in the F400 plot for visual clarity, but still used in the statistics.

FIG. 11: Boxplots of the chosen statistics. Note that two plots are included for the G120 microparticles, one with the plots with dead area, one without. The green line represents the median data point, the box the interquartile range, and the caps the full range. The caps only extend as far as 3 times the interquartile range from the median, as seen in figure 11d.

Microparticle	Surface Average Roughness S_a / µm	RMS Roughness S_q / µm	Autocorrelation Length S_{al} / µm	Texture Ratio S_tr
Noise measurement	7.0×10^{-4}	8.8×10^{-4}	N/A	N/A
Smooth glass	$(3.06 \pm 0.04)\times10^{-3}$	$(4.8 \pm 0.5)\times10^{-3}$	0.8 ± 0.2	0.13 ± 0.04
G120	0.90 ± 0.08	1.11 ± 0.09	5.1 ± 0.2	0.43 ± 0.02
all G120	0.99 ± 0.07	1.21 ± 0.07	5.2 ± 0.2	0.46 ± 0.02
F400	0.83 ± 0.05	1.02 ± 0.07	4.6 ± 0.2	0.45 ± 0.04

TABLE I: Roughness Statistics calculated with methods from sections 2.B.5-7. S_a: average surface roughness; S_q: RMS roughness; S_{al}: autocorrelation length; S_{tr}: texture ratio.

4. DISCUSSION

In this section we will discuss the validity of the experimental procedure and data analysis as well as draw some sensible conclusions from the collected data. First, we look at the interpretations of the statistics presented in the study, followed by a discussion of the limits of the AFM, including artifacts and range limitations. Then, the impacts of the post-processing on the results is considered. Finally, reasonable conclusions are extracted from the study as well as further thoughts on the choice of statistics.

A. Interpretation of Statistics

First, the interpretation of each statistic used in this study is discussed, followed by a justification of how each is expected to differ for the two samples versus how it measurably differed.

To start, the first two statistics used are related to each other. Both the arithmetic average height S_a and root mean square roughness S_q are 'order of magnitude' measures, meaning that they can determine between the size of surface features, but not the type of features. However, we would expect that both measures would be larger for the larger microparticles due to them creating larger surface features. As seen in figure 11 and table I these measures are smaller for the smaller F400 microparticles, but due to the large spread of values they are not statistically unique (see appendix B).

Continuing with the autocorrelation statistics, the autocorrelation length S_{al} represents the distance between significantly different features on the surface, and the texture ratio S_{tr} represents the directionality of the surface (1 being a homogeneous surface and 0 being a completely directional surface). In this investigation, we expect S_{al} to be larger for the larger microparticles as they would create larger surface features. As seen in figure 11 and table I this is the case, but again they are not statistically different (see appendix B). Finally, because of the model we consider, we do not expect the texture ratio to be different for the two microparticles, as both surfaces will have the same directionality from being formed by the hard microparticles carving features into them. This is the case as seen in table I. Both grades of microparticles have the same texture ratio with a 95% confidence interval (see appendix B).

B. Artifacts and Range Issues

As seen in figure 9, much of the area on some of the G120 scans goes off the vertical scale as it is either too high or too low to scan. This is illustrated by the monotone patches of black or grey, and means that the full detail of the profile was not obtained as 10 nm of vertical range was not enough. The larger G120 microparticles had a size much greater than the vertical range, so they were able to carve channels and create slopes which easily exceed the range of the AFM used.

Thus, the statistics S_a and S_q become lower bounds as any 'lost' area would be above or below the maximum or minimum flat planes, respectively. The effects this would have on S_{al} and S_{tr} are harder to deem due to the complex nature of the autocorrelation. This means that with the full range of the sample, the two roughness statistics could have been statistically different as the blue boxplots in figure 11 would be higher up in the plot.

Moreover, the removal of some images results in a sampling bias by not including the G120 scans with flat areas. This is minimised by having a computational-based discarding criterion. However, not disposing of the erroneous images would invalidate the data because they are not representative of the real surface. To solve this dilemma, smaller microparticles could be used that would not exceed the AFM range, or a different AFM could be used. In future this approach would be implemented to yield better final results.

Finally, another source of error is the artifacts in the images. As seen in figure 6a, there are many scars on the initial scans, which result in lost data at these points in the image. This could be due to many reasons, but a likely one is that the glass plates were improperly dried after being sanded, as the smooth glass plate showed fewer of these streaks. Ultimately, however, these do not change the final statistical values too much to make a significant difference.

C. Post-Processing Issues

A large problem in the field of microscopy is what post-processing to use on the images obtained. This is a difficult choice, as post-processing can alter roughness parameters' measured values far from their real values [10, 11]. For example, by making the surface seem flatter than it actually is. In this report, it was decided to not plane-level the images as the surfaces were not assumed to be flat, and the interest was not in specific surface features (like figure 5), but the surface as a whole. Plane-levelling the images may have resulted in different values for the roughness parameters and allowed a different conclusion to be drawn, which is precisely the issue with the selection of post-processing. The number of post-processing steps, therefore, has been minimised for the results in this report.

D. Conclusions on Sample Differentiation

In this section, we discuss whether we can differentiate the two microparticles based on the computed statistics. As mentioned above, S_a, S_q, and S_{al} are not statistically different for each microparticle grade with a 95% confidence interval using the G120 scans. However, when using all of the G120 scans they are statistically unique. So, given the full range, they would also be differentiable. Given one of the glass plates at random, it would be possible to tell which one it was if multiple scans could be taken with the AFM. First off, because these roughness statistics would be different, but also with this

specific AFM because either very few or very many of the scans would go off the vertical scale, so which particle was used could be deduced. With an AFM with a larger vertical range, it is possible that only the statistics would be needed to tell the plates apart as their values would be further from one another.

A final consideration is why the roughness parameters are so similar when the microparticle sizes are so different. The sizes of the particles are not on the range of the horizontal scanning range of the AFM, so in taking these tiny images, the full picture might not be being gleamed. Therefore, an AFM with a larger scanning range might be more beneficial.

5. CONCLUSIONS

Overall, determining which microparticle was used to sand the two samples in the report is possible using an AFM. This may not be possible with a single image but certainly with multiple. They are not very distinct in common roughness parameters, but a difference can clearly be seen in overall range since around 50% of the 19 G120 samples broke the range of the AFM whilst none of the 17 F400 second samples did. Therefore, given an AFM with a larger range, perhaps the roughness parameters would have been more dissimilar. In this case, an atomic-force microscope was a useful tool, but the specific AFM used in this report had some range limitations.

REFERENCES

[1] J. Edward, 1956. *A Short History of the Electron Microscope*. Bios, 27(1), pp.33–37.

[2] G. Binnig, C. Quate & C. Gerber, 1986. *Atomic force microscope*. Physical review letters, 56(9), pp.930–933.

[3] P. Eaton & P. West, 2010. *Atomic force microscopy*, Oxford: Oxford University Press.

[4] N. Chkhalo, N. Salashchenko & M. Zorina, 2015. *Note: A stand on the basis of atomic force microscope to study substrates for imaging optics*. Review of Scientific Instruments, 86(1), p.016102.

[5] E. Gadelmawla, M. Koura, T. Maksoud et al., 2002. *Roughness parameters*. Journal of materials processing technology, 123(1), pp.133–145.

[6] D. Nečas & P. Klapetek, 2013. *One-dimensional autocorrelation and power spectrum density functions of irregular regions*. Ultramicroscopy, 124, pp.13–19.

[7] Thorlabs, 2017. *EDU-AFM1 Educational Atomic Force Microscope User Guide*. Retrieved from https://www.thorlabs.com.

[8] BudgetSensors, accessed Jan. 2021. *HS-100MG height calibration standard*. https://www.budgetsensors.com/height-calibration-standard-hs-100mg

[9] D. Nečas & P. Klapetek, 2012. *Gwyddion: an open-source software for SPM data analysis*. Central European Journal of Physics, 10(1), pp.181–188.

[10] D. Nečas, P. Klapetek & M. Valtr, 2020. *Estimation of roughness measurement bias originating from background subtraction*. Measurement Science and Technology, 31(9), p.15.

[11] D. Nečas, P. Klapetek & M. Valtr, 2020. *How levelling and scan line corrections ruin roughness measurement and how to prevent it*. Scientific Reports, 10(1), p.15294.

Appendix A: Measurement Uncertainties

The roughness parameters were calculated for each image, and then combined by calculating the mean and standard error of the individual measurements. The mean is calculated using the equation,

$$\bar{V} = \frac{1}{N}\sum_{i=1}^{N} V_i, \tag{A1}$$

where $\bar{V}$ is the mean measurement of the voltage V and V_i are individual measurements of V [This equation, like all of the equations included in Appendix A, is based on the error analysis formula given in I. G. Hughes and T. P. A. Hase, *Measurements and Their Uncertainties*, Oxford University Press: Oxford (2010).].

The sample standard deviation, σ_{sample}, of the set of measurements is worked out using the equation,

$$\sigma_{sample} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} d_i^2}, \tag{A2}$$

where $d_i = \bar{V} - V_i$. The uncertainty in the measurement of $\bar{V}$ is taken to be its standard error, α_V, where

$$\alpha_V = \frac{\sigma_{sample}}{\sqrt{N}}. \tag{A3}$$

Appendix B: Statistical Tests

In this report, whenever two distribution means are compared it is using the two-sample *t*-test. This is a method used to test whether the means of two samples are equal or not. First, the pooled variance s_p is calculated using

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}, \tag{B1}$$

where n_1 and n_2 are the number of items of the two distributions, and s_1 and s_2 are the standard deviations of the two distributions. Then, the test statistic t is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}, \tag{B2}$$

where $\bar{x}_1$ and $\bar{x}_1$ are the means of the two distributions. Finally, this is compared with the 95% confidence interval value of $t_{95} = 1.96$. So if $t > t_{95}$ then the means are statistically different.