

# 1 Introduction

Our microscope is supposed to scan a tissue sample, which covers some subset of a slide of size  $L \times W$ . Initially, it performs a quick scan of the whole slide with a small magnification setting. This scan is recorded in `qptiff` format.

Using that image, the microscope determines which parts of the slide actually have tissue in them. It scans those sections under a larger magnification, using high-powered fields of size  $l \times w$ . Covering the whole tissue may require  $O(1000)$  fields.

The scanning is configured with an overlap fraction  $f = 20\%$ , so that some parts of the image are scanned twice. The rightmost  $\frac{1}{2}fl$  of each field overlaps the leftmost  $\frac{1}{2}fl$  of the field to its right. Similarly, the bottom  $\frac{1}{2}fw$  of each field overlaps with the top  $\frac{1}{2}fw$  of the field under it. These overlaps are used to assess differences between the fields, to correct those differences, and ultimately to assess systematic uncertainties based on the remaining differences.

In the process of scanning, the microscope moves from one field to the next. The slide is supposed to move horizontally by  $l(1 - f)$  and vertically by  $w(1 - f)$ . In practice, due to mechanical imprecisions in the microscope, the slide does not always move by exactly this amount, and so a simple, out of the box tiling of the fields (even taking the overlap regions into account by shaving off the outer  $\frac{1}{2}fl$  and  $\frac{1}{2}fw$  from each side) does not give an accurate picture of the tissue sample. If fields move away from each other, cells in the boundary region will appear larger than they should due to duplicated pixels, and in extreme cases they may even be cut in half. If they move towards each other, those cells will appear *smaller* than they should, and for an extreme shift small cells might be completely gone. If they move sideways, the cell's shape might be ruined and it may not be identified properly.

Here, we describe how we measure the actual positions of the fields based on the images themselves.

## 2 Pairwise alignment

The first step is to look at each pair of overlapping fields and measure the relative shift of those two fields. Once we assemble all of those results, we stitch them together to obtain the final result.

### 2.1 Measuring the alignment

For any individual pair of fields that overlap, we can align those two fields by comparing the images to each other. For example, fig. 1(a) shows the overlap of two fields connected by a corner. It is clear that the two fields are misaligned and that the field shown in green needs to move up and to the right with respect to the pink field. The result, with the size and direction of the shift determined using the procedure described below, is shown in fig. 1(b) and indicates much better agreement between the fields.

In order to determine the size and direction of the shift, we compute the *cross correlation* between the two overlap images. The cross correlation  $C$  between images  $A$  and  $B$  is defined as

$$C(\delta\vec{r}) = (A \star B)(\delta\vec{r}) \equiv \sum_{\vec{r}} A(\vec{r})B(\vec{r} + \delta\vec{r}) \quad (1)$$

and describes how similar the two images are when shifted by  $\delta\vec{r}$  with respect to each other. The value of  $\delta\vec{r}$  that maximizes  $C$  gives the optimal shift. We want to achieve precision better than a single pixel, so we fit the region around the tallest peak of the cross-correlation function to a cubic polynomial and find its maximum  $\delta\vec{r}_{\text{max}}$ , as shown in fig. 2(b) for a one-dimensional example. For the real images, we fit the region around the peak to a two-dimensional cubic polynomial.

### 2.2 Uncertainty on pairwise alignment

We also want to estimate the uncertainty on  $\delta\vec{r}$ . This will become important in the next step, when we want to combine all of the alignment results and assign a larger weight to alignments with smaller uncertainty. To obtain this uncertainty, we first estimate  $\sigma_C(\delta\vec{r}_{\text{max}})$ , the uncertainty on the value of the cross correlation

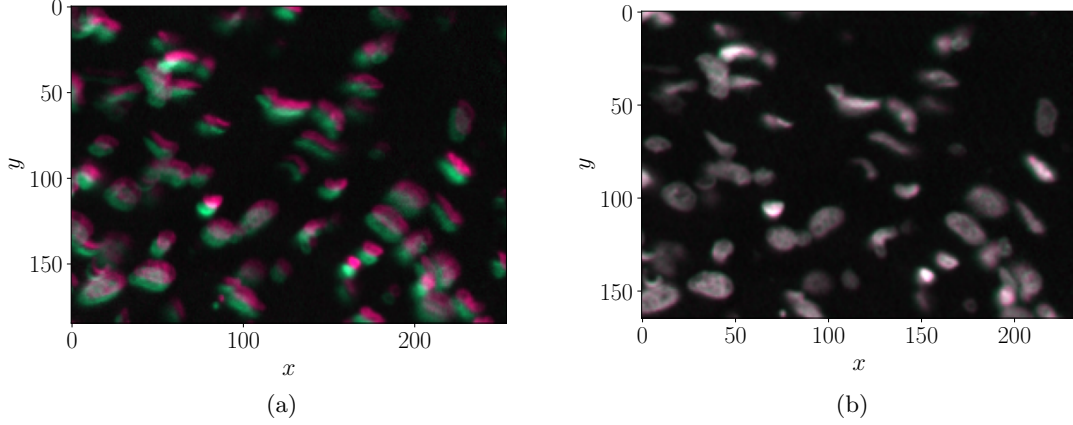


Figure 1: The overlap of two fields (a) before and (b) after alignment. The two fields' images are shown in pink and green, respectively. Because those colors are complementary, gray areas mean that the two fields have the same intensity in that spot.

function at its maximum. We shift the images by  $\pm 0.5\delta\vec{r}_{\max}$ , so that they are now in their best-aligned position (as shown in fig. 1(b)). We can call the shifted images

$$\begin{aligned} A'(\vec{r}) &= A\left(\vec{r} - \frac{1}{2}\delta\vec{r}_{\max}\right), \\ B'(\vec{r}) &= B\left(\vec{r} + \frac{1}{2}\delta\vec{r}_{\max}\right). \end{aligned} \quad (2)$$

Plugging into eq. (1),

$$\begin{aligned} C(\delta\vec{r}_{\max}) &= \sum_{\vec{r}} A'\left(\vec{r} + \frac{1}{2}\delta\vec{r}_{\max}\right) B\left(\vec{r} + \frac{1}{2}\delta\vec{r}_{\max}\right) \\ &= (A' \star B')(0) \end{aligned} \quad (3)$$

At this point, we consider that the two images are as well aligned as we can make them, and are supposed to be two independent measurements of the same thing. We then estimate the error on an individual pixel's intensity, in either  $A'$  or  $B'$ , as

$$\sigma_{A'}(\vec{r}) = \sigma_{B'}(\vec{r}) = |A'(\vec{r}) - B'(\vec{r})|. \quad (4)$$

By standard techniques of error propagation, we then derive

$$\begin{aligned} \sigma_C(\delta\vec{r}_{\max}) &= \sum_{\vec{r}} \left( [A'(\vec{r})\sigma_{B'}(\vec{r})]^2 + [\sigma_{A'}(\vec{r})B'(\vec{r})]^2 \right) \\ &= ([A']^2 + [B']^2) \star \sigma_{A',B'}^2(0) \end{aligned} \quad (5)$$

We now use  $\sigma_C(\delta\vec{r}_{\max})$  to derive  $\sigma_{\delta\vec{r}_{\max}}$ . To motivate the calculation, we first consider the 1-dimensional case shown in fig. 3. As illustrated there, we need to find  $\sigma_{\delta x_{\max}}$  so that  $C(\delta x_{\max} \pm \sigma_{\delta x_{\max}}) = C(\delta x_{\max}) - \sigma_C$ . Assuming  $\delta x$  is small, we can do this by expanding

$$C(\delta x) \approx \delta x_{\max} + \frac{1}{2}C''(\delta x_{\max})(\delta x - \delta x_{\max})^2 \quad (6)$$

and find that

$$\sigma_{\delta x_{\max}}^2 = \frac{2\sigma_C(\delta x_{\max})}{C'''(\delta x_{\max})}. \quad (7)$$

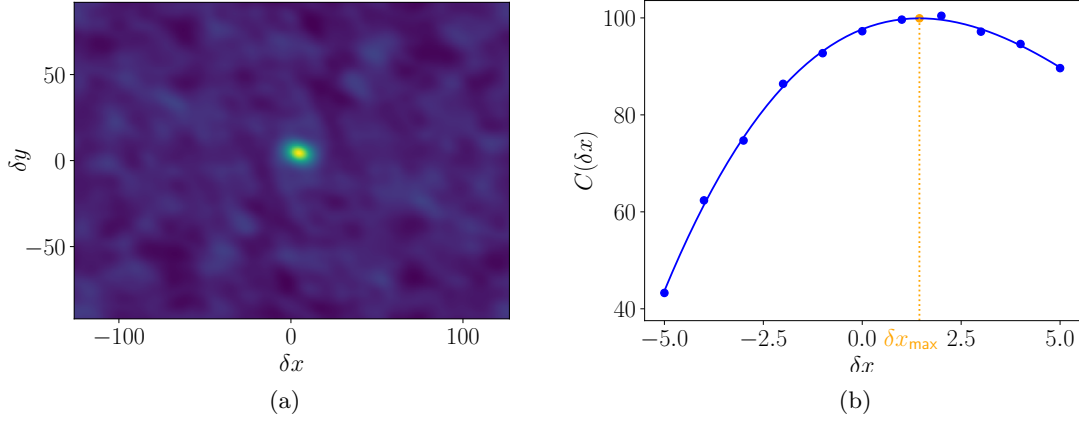


Figure 2: (a) Plot of the cross-correlation function for the overlap shown in fig. 1. (b) Example of the procedure for aligning two images to sub-pixel precision. The blue points show the cross correlation function for pixel shifts, calculated using eq. (1). The blue curve shows a polynomial fit to the cross correlation function, and the orange point and line show the maximum of the polynomial fit.

By an analogous argument, we find the equivalent result for two dimensions. Because  $\delta\vec{r}$  has two components,  $\delta x$  and  $\delta y$ , we have a  $2 \times 2$  covariance matrix,

$$\mathbf{cov}_{\delta\vec{r}_{\max}} = 2\sigma_C(\delta\vec{r}_{\max})\mathbf{H}^{-1}(\delta\vec{r}_{\max}), \quad (8)$$

where  $\mathbf{H} = \begin{pmatrix} \frac{\partial^2 C}{\partial \delta x^2} & \frac{\partial^2 C}{\partial \delta x \partial \delta y} \\ \frac{\partial^2 C}{\partial \delta x \partial \delta y} & \frac{\partial^2 C}{\partial \delta y^2} \end{pmatrix}$  is the Hessian matrix of  $C(\delta\vec{r})$ .

For the final step, we multiply the covariance matrix by a factor determined empirically, as we will describe in section 4.2. This factor is found to be  $\frac{1}{16}$ , which means that the original error estimate was too conservative by a factor of 4. One possibility for this overestimate is that, if the alignment is not perfect, eq. (4) overestimates the error on pixel intensity in regions where cell boundaries do not line up correctly. Whatever the explanation, the plots in section 4.2 will show that the corrected covariance matrix

$$\mathbf{cov}_{\delta\vec{r}_{\max}}^{\text{corr}} = \frac{1}{8}\sigma_C(\delta\vec{r}_{\max})\mathbf{H}^{-1}(\delta\vec{r}_{\max}) \quad (9)$$

is a good estimate of our uncertainty.

### 2.3 Detecting bad overlaps

The error estimation procedure also gives us an opportunity to detect when a particular overlap does not contain enough information to align it. This happens when the overlap region does not contain any cells, and sometimes also when it contains one or two. In this case, the cross correlation function simply consists of noise, as shown in fig. 4.

The procedure in section 2.1 looks at the tallest peak of the cross-correlation function. For a good overlap, there is often only one peak, or one tall peak and several other, much shorter ones. Figure 4(b), on the other hand, shows a central peak near (0,0) and a diagonal structure, similar to the structure in fig. 4, containing several yellow areas. The error estimation tells us whether any of the other peaks are significant: if any other peak at  $\delta r'$  has  $C(\delta r') \geq C(\delta r_{\max}) - 3\sigma_C(\delta r_{\max})$ , we do not know, to  $3\sigma$  confidence, which peak is the correct one, and therefore we mark the overlap as bad. The result from bad overlaps are not used in the final stitching procedure.

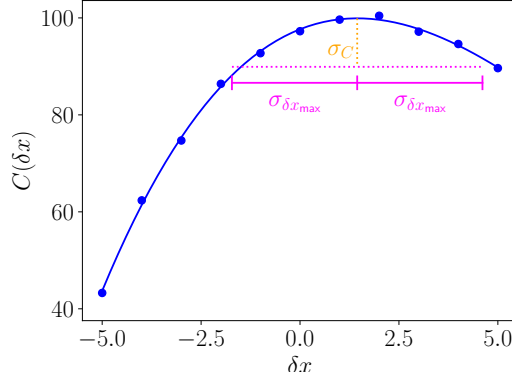


Figure 3: Illustration of the procedure for error estimation using the same example curve as in fig. 2(b). Given the error on  $C$  at the peak,  $\sigma_C$ , shown in orange, we can use the second derivative of the curve to estimate the error on  $x_{\max}$ ,  $\sigma_{x_{\max}}$ , shown in pink.

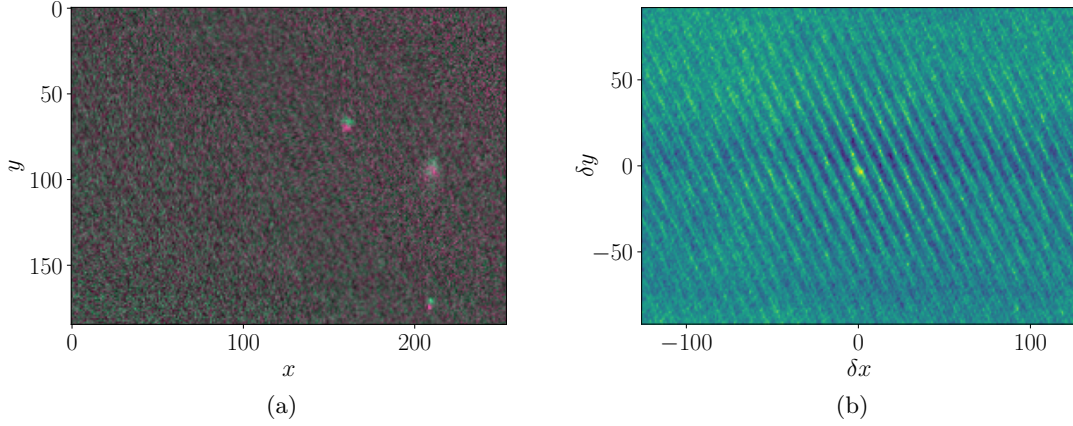


Figure 4: (a) An almost empty image overlap, similar to fig. 1(a), but with the pixel intensities multiplied by a factor of 10 with respect to the ones there. The image contains mostly noise, with three faint features. (b) The cross-correlation function for the overlap in (a).

### 3 Stitching

After we derive the individual alignment results for each overlap, we can “stitch” the results together to obtain the final positions of each field. We see each alignment result as a Gaussian constraint on the relative positions of those two fields, and combine all of those constraints together to find the likelihood  $L$  for alignment positions  $\vec{r}_i$ :

$$-2 \ln L_{\text{alignment}}(\vec{r}_1, \dots, \vec{r}_N) = \sum_o (\vec{r}_{io} - \vec{r}_{jo} - \delta\vec{r}_o - \vec{r}_{io}^n + \vec{r}_{jo}^n)^T \mathbf{cov}_o^{-1} (\vec{r}_{io} - \vec{r}_{jo} - \delta\vec{r}_o - \vec{r}_{io}^n + \vec{r}_{jo}^n) \quad (10)$$

In this parameterization,  $\vec{r}_k$ , with  $k$  running from 1 to the number of fields  $N$ , is the position of the field indexed by  $i$ ;  $\vec{r}_k^n$  is the field’s nominal position, given by the microscope software;  $o$  indexes the overlaps;  $i_o$  and  $j_o$  are the indices of the two fields whose overlap is described by  $o$ ;  $\delta\vec{r}_o$  is the  $\delta\vec{r}_{\max}$  determined for that overlap, as described in section 2.1; and  $\mathbf{cov}_o$  is the covariance matrix from eq. (8).

The likelihood above includes all of the information derived from alignment. However, it is degenerate: shifting all fields by the same amount leaves the equation unchanged. This can be fixed easily by just

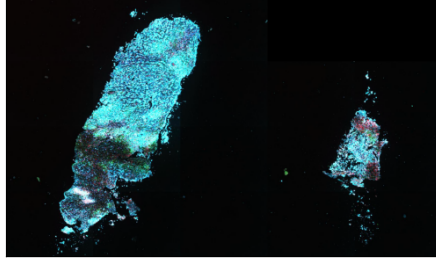


Figure 5: An example of an image, code name M21\_1, taken with the Vectra 3.0 microscope, with two disconnected islands.

choosing one point to remain fixed, but in general the image may contain multiple disconnected “islands”, as shown in fig. 5. In this example, the larger island on the left is covered by 33 fields, and the smaller one on the right is covered by 7. There are no overlaps between fields in one island and fields in the other, and moving one island with respect to the other also leaves eq. (10) unchanged.

A simple solution is to detect the islands using algorithms from graph theory and align each island separately, but there are cases where even this will fail. For example, the islands may be connected by a single overlap which is not “bad”, per the definition in section 2.3, but has a large uncertainty, with the result that the two almost-islands are aligned badly with respect to one another when the single overlap is the only constraint. We would like a procedure that automatically takes all of these cases into account.

To accomplish this, we construct a model that allows for two sources of error in the microscope’s positioning.

1. When the controlling computer signals the microscope to move by  $\Delta\vec{r} = (\Delta x, \Delta y)$ , imperfect calibration may result in an actual movement of  $\mathbf{A}\Delta\vec{r}$ , where  $\mathbf{A}$  is a  $2 \times 2$  matrix to be measured from the data.
2. Each movement of the microscope involves a random error, so that the final, actual movement is  $\mathbf{A}\Delta\vec{r} + \delta\vec{r}$ .

The second source of error is random for each field. However, the first error, parameterized by the matrix  $\mathbf{A}$ , is generally a property of the microscope and is therefore common to all fields. This matrix is the way we connect the islands together, adding more terms to eq. (10) so that it is no longer singular.

We first assume that  $\mathbf{A}$  is close to the identity matrix  $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ . Therefore, although its deviations from the identity may affect the positioning between fields that are far apart by tens of pixels, it has negligible effect on an individual overlap. Therefore, the distribution of  $\delta\vec{r}$  may be approximated as the distribution of  $\delta\vec{r}_o$  from the overlaps  $o$ . We take the standard deviation, over all the overlaps, of the  $x$  and  $y$  components of  $\delta\vec{r}_o$ , and constrain the position of each rectangle using a Gaussian constraint, centered at  $\mathbf{A}\vec{r}_k^n$  and with a width of  $\sigma_x$  for the  $x$  component and  $\sigma_y$  for the  $y$  component. Or, in other words, we add to eq. (10) a term that looks like

$$-2 \ln L_{\text{constraint}}(\vec{r}_1, \dots, \vec{r}_N, \mathbf{A}) = \sum_k (\vec{r}_k - \mathbf{A}\vec{r}_k^n)^T \mathbf{\Sigma}^{-1} (\vec{r}_k - \mathbf{A}\vec{r}_k^n) \quad (11)$$

where  $\mathbf{\Sigma} = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}$ .

The constraint added in eq. (11) turns out to be weak, because the positioning errors  $\sigma_{x,y}$  are significantly larger than the errors from the pairwise alignments. In fact, if we modify  $\sigma_{x,y}$  by multiplying them by a few orders of magnitude up or down, the final alignment is basically unchanged.

The final likelihood function is the sum of eqs. (10) and (11):

$$\begin{aligned}
-2 \ln L(\vec{r}_1, \dots, \vec{r}_N, \mathbf{A}) = & \sum_o (\vec{r}_{io} - \vec{r}_{jo} - \delta \vec{r}_o - \vec{r}_{io}^n + \vec{r}_{jo}^n)^T \\
& \mathbf{cov}_o^{-1} (\vec{r}_{io} - \vec{r}_{jo} - \delta \vec{r}_o - \vec{r}_{io}^n + \vec{r}_{jo}^n) \\
& + \sum_k (\vec{r}_k - \mathbf{A} \vec{r}_k^n)^T \mathbf{\Sigma}^{-1} (\vec{r}_k - \mathbf{A} \vec{r}_k^n)
\end{aligned} \tag{12}$$

The most important feature of this equation is that, because  $\vec{r}_k^n$  are constants and  $\mathbf{cov}_o$  and  $\mathbf{\Sigma}$  are positive definite, it is a *convex quadratic* in the variables we are interested in solving for,  $\vec{r}_k$  and  $\mathbf{A}$ . Assembling  $\vec{r}_k$  and  $\mathbf{A}$  into a single, large vector  $\vec{v}$

$$\vec{v} = (x_1, y_1, \dots, x_k, y_k, A_{xx}, A_{xy}, A_{yx}, A_{yy})^T, \tag{13}$$

we can expand eq. (12) into the form

$$-2 \ln L(\vec{v}) = \vec{v}^T \mathbf{P} \vec{v} + \vec{v} \cdot \vec{q} + r. \tag{14}$$

Writing  $\mathbf{P}$ ,  $\vec{q}$ , and  $r$  explicitly here would result in a mess, but they are easy to compute in code from the numerical values of the variables in eq. (12). Once they are computed, the value of  $\vec{v}$  that minimizes eq. (14), or equivalently maximizes the likelihood, is simply

$$\vec{v} = -\frac{1}{2} \mathbf{P}^{-1} \vec{q} \tag{15}$$

with a covariance matrix

$$\mathbf{cov}_{\vec{v}} = \mathbf{P}^{-1}. \tag{16}$$

## 4 Alignment results

In this section, we show various plots that illustrate the results of the alignment procedure. We include plots for four microscopes:

- (a) **JHUVectra3\_1**, a Vectra 3.0 microscope at Johns Hopkins Hospital . We plot results for two large images, codenames M1\_1 and M2\_3, and one small image, M21\_1\_, all taken from melanoma patients.
- (b) **JHUPolaris\_2**, a Vectra Polaris microscope at Johns Hopkins Hospital . We plot results for the PZ1 image, which is from a pancreas biopsy.
- (c) **BMSPolaris\_1**, a Vectra Polaris microscope at Bristol-Meyers Squibb. We plot results for the lung image named ML1603480\_BMS078\_5\_22.
- (d) **AKYPolaris\_1**, a Vectra Polaris microscope at Akoya Biosciences. We plot results for the image named TS19\_0181\_A\_1\_3\_BMS\_MITRE. The sample was taken by Bristol-Meyers Squibb (hence the ‘‘BMS’’ in the name) and sent to Akoya to be stained and scanned.

### 4.1 Alignment movements

For an alignment with  $N$  fields, we have  $2N$  coordinates to set. Except for the fields at the edges of an island, each field borders 8 others, on its four edges and four corners, for a total of  $O(8N)$  overlaps. We therefore have more constraints than parameters, and it is natural to ask how consistent those constraints are. There is always uncertainty involved, so the final alignment will never be in perfect agreement with all constraints, but does the agreement significantly improve?

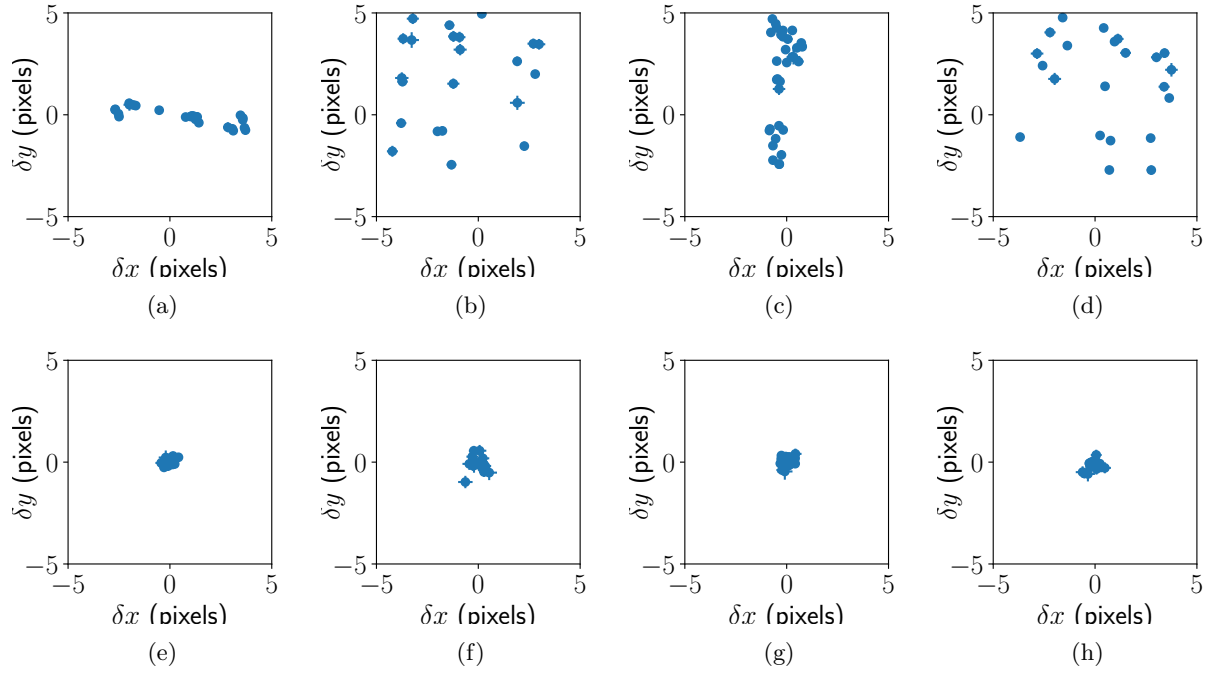


Figure 6: Distributions of  $\delta\vec{r}_o$  for the overlaps in a small image, code name M21\_1, taken with the Vectra 3.0 microscope. The plots show overlaps between one field and the field (a) to its right, (b) under it and to the right, (c) under it, and (d) under it and to the left. Each point with its error bars indicates a single overlap. (e)–(h) The same distributions for the respective overlap directions after stitching has been performed.

#### 4.1.1 JHU Vectra 3.0

Figures 6(a) to 6(d) show the distribution of  $\delta\vec{r}$  for the overlaps, before any stitching has been done. Figures 6(e) to 6(h) show the same distributions after the fields have been moved to their stitched positions, and show significant improvement over the top row.

An interesting feature of the plots is the comparison between figs. 6(a) and 6(c). Fields that are only separated in the  $x$  direction are only misaligned in the  $x$  direction, while fields separated in  $y$  are only misaligned in  $y$ . As illustrated in fig. 7, the microscope scans along a row in the  $x$  direction first, then moves down in the  $y$  direction and goes back to the beginning in  $x$ . This explains fig. 6(a): as the microscope moves in the  $x$  direction, the error introduced by slight movements in the  $y$  direction is small. Figure 6(c) is more surprising, because it indicates that the errors introduced as the microscope moves to the right are almost exactly reversed when it moves back to the left. This indicates that the errors are introduced by some kind of systematic effect, which we will examine further in section 4.4.1. Our alignment procedure is able to correct for this error as well as for the smaller, apparently random error associated with each individual movement.

Figure 8 shows the same distributions for a larger image taken with the same microscope. It shows similar effects, though they are harder to see due to the larger number of overlaps (4527 instead of 109). From these plots, it appears that although most fields are aligned well, there are a few outliers. Furthermore, most of the outliers are corner overlaps (figs. 8(f) and 8(h)) and are misaligned in a diagonal direction. Because there are no such effects in the edge overlaps (figs. 8(e) and 8(g)), the most likely explanation for the discrepancy between the individual alignment and the stitched alignment is that those particular corner overlaps are aligned badly. The error estimation procedure correctly assigns a large error to them, and so they are weighted down in the final fit.

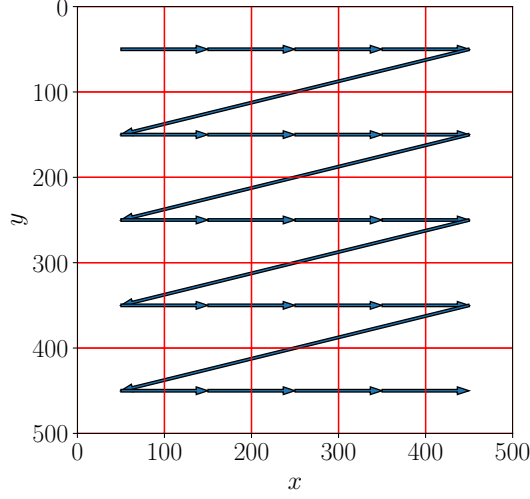


Figure 7: Illustration of how the microscope scans an image. It scans the first row in the  $x$  direction, then moves on to the second, then the third, and so on.

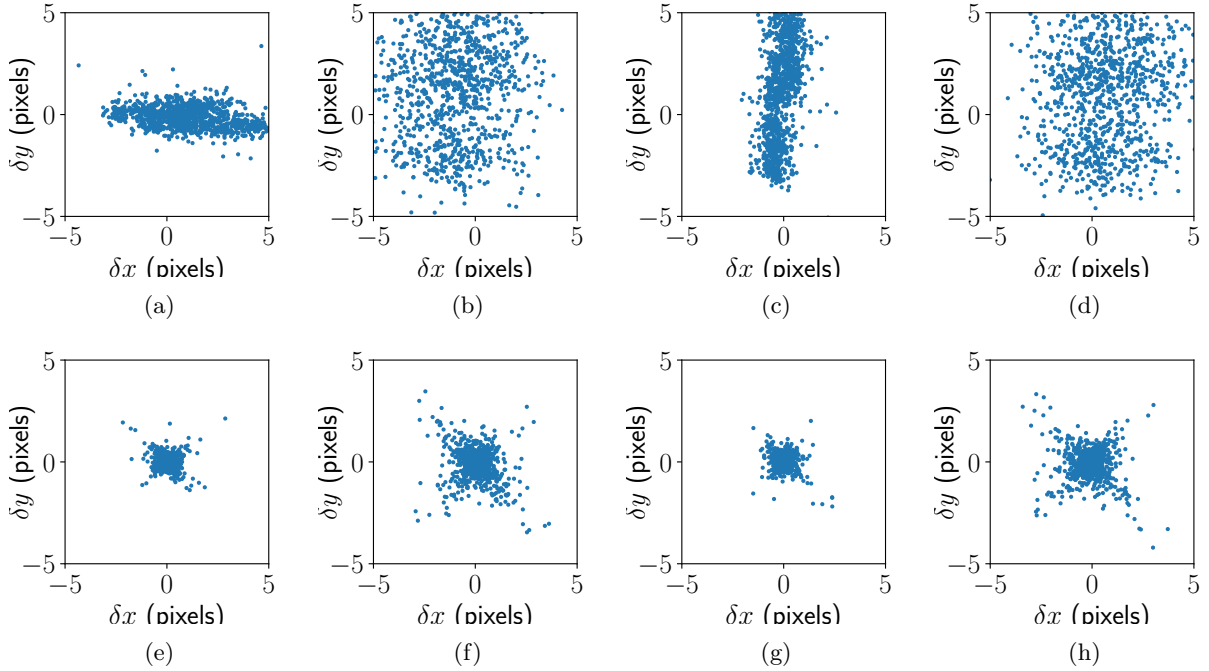


Figure 8: Distributions of  $\delta\vec{r}_o$  for the overlaps in a larger image, code name M1\_1, taken with the Vectra 3.0 microscope. The plots show overlaps between one field and the field (a) to its right, (b) under it and to the right, (c) under it, and (d) under it and to the left. Each point indicates a single overlap. Error bars are omitted because of the larger number of overlaps in this image. (e)–(h) The same distributions for the respective overlap directions after stitching has been performed.



#### 4.1.2 JHU Vectra Polaris

Figure 9 shows the alignment movements from the PZ1 sample.

The effects are smaller than the ones seen in the Vectra 3.0 microscope shown in figs. 6 and 8 and are corrected by the stitching, as shown in the bottom row. Some apparently badly-aligned overlaps form an X shape in the bottom row of plots, but as in the case of the fig. 8, these are mostly limited to corner overlaps and seem not to have a large effect on the overall alignment.

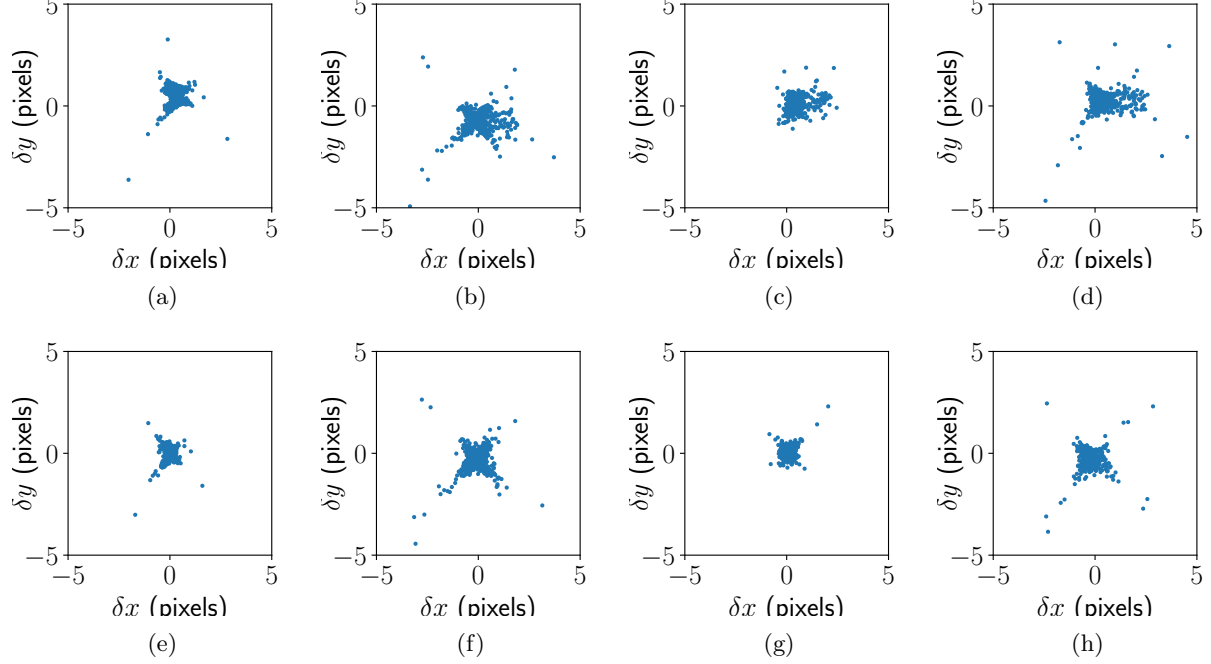


Figure 9: Distributions of  $\delta\vec{r}_o$  for the overlaps in an image, code name PZ1, taken with the JHU Vectra Polaris microscope. The plots show overlaps between one field and the field (a) to its right, (b) under it and to the right, (c) under it, and (d) under it and to the left. Each point indicates a single overlap. Error bars are omitted because of the larger number of overlaps in this image. (e)–(h) The same distributions for the respective overlap directions after stitching has been performed.

#### 4.1.3 BMS Vectra Polaris

Figure 10 shows the alignment movements from the ML1603480\_BMS078\_5\_22 sample. The stitching gives a tighter distribution in the bottom row of plots than in the top row.

#### 4.1.4 Akoya Vectra Polaris

Figure 11 shows the alignment movements from the TS19\_0181\_A\_1\_3\_BMS\_MITRE sample.

All four of the plots in the top row show two clusters of overlaps, indicating that there are systematic effects present in addition to the random movement. Section 4.4.4 will investigate those effects in greater detail.

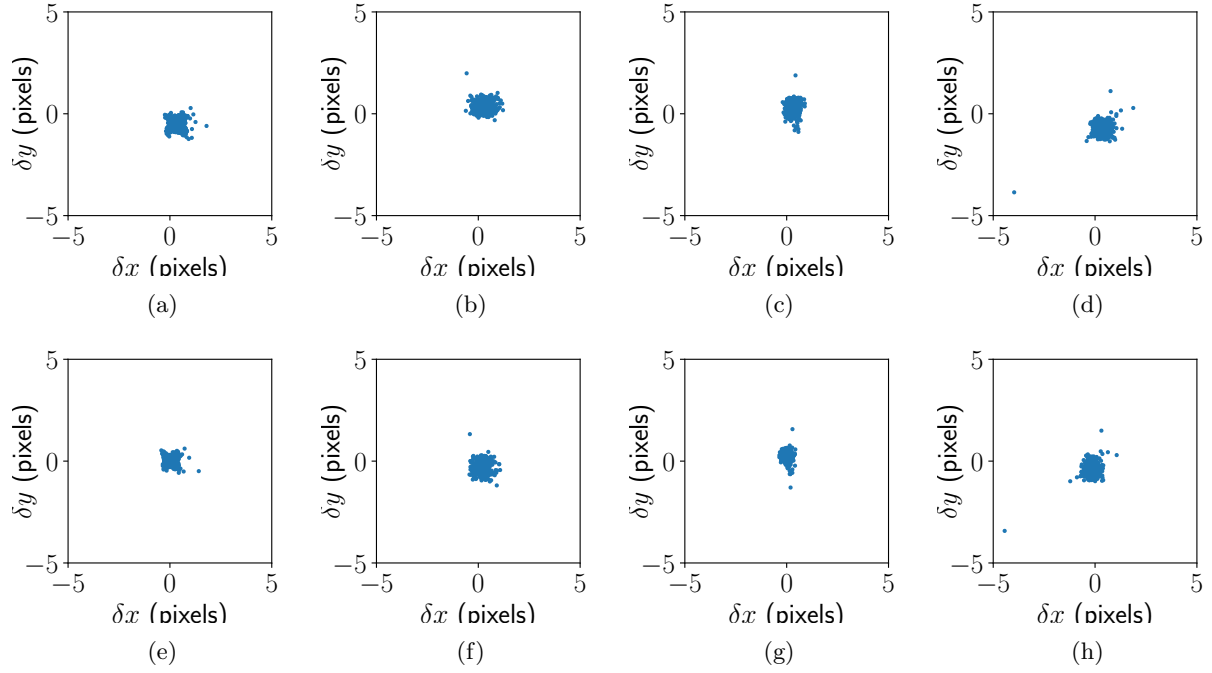


Figure 10: Distributions of  $\delta \vec{r}_o$  for the overlaps in an image, code name TS19\_0181\_A\_1\_3\_BMS\_MITRE, taken with the BMS Vectra Polaris microscope. The plots show overlaps between one field and the field (a) to its right, (b) under it and to the right, (c) under it, and (d) under it and to the left. Each point indicates a single overlap. Error bars are omitted because of the larger number of overlaps in this image. (e)–(h) The same distributions for the respective overlap directions after stitching has been performed.

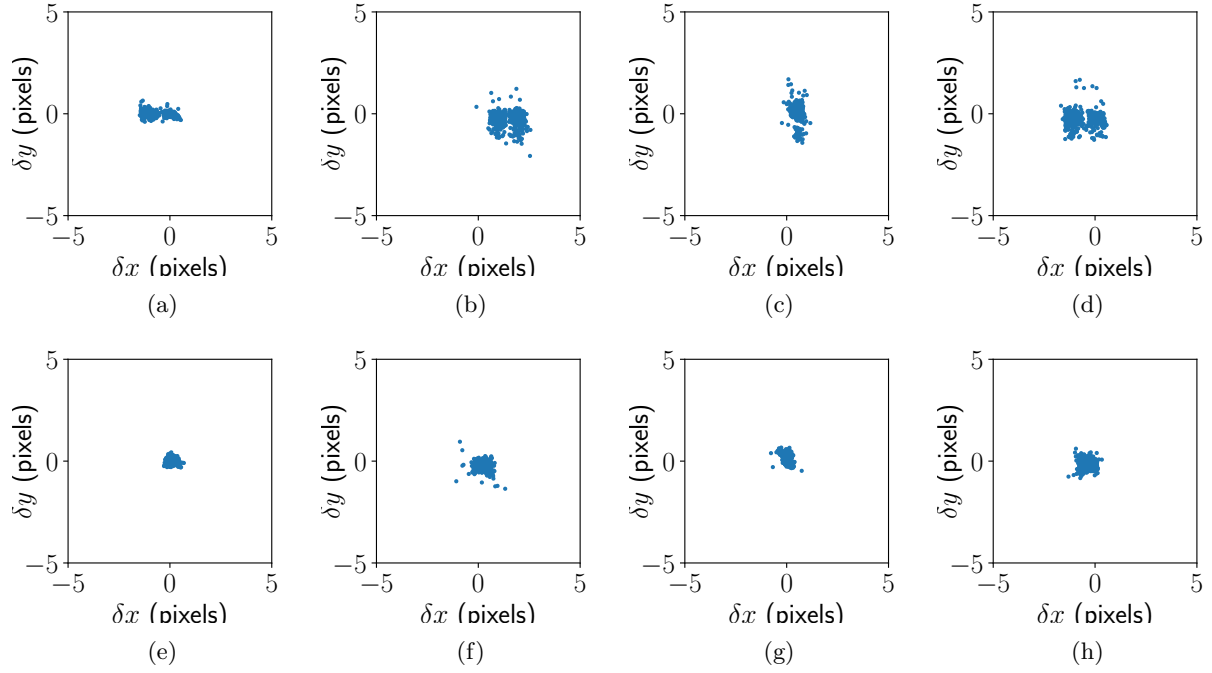


Figure 11: Distributions of  $\delta \vec{r}_o$  for the overlaps in an image, code name TS19\_0181\_A\_1\_3\_BMS\_MITRE, taken with Akoya's Vectra Polaris microscope. The plots show overlaps between one field and the field (a) to its right, (b) under it and to the right, (c) under it, and (d) under it and to the left. Each point indicates a single overlap. Error bars are omitted because of the larger number of overlaps in this image. (e)–(h) The same distributions for the respective overlap directions after stitching has been performed.

## 4.2 Consistency of the errors

We also want to ensure that our estimate of the errors on the alignment are correct. To do this, we use what is known as a “pull” distribution. The idea is to find a quantity  $x$  that we expect to be 0. Any actual measurement of  $x$  will not necessarily be 0, but we perform multiple measurements  $x_i$ , we expect to end up with a normal distribution centered at 0 and with a width of  $\sigma$ . If we instead look at the distribution of  $x_i/\sigma$ , the width of the normal distribution will be 1.

In general, we do not know  $\sigma$  in advance, but we have an estimate  $\sigma_i$  of the error on each individual measurement. The idea of a pull distribution is that we plot  $x_i/\sigma_i$  and look at the width of the resulting distribution. If it is close to 1, that indicates that our error estimate was correct.

### 4.2.1 JHU Vectra 3.0

Figure 12 shows the pull distributions for the total displacement from pairwise alignments for moving around in a square or diamond. For a perfect alignment, this quantity would be 0, so it is a good candidate for a pull distribution. As the figures show, the standard deviations of all pull distributions are close to 1, indicating that our error estimates for both edge and corner overlaps are reasonable.

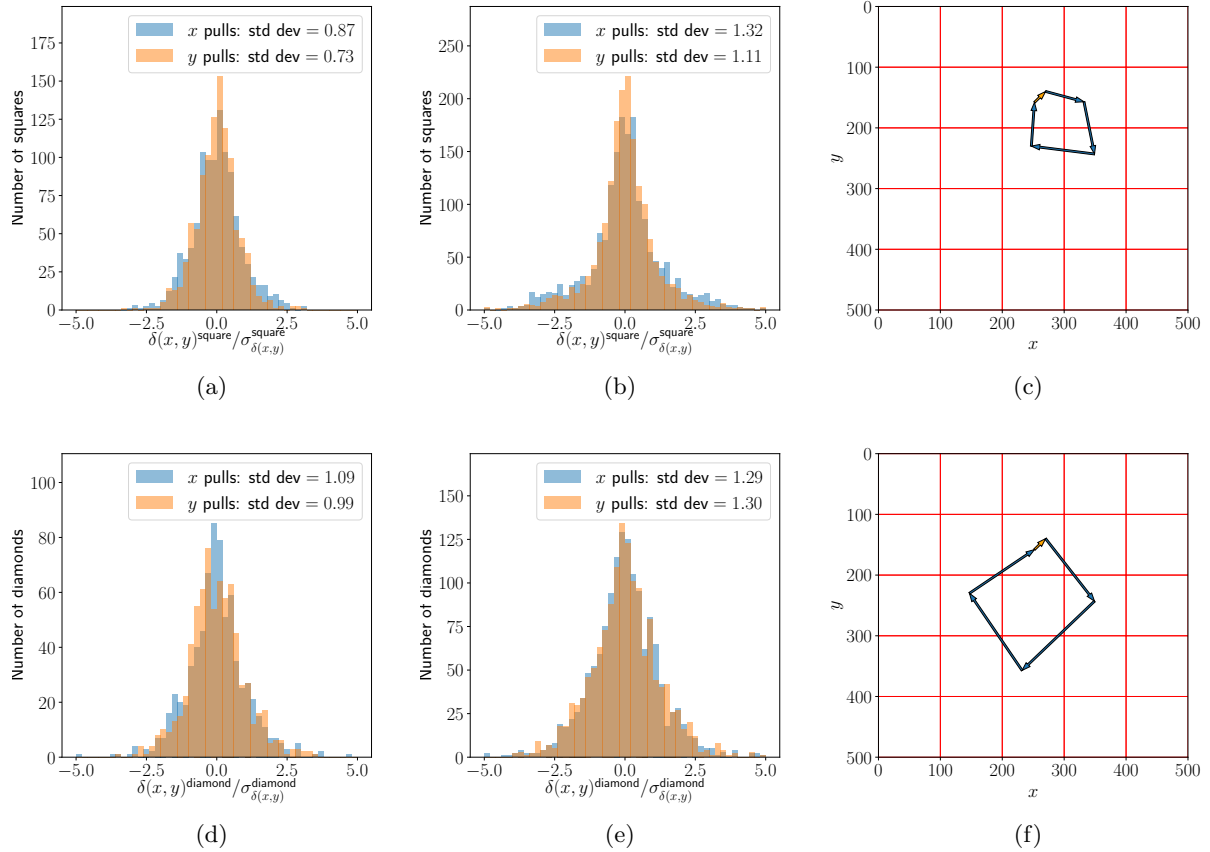


Figure 12: (a), (b) The pull distributions in  $x$  and  $y$  for the total displacement obtained from pairwise alignment when moving around in a square, as shown in (c), for two different images, code names M1.1 and M2.3, respectively, collected with the Vectra 3.0 microscope. (d), (e) The corresponding pull distributions for those two datasets when moving in a diamond, as shown in (f).

Figure 13 shows a different metric for our error estimation. In these plots, we calculate the misalignment for each overlapping pair of fields after stitching, as in figs. 6(e) to 6(h), and create pull distributions for the  $x$  and  $y$  residuals. Again, the standard deviations of the pull distributions are close to 1.

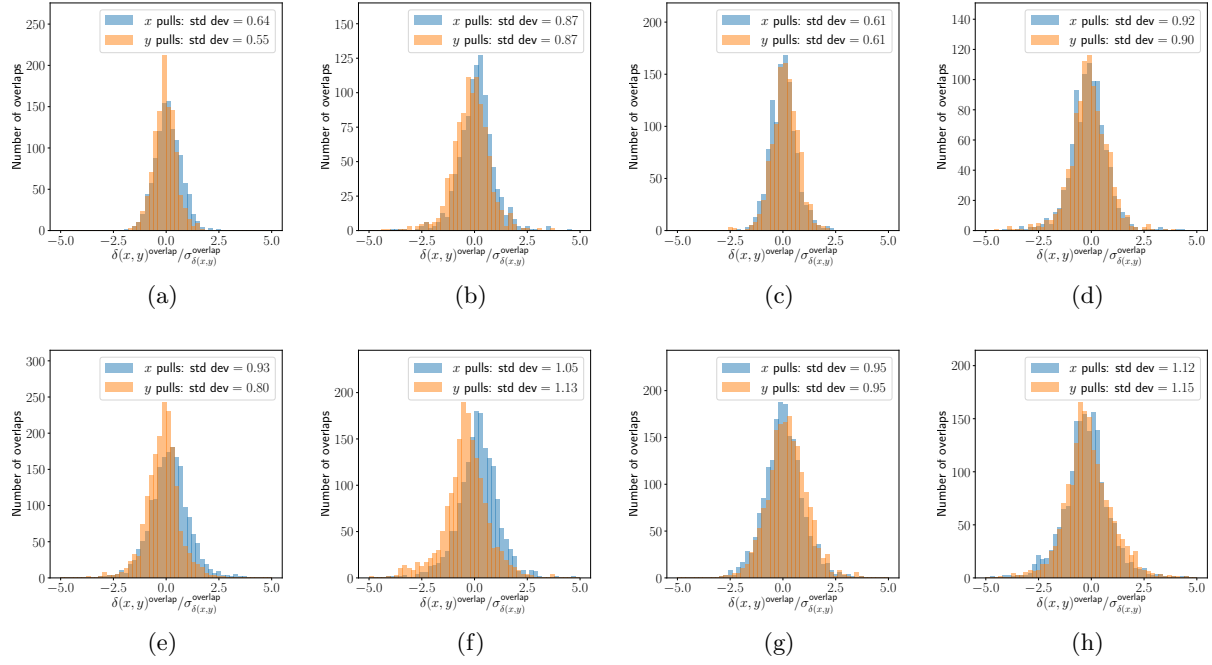


Figure 13: Pull distributions in  $x$  and  $y$  for the residual pairwise misalignment after stitching, as shown in figs. 6(e) to 6(h), for two different images, M1\_1 and M2\_3, collected with the Vectra 3.0 microscope. The plots show distributions for overlaps between one field and the field (a), (e) to its right, (b), (f) under it and to the right, (c), (g) under it, and (d), (h) under it and to the left.

The standard deviations of the pull distributions were used to determine the factor of 1/16 used to correct eq. (8) to eq. (9). The fact that, after making this correction, the standard deviations are 1 for multiple different images as well as for pull distributions of two different quantities lends confidence to the *method* of error estimation. Remaining deviations from 1 indicate that there is still some amount of room for improvement.

### 4.3 Linear transformation matrix

As described in section 3, when we stitch the alignment we also obtain the linear matrix  $\mathbf{A}$ , which describes the relationship between the distance that the microscope is supposed to move  $\Delta\vec{r}$  and the average distance it actually moves  $\mathbf{A}\Delta\vec{r}$ . We expect it to be more or less constant for different images taken with the same microscope, although of course the mechanical parts can also degrade over time, change due to maintenance, or be replaced.

The primary purpose of the alignment and stitching is to determine the relative locations of the high-powered fields, and when only a single island is present  $\mathbf{A}$  is not necessary for this purpose. When there are multiple islands, as in fig. 5, including  $\mathbf{A}$  in the fit allows us to obtain those islands' approximate relative locations. In either case, we do not need the precise values of  $\mathbf{A}$ 's entries for any analysis, but the matrix still provides interesting information about the mechanical calibration of the microscope.

#### 4.3.1 JHU Vectra 3.0

For the M1\_1 dataset, we find

$$\mathbf{A} = \begin{pmatrix} 0.999\,132(9) & 0.000\,173(13) \\ 0.000\,103(10) & 0.998\,652(14) \end{pmatrix} \quad (17)$$

$$= \mathbf{I} + \begin{pmatrix} -0.868(9) & 0.173(13) \\ 0.103(10) & -1.348(14) \end{pmatrix} \times 10^{-3} \quad (18)$$

The Vectra 3.0 scans fields of size  $1344 \times 1004$  pixels<sup>2</sup>. The entries of  $\mathbf{A}$  tell us that on average, when moving one field to the right, the microscope ends up 0.9 pixels to the left and 0.1 pixels below where it should, and when moving down, it ends up 0.1 pixels to the right and 1.1 pixels above where it should. Although the top row of fig. 6 shows that the random scatter is a larger effect, the average shift is noticeable.

#### 4.3.2 JHU Vectra Polaris

From the sample named PZ1, we find the linear transformation matrix to be

$$\mathbf{A} = \begin{pmatrix} 0.999\,782(4) & -0.000\,308(5) \\ -0.000\,327(4) & 0.999\,994(5) \end{pmatrix} \quad (19)$$

$$= \mathbf{I} + \begin{pmatrix} -0.218(4) & -0.308(5) \\ -0.327(4) & -0.006(5) \end{pmatrix} \times 10^{-3} \quad (20)$$

The Vectra Polaris scans fields of size  $1876 \times 1404$  pixels<sup>2</sup>, slightly larger than the Vectra 3.0. The result is that when moving one field to the right, the microscope undershoots by an average of 0.3 pixels and moves up by 0.5 pixels. When moving down, it moves to the right by an average of 0.3 pixels.

#### 4.3.3 BMS Vectra Polaris

From the sample named ML1603480\_BMS078.5\_22, we find the linear transformation matrix to be

$$\mathbf{A} = \begin{pmatrix} 0.999\,829\,0(35) & -0.000\,253(4) \\ 0.000\,364(4) & 0.999\,904(5) \end{pmatrix} \quad (21)$$

$$= \mathbf{I} + \begin{pmatrix} -0.1710(35) & -0.253(4) \\ 0.364(4) & -0.096(5) \end{pmatrix} \times 10^{-3} \quad (22)$$

When moving one field to the right, the microscope undershoots by an average of 0.2 pixels and also moves down by 0.5 pixels. When moving down, it undershoots by an average of 0.1 pixels and moves to the left by 0.3 pixels.

#### 4.3.4 Akoya Vectra Polaris

From the sample named TS19\_0181\_A\_1\_3\_BMS\_MITRE, we find the linear transformation matrix to be

$$\mathbf{A} = \begin{pmatrix} 1.000\,503(5) & -0.000\,454(5) \\ 2.44(31) \times 10^{-5} & 1.000\,044(4) \end{pmatrix} \quad (23)$$

$$= \mathbf{I} + \begin{pmatrix} 0.503(5) & -0.454(5) \\ 0.0244(31) & 0.044(4) \end{pmatrix} \times 10^{-3} \quad (24)$$

When moving one field to the right, the microscope overshoots by an average of 0.7 pixels, as expected from fig. 11(a), and when moving down, it ends up 0.5 pixels to the left of the desired location, as expected from fig. 11(c).

## 4.4 Two-dimensional dependence of shifts

Finally, we can plot the movements of each field in the image. These plots give a full picture of the alignment in all of its detail, and will shed more light on the features observed in the previous sections.

In these plots, we show  $\vec{r}_k - \mathbf{A}\vec{r}_k^n$ , subtracting the effect of the affine matrix  $\mathbf{A}$ . In this way, we measure the random shifts as well as systematic effects not described by the affine matrix. Each field is assigned a color in  $x$  and  $y$  based on the size of the alignment shift.

For each individual microscope, the color scales for  $x$  and  $y$  are set to be the same. However, we do not set those scales to be the same between all microscopes because the sizes of the relevant shifts are different for each one, as will be immediately clear from the plots.

### 4.4.1 JHU Vectra 3.0

Figure 14 shows the two-dimensional distribution of the shifts for the M1\_1 dataset. This plot highlights periodic features, with an amplitude of several pixels, in the 2D distribution.

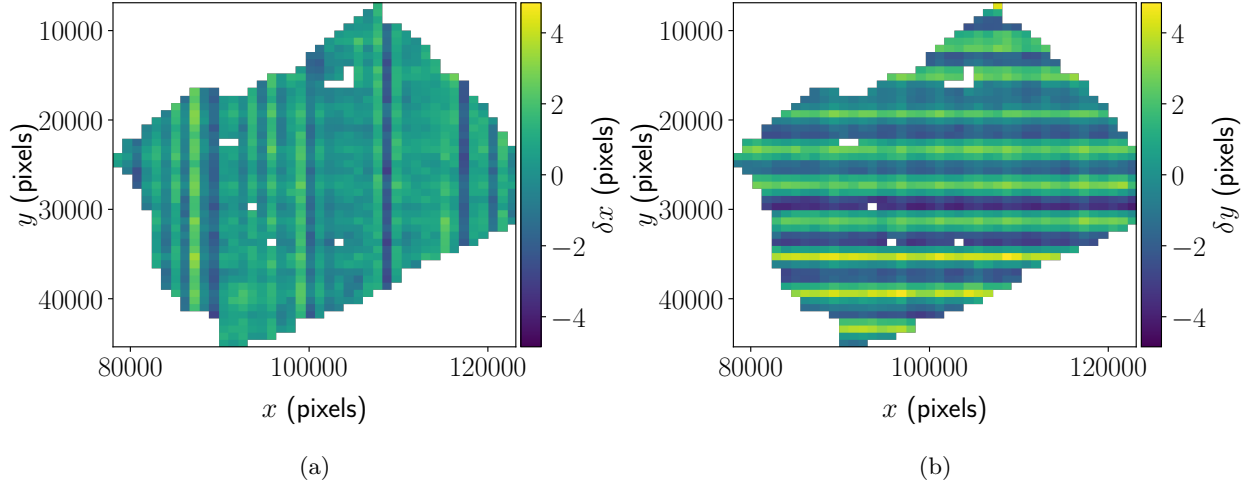


Figure 14: The shifts in (a)  $x$  and (b)  $y$  for the fields in M1\_1.

To analyze the sinusoidal pattern more quantitatively, we can project onto the  $x$  and  $y$  axes and make a profile plot. Each point in fig. 15(a) represents a column in fig. 14(a): the marker gives the average shift for that column and the error bar gives the standard deviation. Similarly, each point and associated error bar of fig. 15(b) show the average and standard deviation of the shifts in the corresponding column of fig. 14(b). The points are fitted to a sine wave.

As a cross check, similar plots are also shown for a second dataset in figs. 15(c) and 15(d), and the amplitudes and wavelengths of the sine waves are in agreement between both datasets. This indicates that the effect is probably a property of the microscope. For instance, the cam wheels controlling the microscope movement may have a slight eccentricity instead of being exactly circular.

### 4.4.2 JHU Vectra Polaris

Figure 16 shows the two-dimensional distribution of the shifts for the PZ1 dataset. This plot shows that at the beginning of the scan, the microscope took some time to gradually “settle in” to its final position. The alignment procedure accounts for this settling.

### 4.4.3 BMS Vectra Polaris

Figure 17 shows the two-dimensional distribution of the shifts for the ML1603480\_BMS078.5.22 dataset. This dataset is very stable and only contains some seemingly random noise, all less than one pixel.



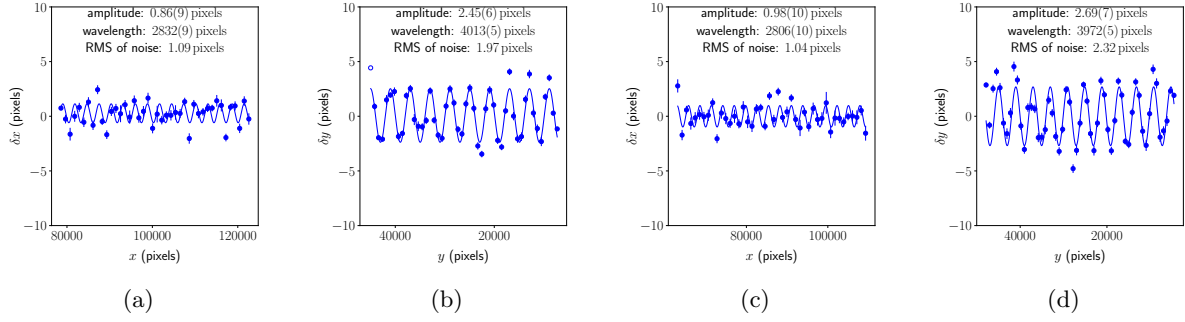


Figure 15: Profile plot of the measured alignment shifts for overlaps in a particular row or column. Each data point shows the average shift for the row or column, and the error bars show the standard deviation. (a) shows shifts in the  $x$  direction as a function of  $x$  and (b) shows shifts in the  $y$  direction as a function of  $y$ , both for the M1\_1 dataset. (c) and (d) show the same for the M2\_3 dataset. The plots are fitted to a sine curve, whose amplitude and wavelength are indicated on the plot.

Figure 18 shows the two-dimensional distribution of the shifts for the ML1603474\_BMS069\_5\_21 dataset, a different image taking with the same microscope. The most striking feature of this plot is that the microscope apparently jumped to the left by 10 pixels in  $x$  for several rows in the middle of the scan, then jumped again by 30 pixels to the right, then moved back closer to its original position. This kind of shift most likely only happens occasionally, but can severely degrade data quality if left uncorrected.

#### 4.4.4 Akoya Vectra Polaris results

Figure 19 shows the two-dimensional distribution of the shifts for the TS19\_0181\_A\_1\_3\_BMS\_MITRE dataset. This dataset is quite stable. It shows a 1-pixel settling effect at the beginning of the scan and periodic modulations with amplitudes of less than 1 pixel.

Figure 20 shows projections of these plots onto the  $x$  and  $y$  axes, which help to more quantitatively identify the periodic pattern.

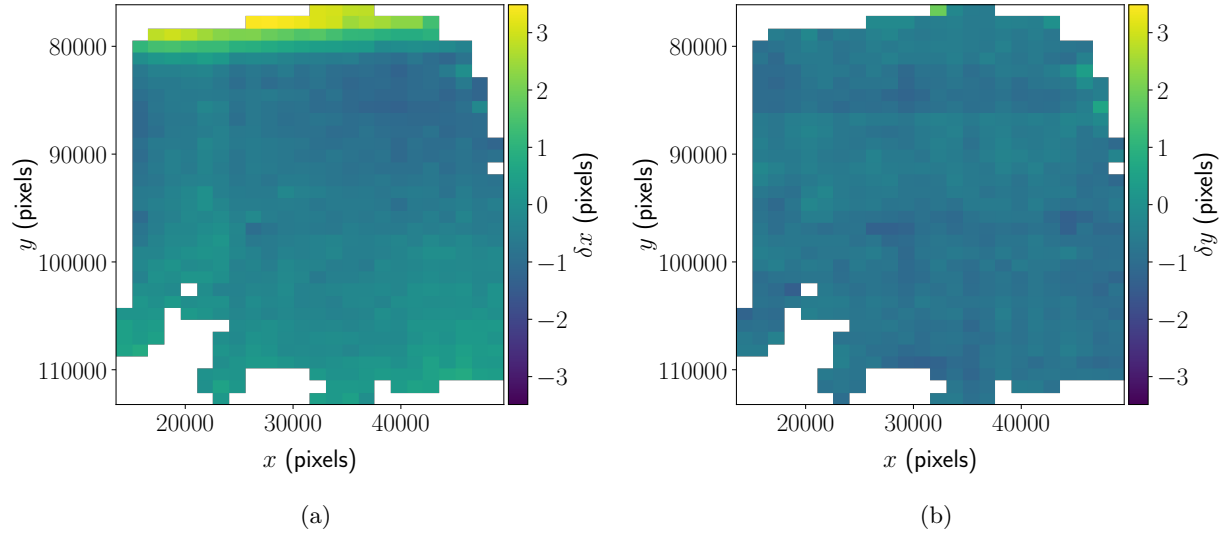


Figure 16: The shifts in (a)  $x$  and (b)  $y$  for the fields in PZ1.

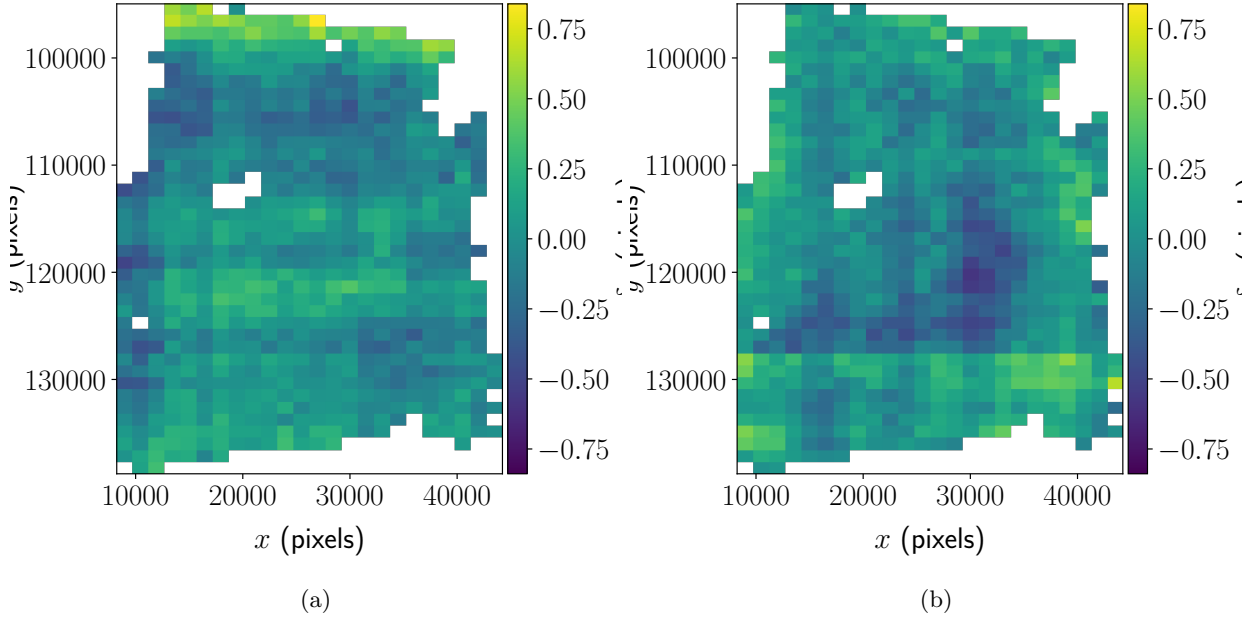


Figure 17: The shifts in (a)  $x$  and (b)  $y$  for the fields in ML1603480\_BMS078\_5\_22.

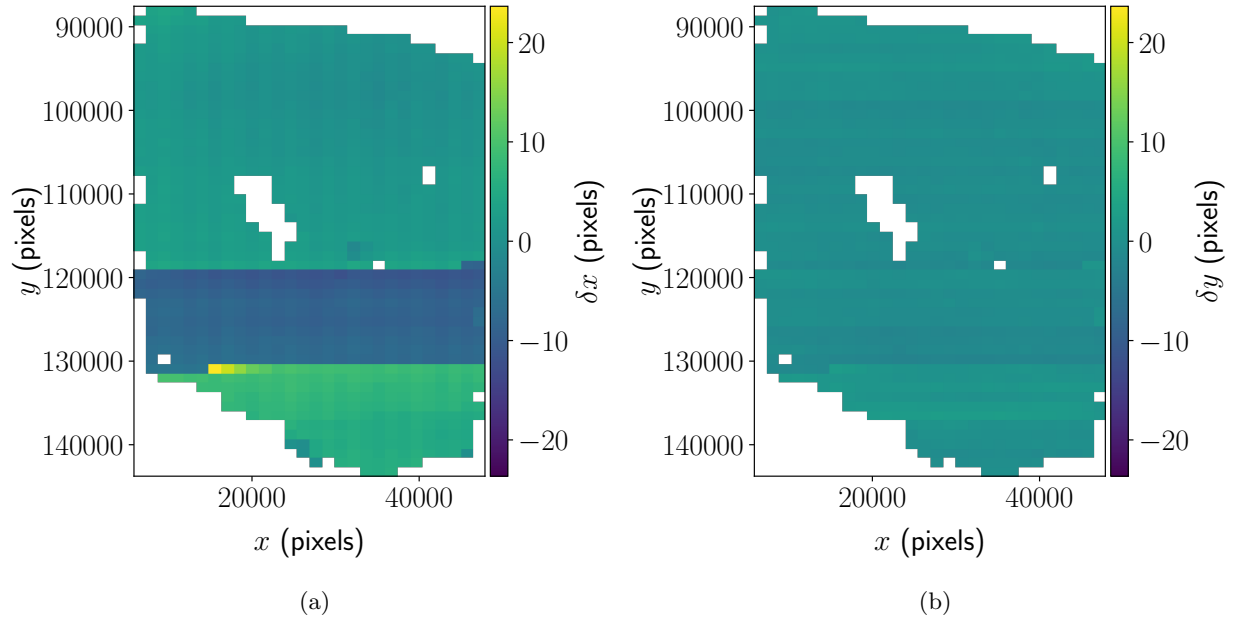


Figure 18: The shifts in (a)  $x$  and (b)  $y$  for the fields in ML1603474\_BMS069\_5\_21.

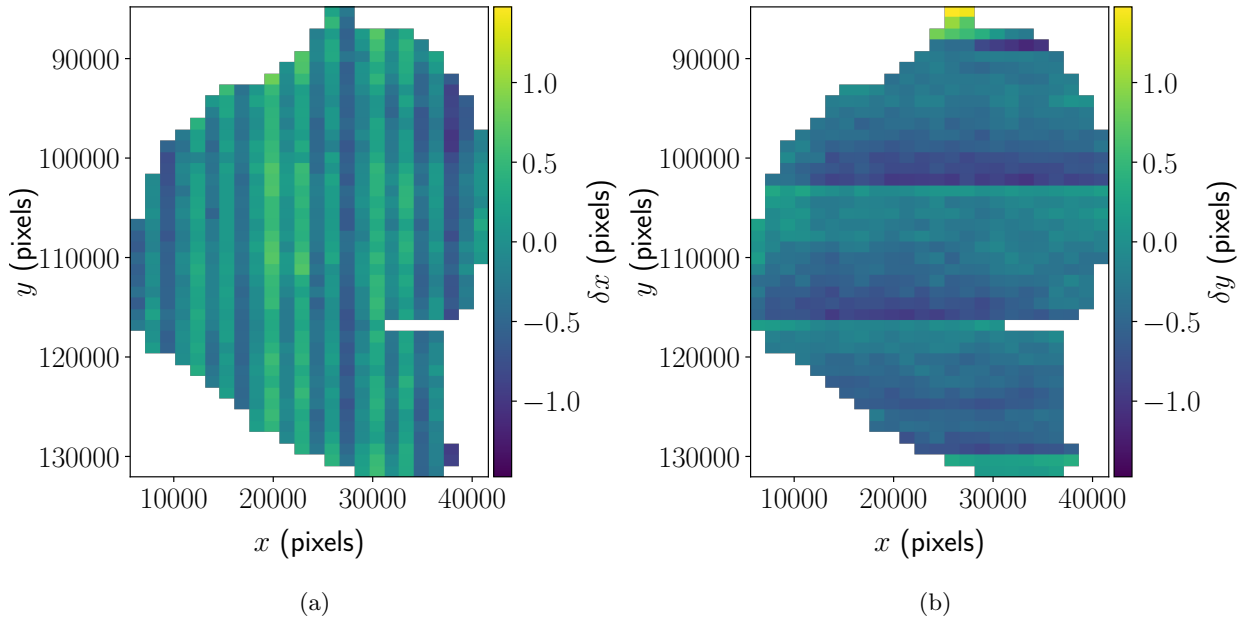
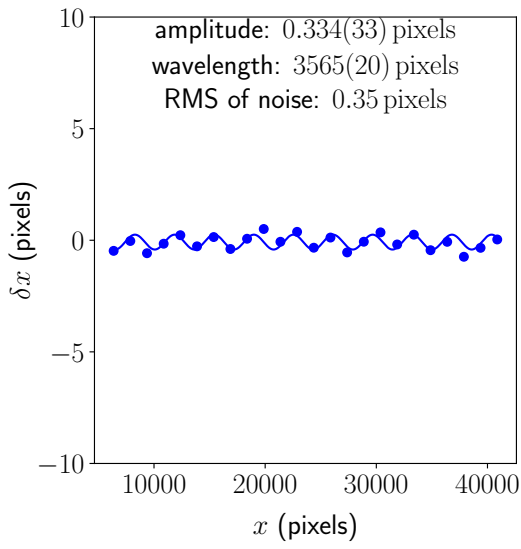
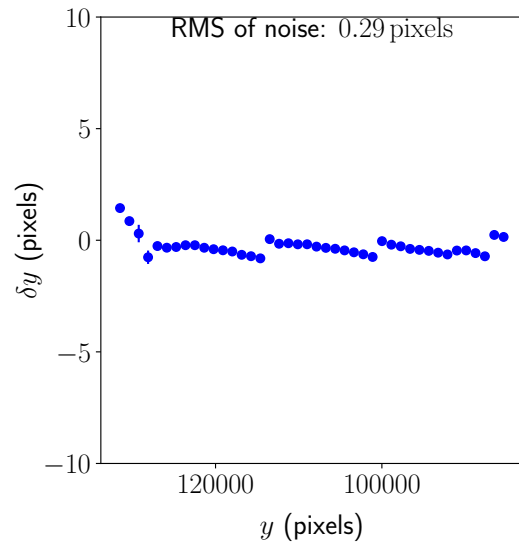


Figure 19: The shifts in (a)  $x$  and (b)  $y$  for the fields in TS19\_0181\_A\_1.3\_BMS\_MITRE.



(a)



(b)

Figure 20: Profile plots of the measured alignment shifts for overlaps in a particular row or column. Each data point shows the average shift for the row or column, and the error bars show the standard deviation. (a) shows shifts in the  $x$  direction as a function of  $x$  and (b) shows in the  $y$  direction as a function of  $y$ , both for the TS19\_0181\_A\_1\_3\_BMS MITRE dataset. (a) is fitted to a sine curve, whose amplitude and wavelength are indicated on the plot.

## 4.5 Isotropy of residual misalignments

As figs. 8(e) to 8(h) show, the residual misalignments after stitching are not isotropic. To some extent, this is to be expected, because the input to the alignment, the edges and corners, are not isotropic either. Nevertheless, the lack of isotropy might indicate an imperfection in previous steps of the calibration (such as flatfielding or warping) or may come from an incorrect relative weight assigned to the edge and corner overlaps in the stitching procedure. We can note that the pull distributions in section 4.2 are generally wider for the corner overlaps than for the edge overlaps, lending some credence to this second possibility.

### 4.5.1 JHU Vectra 3.0

In fig. 21, we attempt to quantify this anisotropy by producing a histogram of the angle  $\phi$  of the overlaps in figs. 8(e) to 8(h). Each histogram is fitted to a Fourier series, and the first four terms are shown on the plot. An isotropic distribution would produce flat histograms.

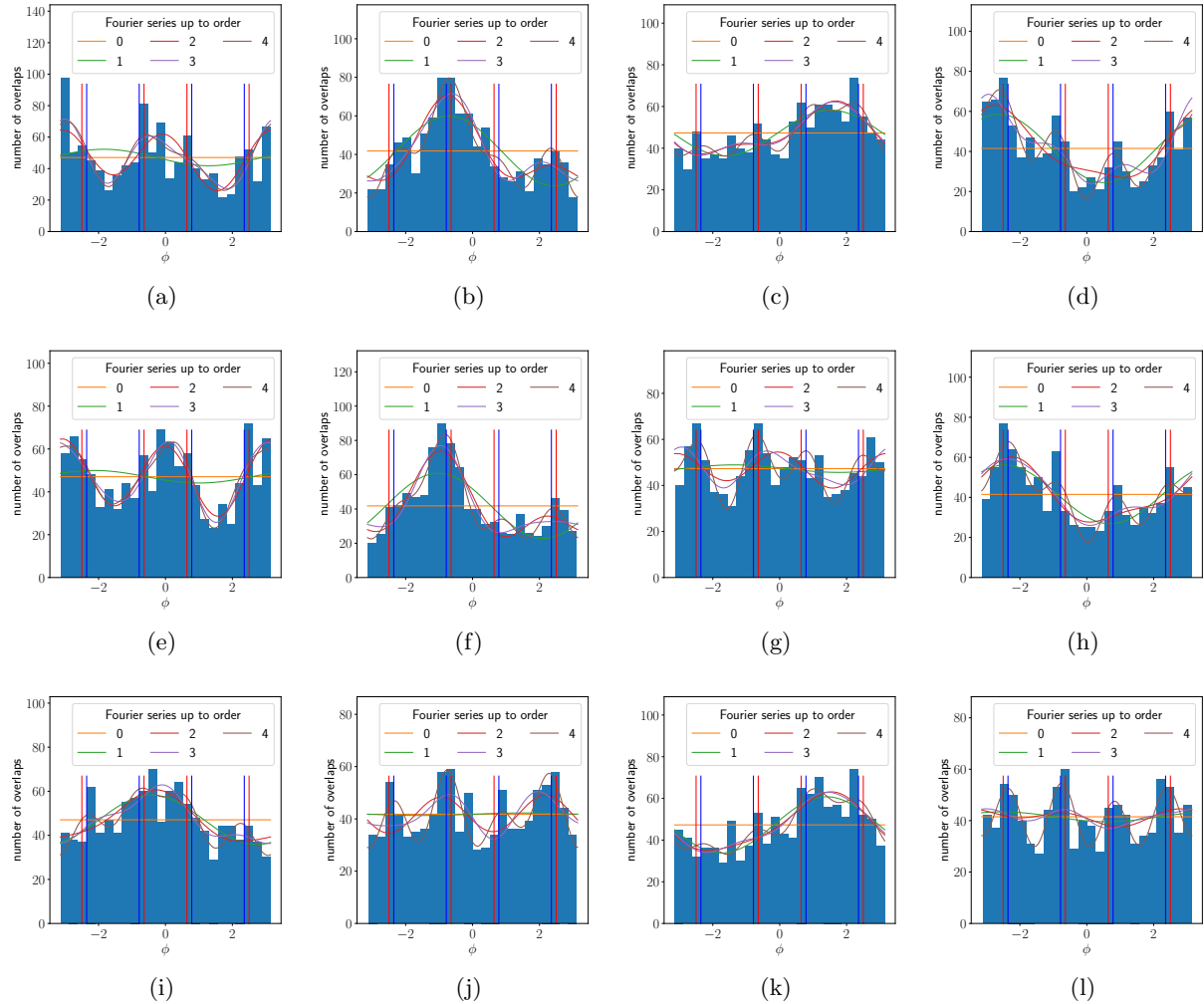


Figure 21: Distributions of the angles  $\phi$  of the residual misalignments, after stitching, for the overlaps in the M1.1 dataset, taken with the Vectra 3.0 microscope. The plots show overlaps between one field and the field (a) to its right, (b) under it and to the right, (c) under it, and (d) under it and to the left. The histograms count the number of overlaps, and the curves show the first few terms in the Fourier series. (e)–(h) The same distributions when stitching using only edge overlaps. (i)–(l) The same distributions when stitching using only corner overlaps.

## 5 Summary

In this note, we describe the alignment and stitching of a microscope image composed of up to 1000 fields. Our procedure includes an estimation of the error on each field, which are propagated throughout the procedure to obtain a final error estimate and covariance matrix for the positions of the fields. To demonstrate the method, we run it over several data samples collected by different microscopes. Table 1 compares the alignment behavior of the different microscopes used. The method provided here, which runs in a few minutes, forms an important part of the data calibration in the astropath framework.

Model		Vectra 3.0	Polaris	Polaris	Polaris
Location		JHU	JHU	BMS	Akoya
Average shift/field (pixels)	$\delta x$ , $x$ overlap	−0.9	−0.3	−0.2	0.7
	$\delta y$ , $x$ overlap	0.1	−0.5	0.5	0.0
	$\delta x$ , $y$ overlap	0.1	0.3	−0.3	−0.5
	$\delta y$ , $y$ overlap	−1.1	0.0	−0.1	0.0
RMS of overlap alignment results (pixels)	$\delta x$ , initial positions	1.7	0.6	0.4	1.0
	$\delta y$ , initial positions	2.3	0.6	0.6	0.4
	$\delta x$ , after affine shift	1.6	0.5	0.3	0.6
	$\delta y$ , after affine shift	2.1	0.5	0.4	0.4
	$\delta x$ , after stitching	0.8	0.4	0.3	0.3
	$\delta y$ , after stitching	0.8	0.5	0.4	0.3
RMS of field shifts (pixels)	$\delta x$	1.1	0.8	0.2	0.4
	$\delta y$	2.0	0.2	0.2	0.3

Table 1: Comparison between the microscopes studied here.