

Lecture 6:

The Continuous-Factor Model

To accompany <https://youtu.be/cgAN5oK7AP8>

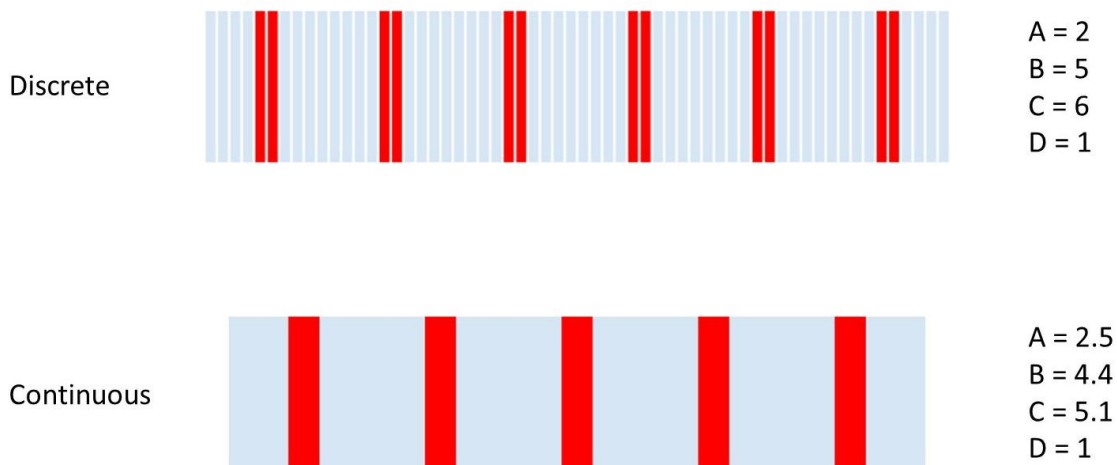
v1, April 13, 2022

Paul Mirsky

paulmirsky633@gmail.com

1 Introduction

1.1 Discrete and continuous factors



Welcome to Lectures on Symmetry Optics. I'm Paul Mirsky. This is Lecture 6 of the series, and the topic is: The Continuous-Factor Model.

Up to now in this lecture series, we have mostly used *discrete factors*. In other words, the factor sizes are whole numbers, and the patterns are made up of a whole number of discrete patches. This is because a factor is represented mathematically by a vector, and a vector can only have a whole number of dimensions. In this lecture, we'll introduce *continuous factors*. These can have any positive real number as a size.

1.2 Contrasting discrete vs continuous factors

| Discrete factors | Continuous factors |
|-----------------------------------|------------------------------|
| Whole-number sizes only | Any (positive, real) number |
| Source-target grid | Quasi-Boolean rules |
| Visualize a point-spread function | No mental picture |
| Easier to beginner to learn | Easier for expert to use |
| Coincidences | No coincidences; trends only |

This is more than just a minor tweak; it's a fundamental change to the algorithm. Let's contrast the two models.

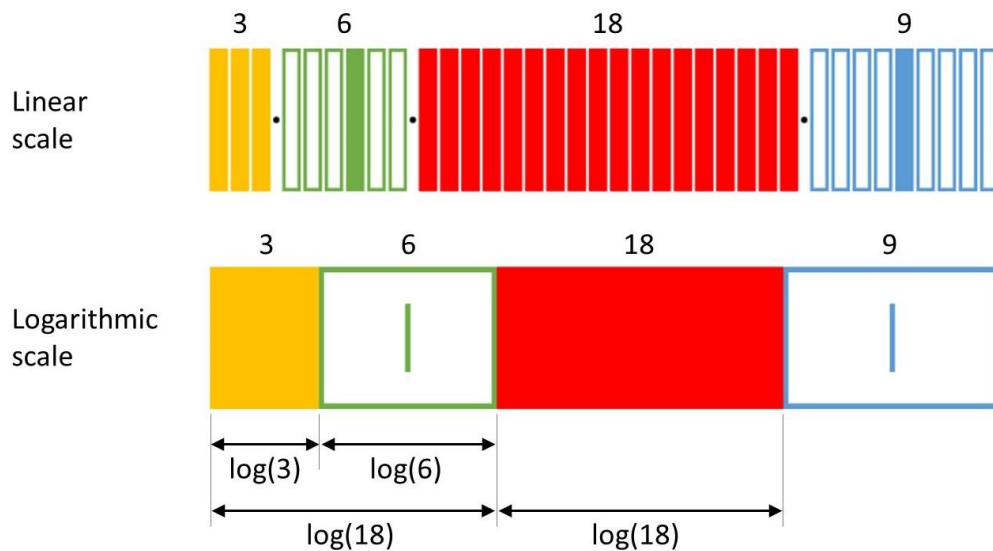
Essentially, we will no longer tabulate a source-target grid. The grid is essential to the discrete model, but in the continuous model we use quasi-Boolean rules instead.

Without the source-target grid, we lose the ability to visualize what's happening. With the grid, we could visualize the live as a point-spread function, and we could imagine the calculation is something analogous to Huyghens' principle. But in the continuous-factor model, there is no mental picture to visualize. It's just a mathematical formula that you need to apply.

Because of that mental picture, I think that the discrete model is easier for a beginner to learn. But once you've already mastered both models, the continuous one is actually easier to use. Also, it's a much more elegant and powerful way to think about it. It's conceptually cleaner. I made the pedagogical choice to teach discrete factors first, and I hope it was the right choice, but regardless – you probably need to know both models in order to really understand either of them well.

Finally, discrete factors are necessary for calculating coincidences. The continuous-factor model is not able to compute coincidences, only overall trends.

1.3 Logarithmic vs linear scales



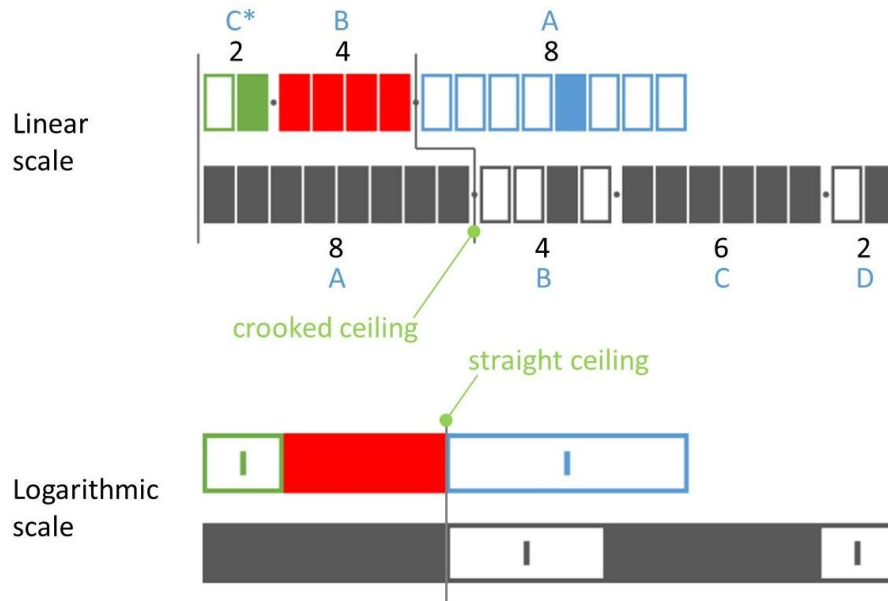
To accompany the new model, we'll also modify the factor diagrams. The older diagrams looked like the one on top; the new ones look like the one on bottom. Strictly speaking, the type of diagram we draw is independent of whether we use discrete or continuous factors. But for a variety of reasons, it's more convenient to use this new type of diagram along with continuous factors.

We're actually making two different changes: Firstly, we'll draw each factor as a single large rectangle, rather than as a series of smaller rectangles. Each large rectangle can be either open or filled, to represent a singular or plenary factor. This is partly just a cosmetic change.

The more important change is that the length of each factor will be drawn with a logarithmic scale, rather than a linear scale. So, for example, in the older diagram, a size-3 factor next to a size-6 factor takes up about as much room as a size-9 factor, because 3 *plus* 6 is 9. And the discrepancy is just due to these dots and spaces between factors. But on a logarithmic scale, the size-3 and size-6 factors take up as much room as a size-18 factor, because 3 *times* 6 is 18.

A logarithmic scale can be counterintuitive. For example, if there were a size-1 factor, it wouldn't take up any space at all. And you can't think of this size-18 factor as being broken into 18 identical pieces, the way this one is – if you did break it up, the pieces would be larger on the left, and smaller on the right. That's inherent in how a logarithmic scale works.

1.4 Crooked vs straight ceilings and floors



The older style diagrams are a little bit like ‘training wheels’ because they contain built-in reminders to help you visualize the patterns.

But, the logarithmic scale actually better reflects the logarithmic logic of symmetry optics. For instance, on a linear scale, a shared ceiling may appear crooked. Here, the live factors are size 2 and 4, so live B’s ceiling is at 8. Static A is also size 8, so live B shares a ceiling with static A. But that line is drawn crooked, because the length of each factor is drawn on a linear scale. If we look at the same thing drawn on a logarithmic scale, the line appears perfectly straight. Here, horizontal position is a true representation of size.

2 Continuous-Factor Beam

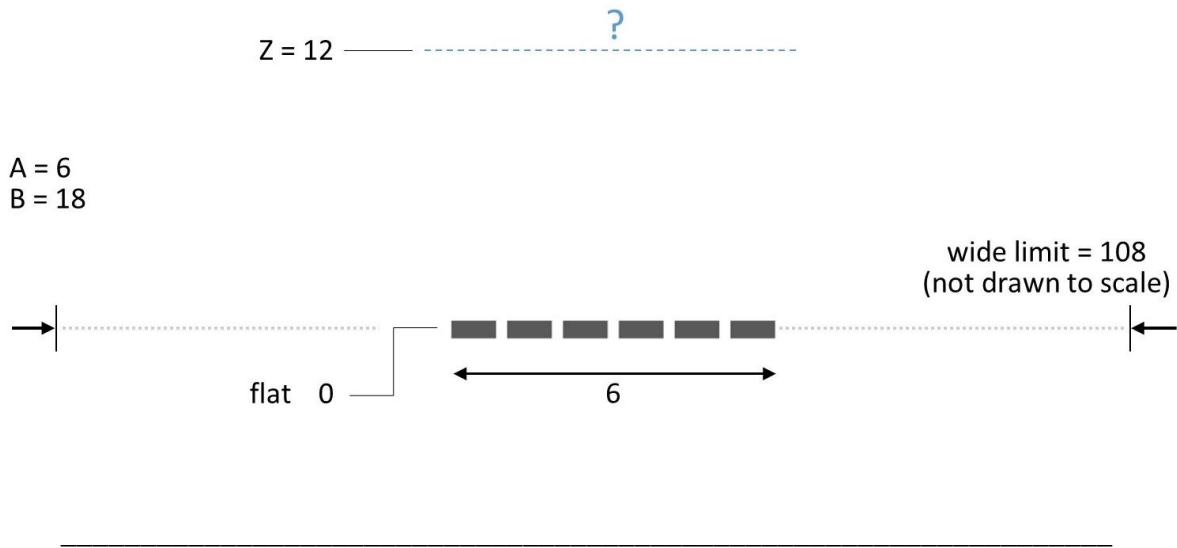
2.1 Overview

In the coming section, we're going to learn how to do calculations with the continuous-factor model. The inputs and outputs of these calculations are the same as in the previous lectures.

The first input is the static factor chain at the flat. The second input is the distance Z from the flat to the plane that we're interested in.

The output of the calculation is the distribution, which consists of two parts: the first is the screen pattern, which is the spatial pattern that you can observe on a screen or a camera. The second is the roundness, which is the number of different angles that are present in the light. Alternatively, it's the number of source patches that emit light to each target patch.

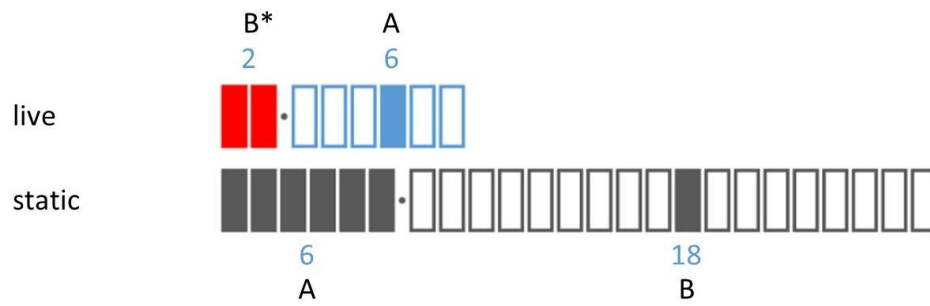
2.2 Review: discrete-factor beam



We're going to step through this method by applying it to the beam. Before we start, let's review a beam problem using discrete factors, for the sake of comparison.

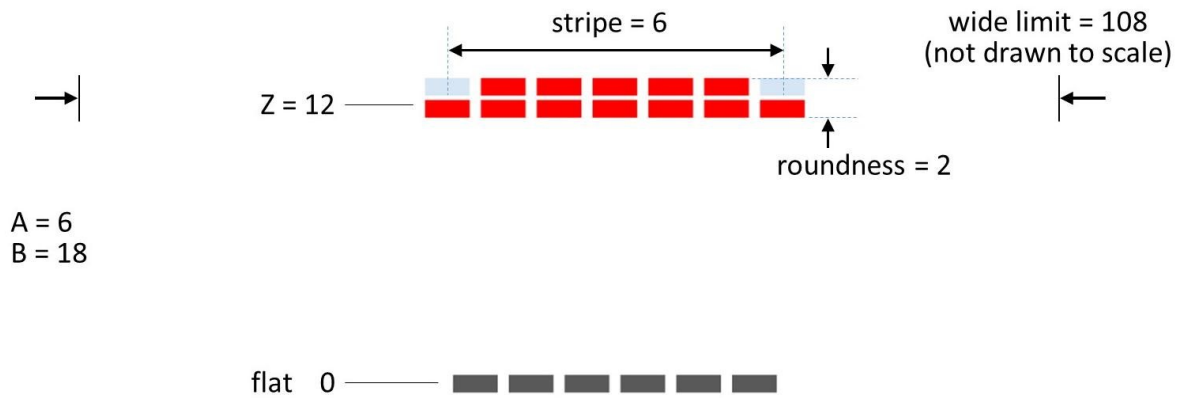
We're given a beam, where the stripe is 6 patches wide at the flat, so static A is size 6. The wide limit is 108 patches, so static B is size 18. We're seeking the distribution at $Z = 12$

2.3 Review: discrete-factor beam, factor chain



The first step is to calculate the live at $Z = 12$. We won't go through the details, but the result looks like this when it's drawn with a linear scale.

2.4 Review: discrete-factor beam, distribution



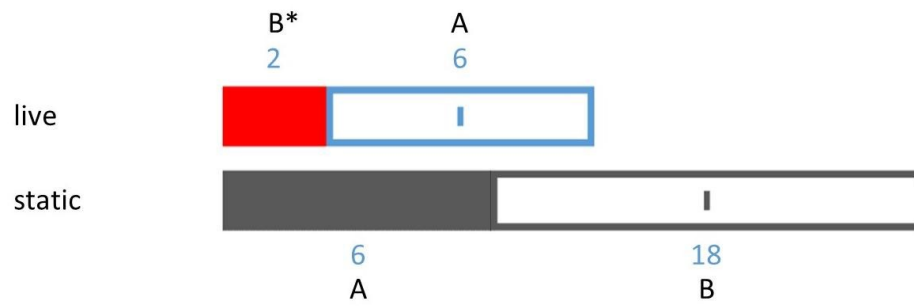
We tabulate the source-target grid, and eventually, we find that the screen pattern stripe is 6 patches wide, the wide limit is 108 patches wide, and the roundness is 2. That was the result using discrete factors.

Now, we'll do the same calculation using continuous factors.

2.5 Steps of CFM calculations

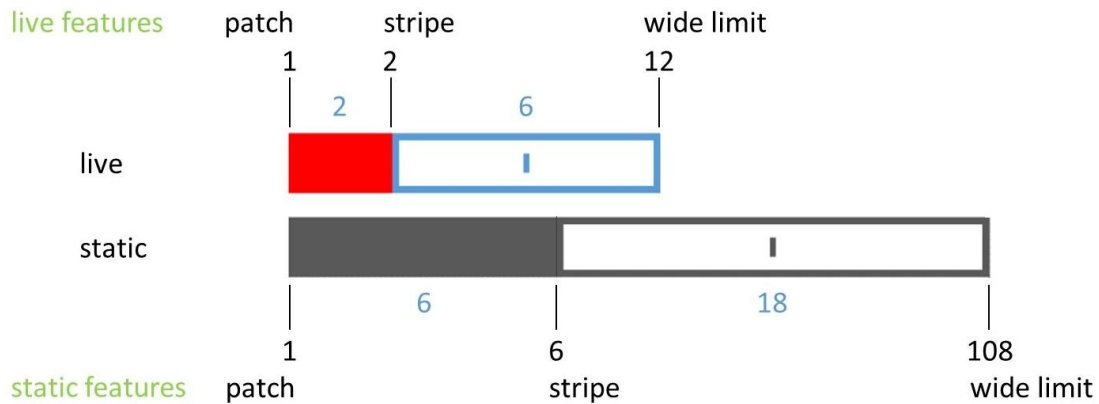
1. Calculate static and live factor chains
2. Split *each* chain into sub-factors, using the *other* chain's features as breakpoints.
3. At each rank, the screen pattern is plenary, if-and-only-if *either* static or live is plenary (OR)
4. At each rank, the roundness is plenary, if-and-only-if *both* static and live are plenary (AND)
5. Merge factors in the screen pattern
6. Plot

2.6 Factor chains



The first step, again, is to calculate the live at $Z = 12$. This step is done exactly like in the discrete-factor model, except that now we're drawing it using a slightly different type of diagram. But the calculation that it represents is exactly the same.

2.7 Splitting factors

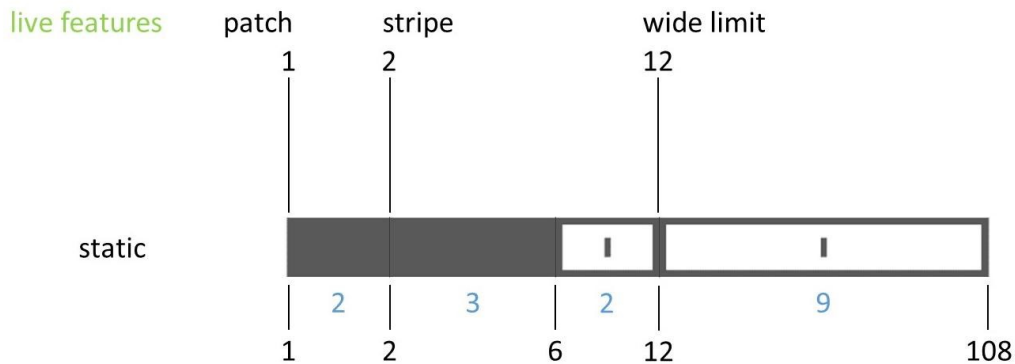


The following step is to split each chain into sub-factors, using the other chain's features as breakpoints. That's *features*, not factors. Let's see how this works.

These are the factor sizes, here in blue. These are the feature sizes, in black. Recall that each feature consists of the cumulative product of all the factors in the chain, from a single patch up to the ceiling or floor of some factor. The live has a stripe at 2, and a wide limit at 12 which is the product of the factor sizes 2 and 6. The static has a stripe at 6, and wide limit at 108 which is the product of the factor sizes 6 and 18. Everything always has a patch of 1.

Next, we'll split the static into sub-factors.

2.8 Splitting static



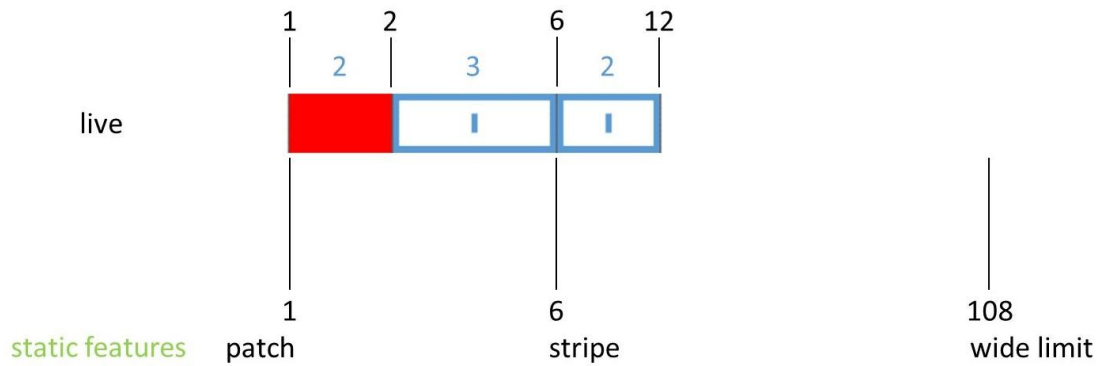
We start with the lowest-ranking factor. Its floor is at 1 and its ceiling is at 6. We split it by any live features that are greater than 1 and less than 6. In this case, there is one such feature at 2.

To split, we just include 2 as a new static feature. Then, to determine the sizes of the sub-factors, we just take the ratios of adjacent features. 2 divided by 1 is 2, and 6 divided by 2 is 3. So the sub-factors are sizes 2, and 3. The original factor was plenary, so the sub-factors are also both plenary.

Now we proceed to the next factor. Its floor is at 6, and its ceiling is at 108, so we look for all live features between 6 and 108. There is one, which is 12. To split the singular factor, we include 12 as a new static feature. The size of the lower subfactor is 12 divided by 6, which is 2. The size of the higher subfactor is 108 divided by 12, which is 9. The subfactors are both singular, like the original factor.

Now, the static chain has been split into four factors.

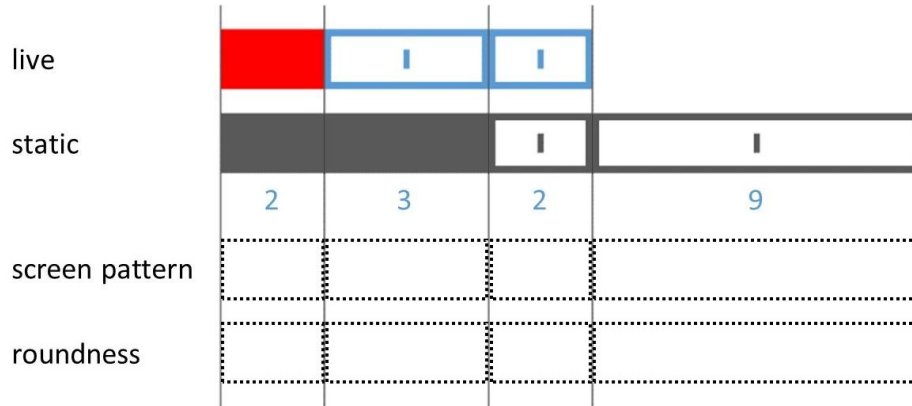
2.9 Splitting live



Next, we will split the live chain by the static features. The first factor does not get split, because there is no static feature between 1 and 2. The second factor has a floor at 2 and a ceiling at 12. This contains the static feature at 6, and so we split the live at 6, into 2 sub-factors, sizes 3 and 2. Now we're done splitting the live.

Note that the static feature at 108 is higher than any of the live factors, so it doesn't do any splitting.

2.10 Common ranks for static and live



If you split each chain with the other one's features, the two chains necessarily end up sharing a set of common ranks, so that the factors for both chains have the same sizes 2-3-2, although the static usually also has some additional high-ranking factors. But while the two factors at every rank share the same size, they don't necessarily have the same type.

The screen pattern and the roundness will also be expressed as factor chains, and they will also share the same common ranks as the static and live. We simply need to determine whether each of these factors is of plenary or singular type.

2.11 Screen pattern from quasi-Boolean rules



The types are determined using *quasi-Boolean rules*. The first one is: *At each rank, the screen pattern is plenary, if-and-only-if either static or live is plenary*. In other words, it is similar to an 'or' operation in Boolean logic.

Let's work the example. At the lowest rank, static and live are both plenary, so the screen pattern is plenary.

At the next rank up, only the static is plenary, but for an 'or' operation, 1 plenary factor in either chain is enough, so the screen pattern is plenary again.

At the next rank, both chains are singular, so the screen pattern is singular.

And finally, at the top rank, there is no live factor, but for this purpose we treat an empty rank like a singular factor; so, again, both chains are singular, and the screen pattern is singular. Now we've now calculated the factor chain for the screen pattern.

2.12 Roundness from quasi-Boolean rules



Next we apply the second quasi-Boolean rule: *At each rank, the roundness is plenary, if-and-only-if both static and live are plenary.* In other words, it is similar to an 'and' operation in Boolean logic.

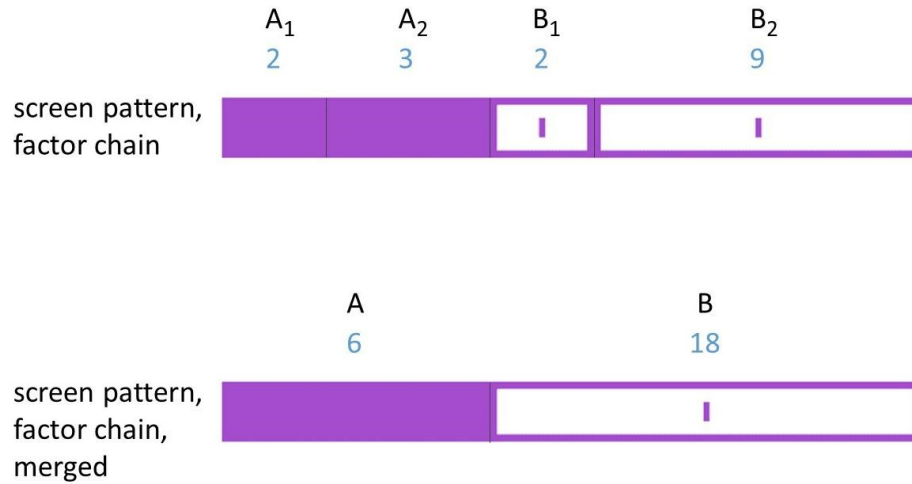
At the lowest rank, the static and live are both plenary, so the roundness is plenary.

At the next rank up, only 1 chain is plenary, not both, so the roundness is singular.

And at the next two ranks up, there are no plenary factors in either chain, so the roundness is singular.

Here we have roundness described by a chain, but actually we want to convert that into a single value. To do this, we take the product of all the sizes of all the plenary factors in the roundness chain. Here, there is only one plenary factor, and its size is 2 – so, the roundness is 2, which agrees with the result from the discrete-factor model.

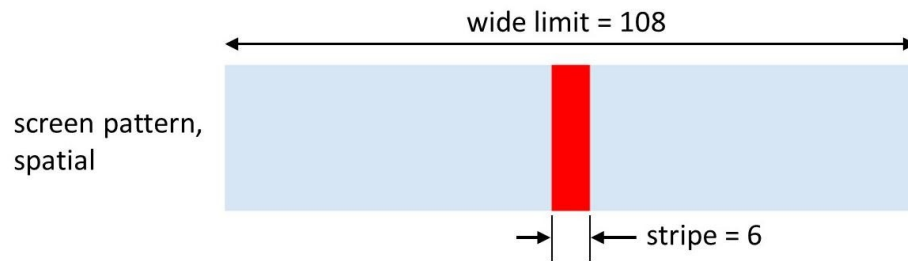
2.13 Merging factors



The next step is to merge factors in the screen pattern. Merging is basically splitting, in reverse. Whenever there are multiple factors of the same type at adjacent ranks, they can be combined into a single factor by multiplying the sizes. Factors A_1 and A_2 are both plenary, so they can be merged. Factors B_1 and B_2 are both singular, so they can also be merged. And this yields a chain with just two factors: A , plenary, size 6, and B , singular, size 18.

In this particular case, the splitting and merging has brought us back to where we started, so it may seem like a lot of trouble for nothing. But this is not always so.

2.14 Plot spatial pattern



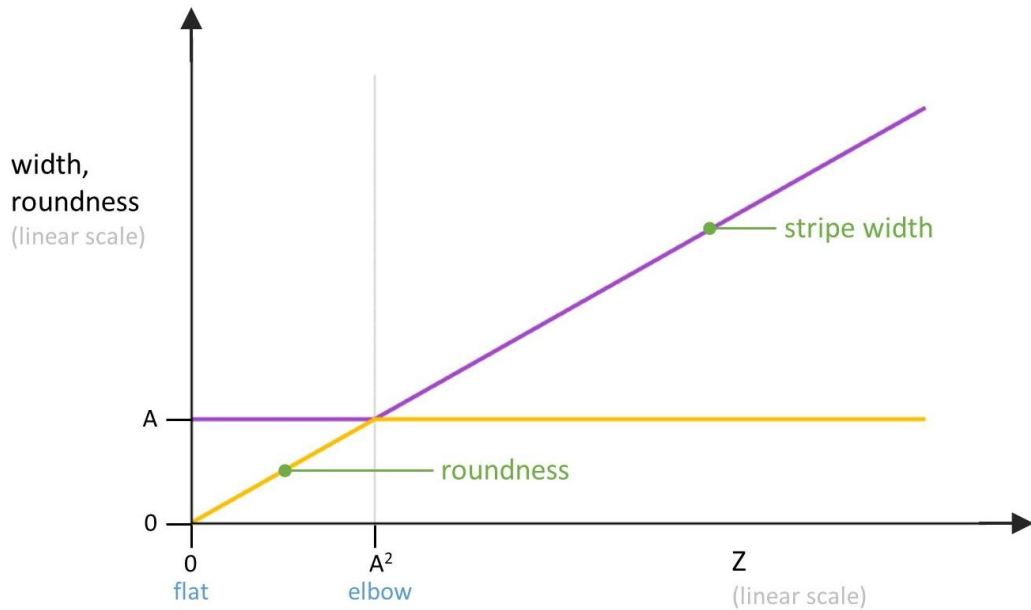
(from drawContinuousFeaturesSpatial.m)

The reason for merging is to prepare for plotting the result. The plotting function is part of the companion code. It expects the factor chain to match a certain format. For the beam, it expects two factors – a plenary and a singular.

This brings us to the final step, which is to plot. We plot using the function *drawContinuousFeaturesSpatial.m* which is provided in the companion code. But it's also pretty easy to calculate the stripe and the wide limit by hand. The stripe is 6 patches wide, and the wide limit is 108 patches wide; this again matches the result from the discrete-factor model.

We're done calculating the distribution at $Z = 12$. In a moment, we'll do the same thing for a few other planes to see how width and roundness evolve over distance in the beam.

2.15 Stripe width and roundness vs Z



But we'll start first with a preview. This is a plot of the stripe width vs distance. Note that we're putting Z on the horizontal axis, whereas previously we'd always drawn Z going vertically.

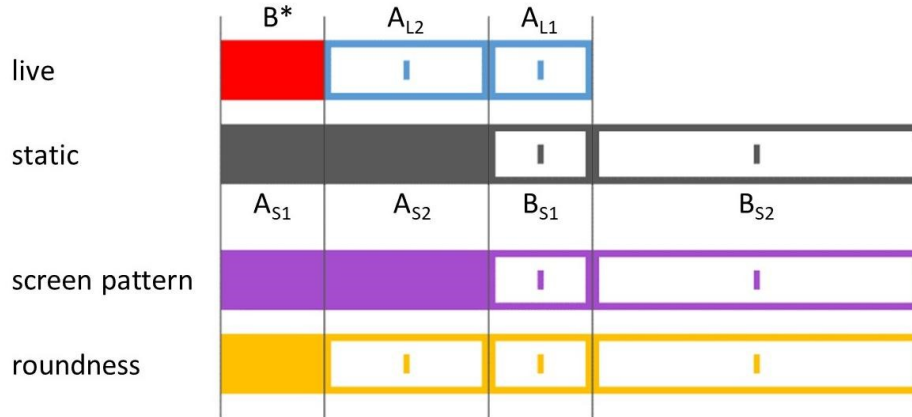
This plot shows that as the beam propagates from the flat, the stripe width remains constant, up to the elbow. Then, it begins to grow linearly in proportion to distance..

The roundness starts at zero, and then rises in proportion to distance. When it reaches the elbow, it's equal to the stripe width. Then it remains constant from that point onwards.

These same general trends appeared in the discrete-factor model of the beam, too – but there, it was more of an approximation. There was some ambiguity in how we defined the width of the beam, and the roundness was also not perfectly uniform over the whole pattern, so the trends were not as stark as they are here. But in the continuous-factor model, both of these functions have a sharp kink at the elbow. In fact, that's the source of the term 'elbow'.

Now let's see how the continuous-factor model leads to these results.

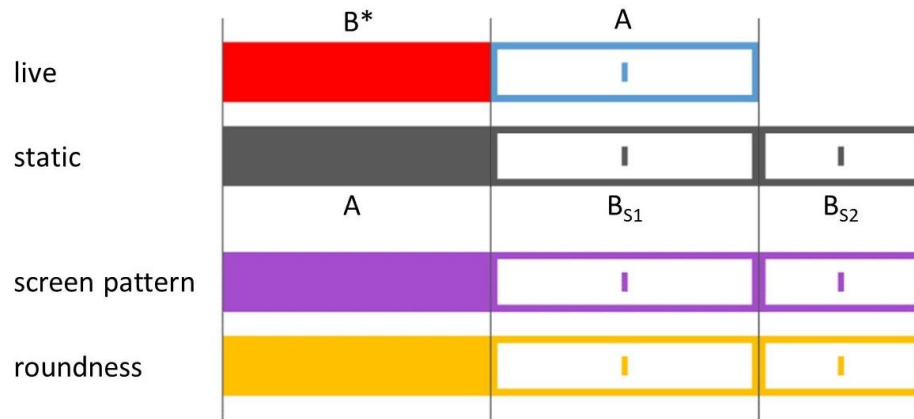
2.16 Near field



The example we just did at $Z = 12$ is a typical plane in the near field. Here, live B can expand without affecting the screen pattern at all. It only takes 1 plenary factor to fulfill the quasi-Boolean OR, and static A is already plenary at this rank. So, live B is just redundant.

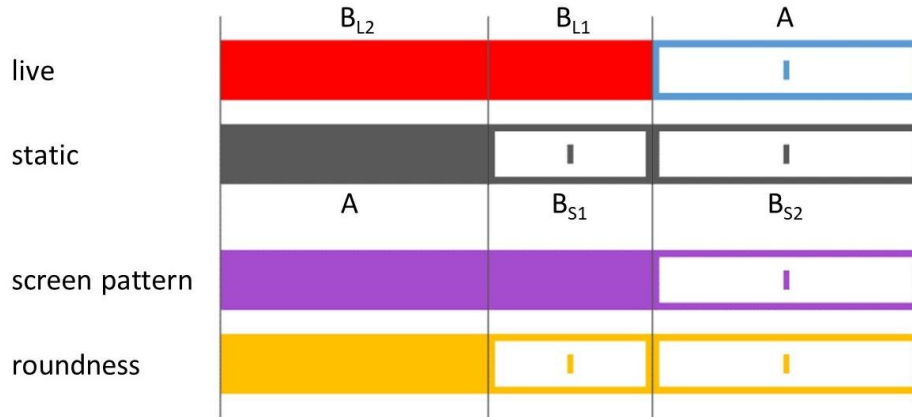
But, when live B expands it does increase the roundness. Static A already fulfills one half of the quasi-Boolean AND, so wherever B is plenary the rule is fulfilled and the roundness is plenary.

2.17 Elbow



Next comes see the elbow, where live B shares a ceiling with static A. In one sense, this is simply an extrapolation of the trend we saw in the near field. Here, stripe width and roundness are both equal to A. But the elbow's main significance is that it marks the transition to the far field. Any distance after the elbow will be in a different regime.

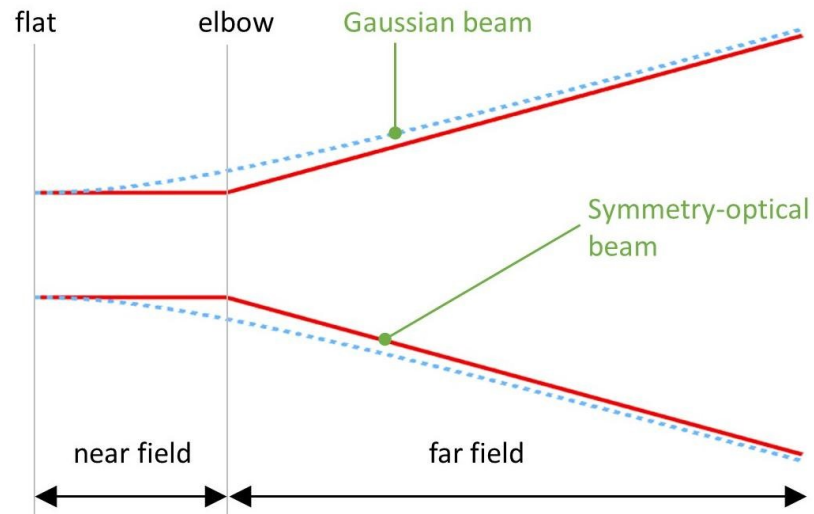
2.18 Far field



Here we see the far limit, which is one example of a plane in the far field. After the elbow, any additional expansion in live B *does* make a difference to the screen pattern, because the static at that rank is singular. Instead of being redundant, the live is the one plenary factor and so *it* determines the screen pattern.

The roundness stays constant by a similar logic. Because two plenaries are required to fulfill the quasi-Boolean AND, live B can expand at this rank, but by itself has no effect on the roundness.

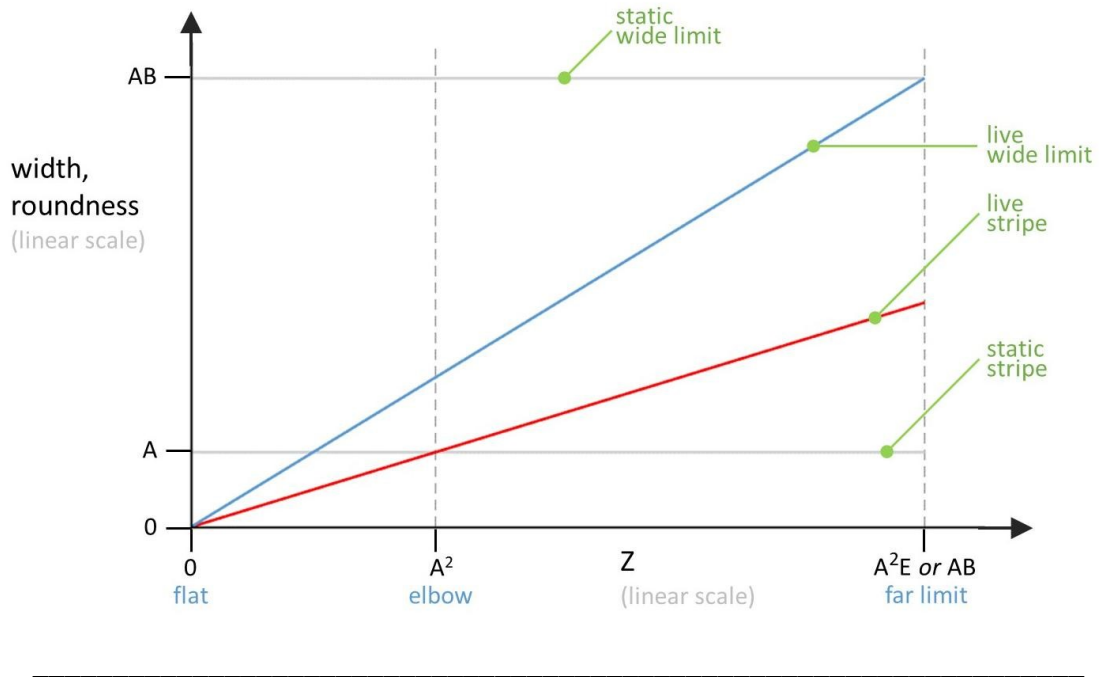
2.19 Comparison with Gaussian beam



We think of the elbow as a sharp kink more for the purposes of conceptual clarity and simplicity, rather than physical accuracy. If you actually observe a real beam in the laboratory, it typically looks more like a Gaussian beam, which is drawn here in dotted blue. The Gaussian beam is a more conventional model. It changes gradually from a straight section to a cone-like section.

The two models agree very close to the flat, and very deep into the far field. The greatest discrepancy between the two occurs at the elbow, which is also called Rayleigh range. There, the Gaussian beam is larger by a factor of root-2.

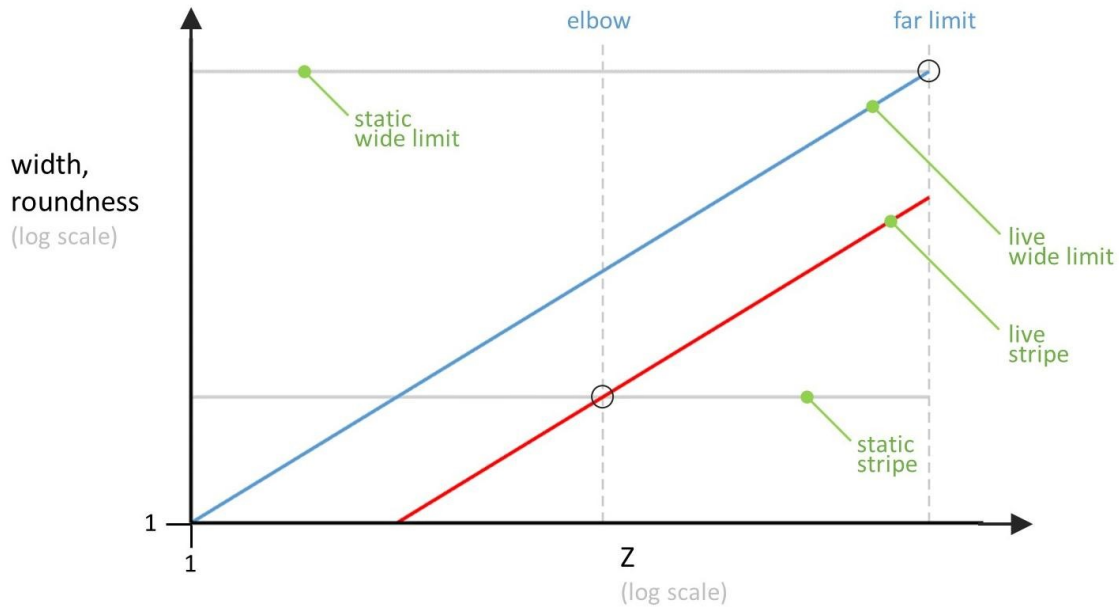
2.20 Beam features vs Z, linear scale



Let's look at the static and live features as a function of Z. For now, these are plotted using a normal, linear scale.

Note a couple of things: First of all, the live stripe and the live wide limit are NOT parallel lines. They clearly have different divergences. Secondly, the origin here is zero. This shows that at the flat, which is $Z=0$, the live features have width zero.

2.21 Beam features vs Z, log scale



Next, we're going to draw the same functions, the same physical thing – but this time, we draw using log scales for both axes. Note that live wide limit and the live stripe appear as parallel lines on the log scale, even though they actually have different divergences. Also, the origin of this plot is the point 1,1. This line represents the plane 1 wavelength from the flat, and this line represents a width of 1 patch. So, this shows us that at 1 wavelength from the flat, the live wide limit is just one patch wide. And what about being size-zero at the flat? The flat is located at minus-infinity to the left, and zero feature size is minus-infinity downwards. For practical purposes, we trim the plot here.

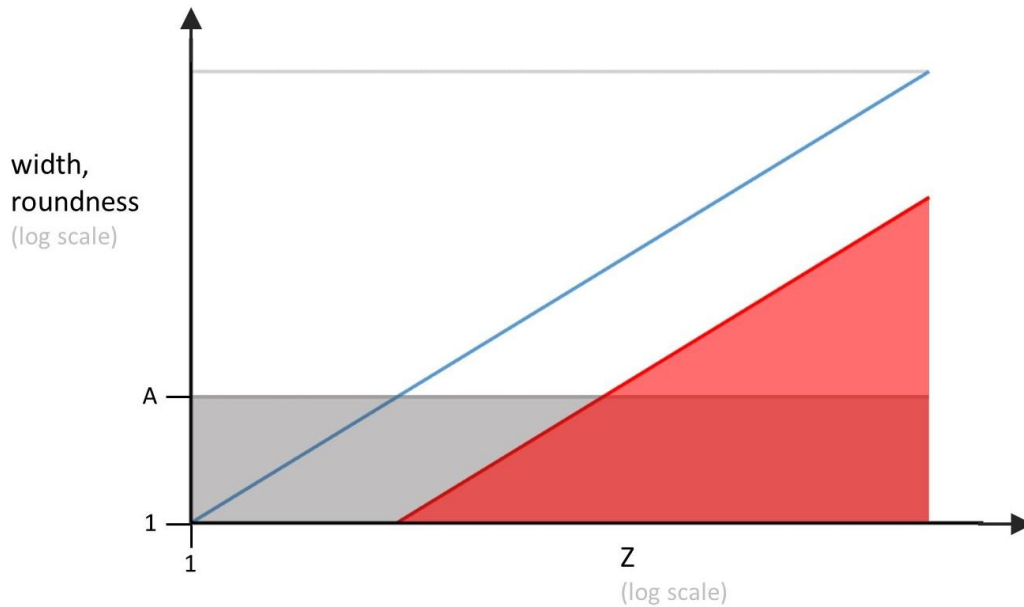
The static features remain constant over distance, so while the log scale changes their positions, their qualitative appearance is just like on the linear scale.

We can imagine any plane as a vertical slice through these lines. And each and every slice contains the same information as a factor chain diagram. So, the plot as a whole is actually a compact representation of the factors at all planes, from 1 wavelength all the way up to the far limit.

The critical planes occur when a live factor shares a ceiling with a static factor. Equivalently, it's where a live feature is as wide as a static feature. These planes occur where lines cross on the plot, as highlighted by these circles. Here, the live stripe is as wide as the static stripe, so this plane is the elbow. Here, the live wide limit is as wide as the static wide limit, so this plane is the far limit.

I said earlier that in the continuous factor model, you can't really visualize what's going on. That's true, in the sense that you can't easily picture how the light physically appears in space. But you can use visual techniques to represent the math, as we've been doing all along. And it turns out that this plot of features on a log scale is very true to the quantitative relationships we're trying to understand.

2.22 Shaded plenary-factor regions



Because it's so much more convenient to think in visual terms, here is another trick that can make it easier to interpret the plot.

If we shade in the region that represents the static plenary factor, it makes it much easier to identify. We can also do the same for the live plenary factor. And, if we superimpose the shaded regions, we can see at a glance how the stripe and roundness evolve.

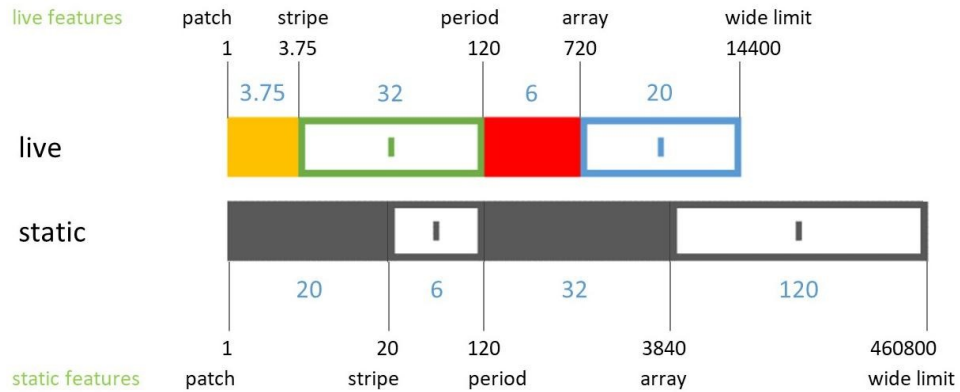
If there is shading in either color, then it contributes to the screen pattern. So the screen pattern starts at width A , continues at a constant, and then at the elbow, it starts to expand.

Where the two shaded regions overlap, then there is roundness. This is where roundness increases up to the elbow and then remains constant.

That's it. We've now learned all the principles of the continuous-factor model, in the context of the beam.

3 Many-Slit Interference

3.1 Factor chains

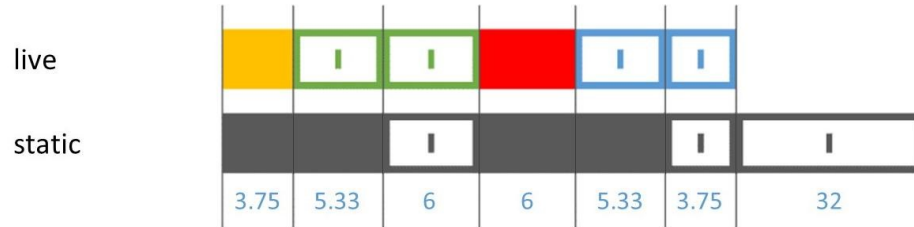


Now, let's apply these very same principles to Many-slit interference. We'll focus on this example. We're given these static factor sizes. From lowest to highest rank, they are 20, 6, 32, and 120, and we want to calculate the distribution at the interval elbow, which occurs at a distance of 14,400.

We start by calculating the live factor chain at 14400.

Next we calculate all the features, in both chains. Again, the features are the cumulative products of the factor sizes – so for instance, the static array is $20 \times 6 \times 32$, or 3840

3.2 Splitting factors



Next we split the static by all of the live features, and then split the live by all of the static features, which gives us a set of common ranks.

The screen pattern and the roundness will also share this set of ranks.

3.3 Applying quasi-Boolean rules



We calculate those by applying the quasi-Boolean rules. For the screen pattern, the factor types are determined by a quasi-OR operation, so the types are: plenary, plenary, singular, plenary, plenary, singular, singular.

For the roundness, the factor types are determined by a quasi-AND operation, so the types are: plenary, singular, singular, plenary, singular, singular, singular. To convert the roundness chain into a number, we multiply the plenary sizes. So, that is 3.75 times 6, which is 22.5

This last step raises an interesting question: is roundness simply a number, like 22.5? Or does it have any other kind of structure, described by the factor chain? For example, suppose that rather than having plenary factors at the 1st and 4th ranks, we instead had plenary factors at the 1st and 3rd ranks? The 3rd and 4th ranks are both size 6, so in both cases the roundness would be 22.5. But, the chains would be different. Shouldn't that imply some kind of difference?

This question has not yet been fully explored. The roundness chain *does* seem to describe the pattern within each column of the source-target grid. But we haven't taken those patterns into consideration; instead, we've simply collapsed each column, and counted the bright cells. So for now, roundness is just a single number. This may change in the future as the field develops.

3.4 Screen pattern factor chain, spatial pattern



We have the screen pattern as a factor chain, but we still want to convert it into a spatial pattern. The chain contains adjacent factors of the same type, so we merge them into just 4 factors, with the types plenary-singular-plenary-singular.

And that chain is the input to a plotting function, which gives us this spatial pattern. It's hard to make out the details, because the pattern is more than 99% dark space, it's actually 119 / 120ths dark space.

So, to focus on the interesting part, we can use the option of trimming the excess dark space and showing only the array. However we choose to display it, here's our result. The calculation is done.

3.5 Companion code

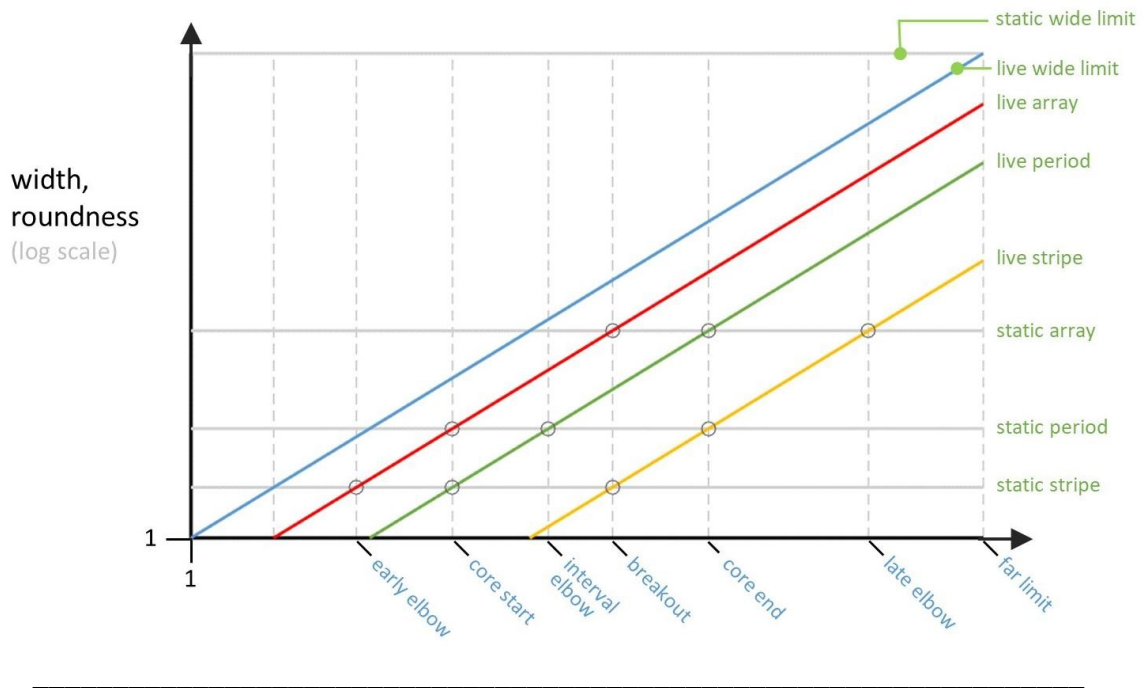
Companion code is available to let you try all of these calculations yourself. Here's just one example, in the file *factorsAndScreenPatternMSI.m*. Running this file calculates and draws the factor chain and the screen pattern for any plane except the flat.

By the way, note that there is some ambiguity in where each period starts and ends, so you need to take the visualizations with a grain of salt. For instance, in this case, C is 5, so there are 5 full periods. Let's say we reduce C to 4.4. Where do we trim? These algorithms trim symmetrically from the edges, so it only cuts off dark space and we're still left with 5 full stripes. This could be a little bit misleading.

If we reduce C to 4, or any even number, it cuts the first and last periods in the middle of the stripe. This is also a little misleading, because all of the stripes are the same size in principle. In other words, don't interpret it too literally. The benefit of drawing it like this is that it's easy to see the exact boundary of the array. For cosmetic purposes, you may wish to set the variable `centerOdd` to false, and this will center even numbers instead.

In all cases, remember not to attach too much importance to precisely how the first and last periods are drawn.

3.6 MSI features vs Z, log scale

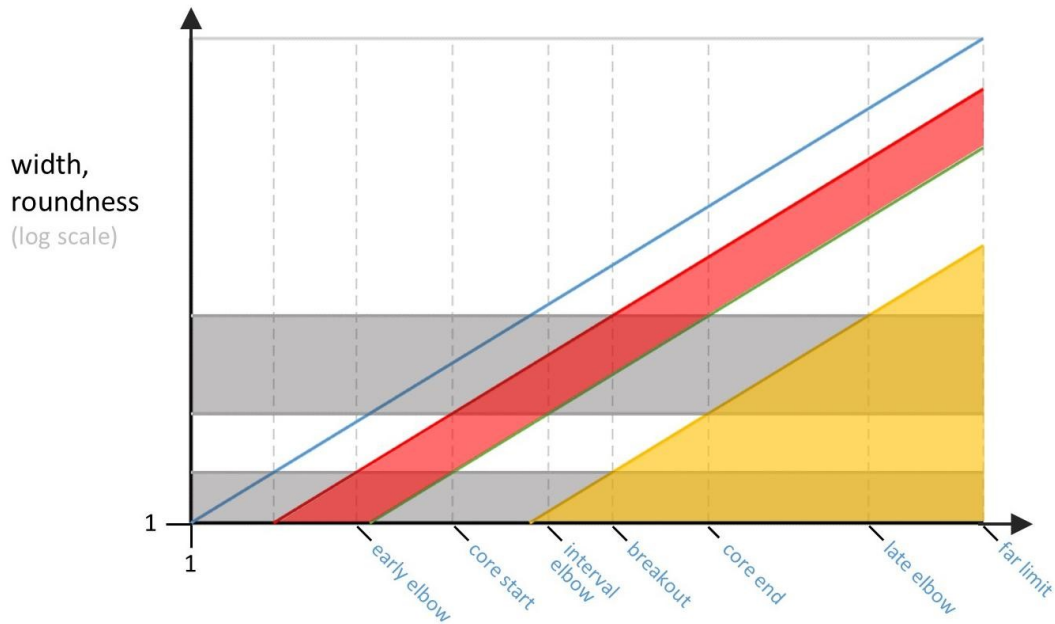


Now let's plot all the features on a log scale, just as we did for the beam. Again, the live features are a set of four straight, parallel lines which rise at an angle. Remember that even though the lines appear parallel on the log scale, they actually have four different physical divergences. The static features, again, are plotted as four horizontal lines.

And, when we plot both the static and the live, and find the points where they intersect one another, we find the critical planes. For instance, this point is where the live period intersects the static array. This occurs at this plane, which is the core end.

If there is ever a textbook on Symmetry Optics, this image will be on the cover, because it very accurately shows the big picture of Many-Slit Interference as well as the fine details, all condensed into one elegant, powerful diagram. It's worth perusing in detail.

3.7 Shaded plenary-factor regions

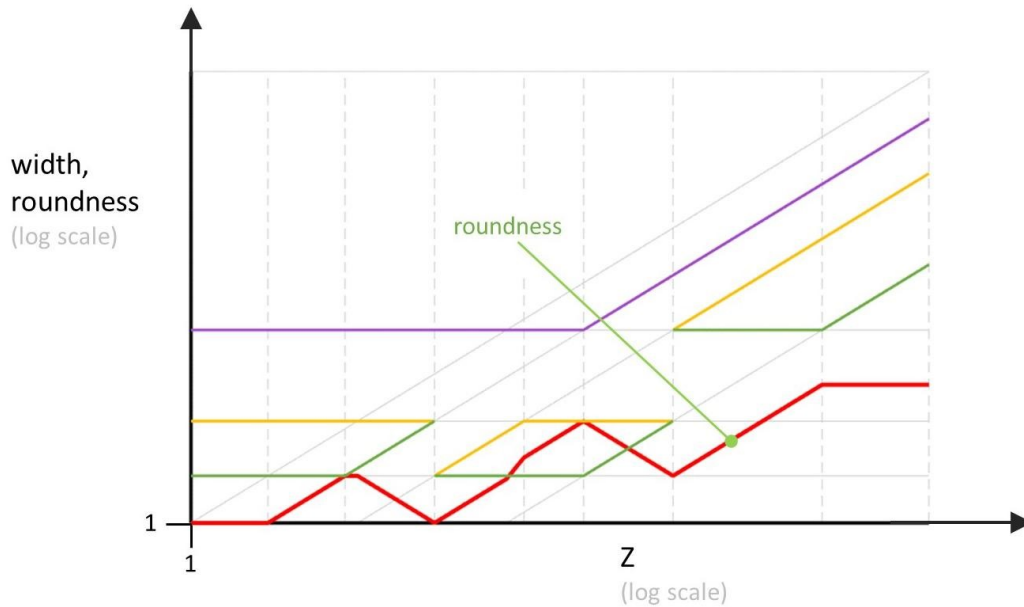


Just like with the beam, it's handy to shade in the regions that represent the plenary factors, because it makes it easier to interpret at a glance. To find the array of the screen pattern, we trace along the top edge of the shaded region.

To find the period, we trace the top edge of the open region. It jumps at the core start, and then jumps again at the core end.

To find the stripe, we trace along the bottom edge of the open region. At the core start, it meets up with the period, and it jumps opposite to the period – and, the same thing happens at the core end.

3.8 Screen-pattern features, plus roundness



This is what the features of the screen pattern look like.

To find the roundness, we look for the regions that are doubly shaded, with grey and color overlapping. So it starts here, not at zero but at 1, because in this logic, not having any factors equals 1. Then here we see it increasing, then decreasing back to zero, then increasing again. Eventually, there are contributions from two different regions of overlap.

The roundness is plotted here in red. It rises and plateaus and falls in a fairly complicated way. While the trends in the screen pattern are qualitatively similar for all cases, the roundness trends can vary a fair amount from one case to another.

We're now finished calculating the distribution for Many-Slit Interference.

4 Additional topics

4.1 Overview

Now that we've covered the basics, let's explore a few additional topics:

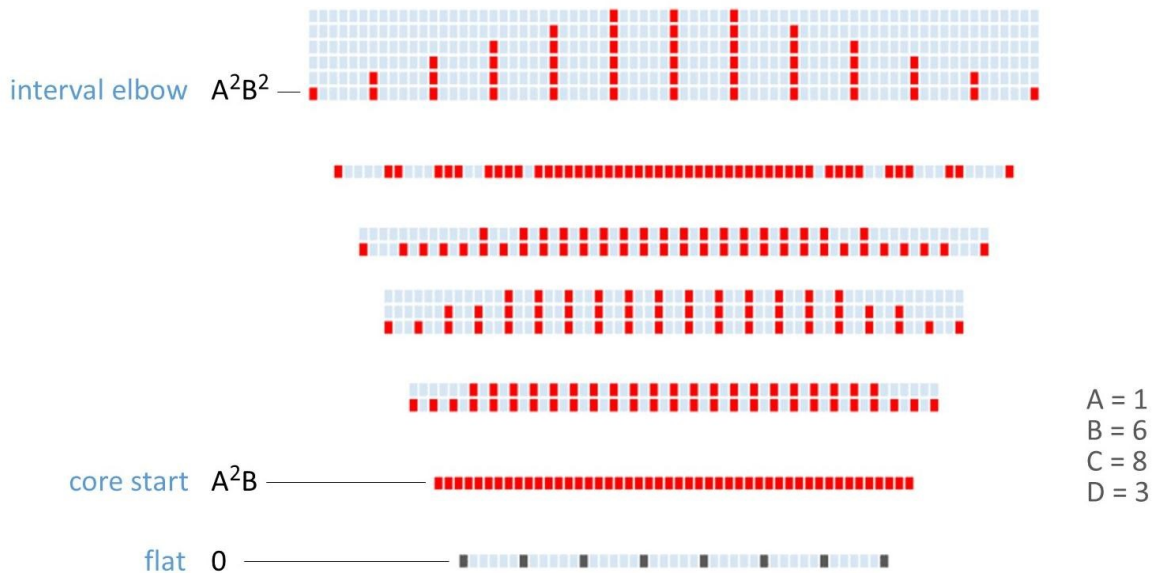
First, we'll discuss coincidences.

Then, we'll learn about early- vs late-breaking interference.

Next, we'll discuss conceptual beams

And finally, we'll talk about confirmation calculations for the continuous-factor model.

4.2 Coincidences occur only in discrete-factor model



The continuous-factor model has one important drawback, which is that it fails to model coincidences. To illustrate, here's a slide from the discrete-factor model. From the core start to the interval elbow, the general overall trend is that the period gets wider. It starts at 1, then grows to 2, 3, and eventually 6.

But it's not simply a monotonic increase – it actually goes 1, 2, 3, 2, 1, 6.

The continuous factor model only gets the overall trend correct, and so it will find the period as 1, 2, 3, 4, 5, 6. For this reason, it needs to be applied with some judgment.

4.3 Early- vs late-breaking interference

Early-breaking ($B > C$)

| Plane | Z |
|----------------|-----------------------------|
| far limit | $A \cdot B \cdot C \cdot D$ |
| late elbow | $A^2 \cdot B^2 \cdot C^2$ |
| core end | $A^2 \cdot B^2 \cdot C$ |
| interval elbow | $A^2 \cdot B^2$ |
| breakout | $A^2 \cdot B \cdot C$ |
| core start | $A^2 \cdot B$ |
| early elbow | A^2 |
| flat | 0 |

Late-breaking ($C > B$)

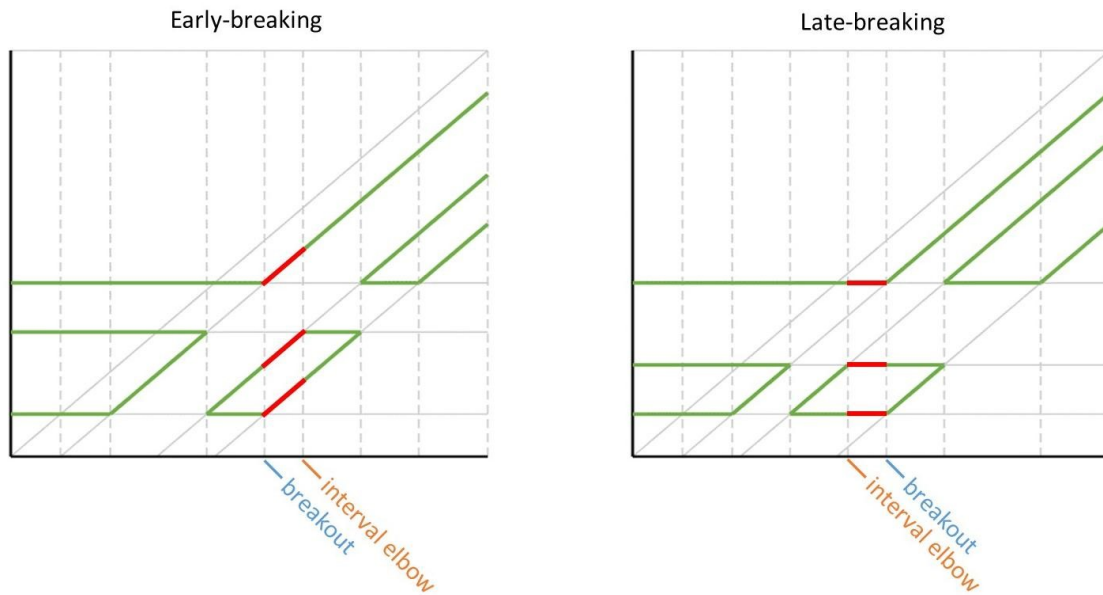
| Plane | Z |
|----------------|-----------------------------|
| far limit | $A \cdot B \cdot C \cdot D$ |
| late elbow | $A^2 \cdot B^2 \cdot C^2$ |
| core end | $A^2 \cdot B^2 \cdot C$ |
| breakout | $A^2 \cdot B \cdot C$ |
| interval elbow | $A^2 \cdot B^2$ |
| core start | $A^2 \cdot B$ |
| early elbow | A^2 |
| flat | 0 |

Next we'll discuss the differences between early-breaking and late-breaking interference. This chart shows all the critical planes, and it shows the distances where they occur. For the most part, they always occur in this order— but, breakout is the one exception. The interval elbow is at $A^2 B^2$, and breakout is at that same distance *times* the ratio C/B . But, what if C is actually smaller than B ? In that case, breakout would actually be located *before* the interval elbow.

To get a better feel for this, we can re-express the breakout distance as $A^2 \cdot B \cdot C$. This shows more clearly that breakout is always a factor C after the core start, and always a factor B before the core end. But where breakout lies in relation to the interval elbow depends on whether B or C is larger.

This lets us categorize all cases of interference into several basic types: First is the early-breaking case, where $B > C$, and breakout occurs before the interval elbow. Second is the late-breaking case, where $C > B$, and breakout occurs after the interval elbow. And of course, there is also a third case where $B = C$, and then the interval elbow and breakout occur at the same plane.

4.4 Early- and late-breaking interference on log scale



On the log feature plot, the difference is visible in the stage between breakout and interval elbow. When the features break out early, they all expand uniformly up to the interval elbow. But when the features break out late, they all remain constant from the interval elbow up to breakout. The roundness actually grows in this case.

4.5 Early- and late-breaking interference in simulated videos

Let's see how this difference appears in spatial terms. This video shows a simulation, not real data. Like in the previous lecture, we're seeing one object at two different magnifications. The object here is a system of many-slit interference, but because we're seeing the core start it appears as just one solid region rather than as a repeating-stripe pattern. We're going to see the trends in 3 different stages, which together constitute the core.

(early-breaking.avi)

In the first stage, the period grows wider, but the stripe and the array both stay at a constant width. We can think of those features as being constrained by the static. But they begin to break out of those constraints, naturally, at breakout. In the 2nd stage, the pattern 'breaks out' and appears to grow by a single overall magnification factor. Superficially, this looks very similar to the far field, and so we might imagine that this would continue indefinitely. But actually, the trend stops at the interval elbow, where it reaches the period of the original grating. Then in the 3rd and final stage, the period remains constant while the other features grow, until we reach the core end.

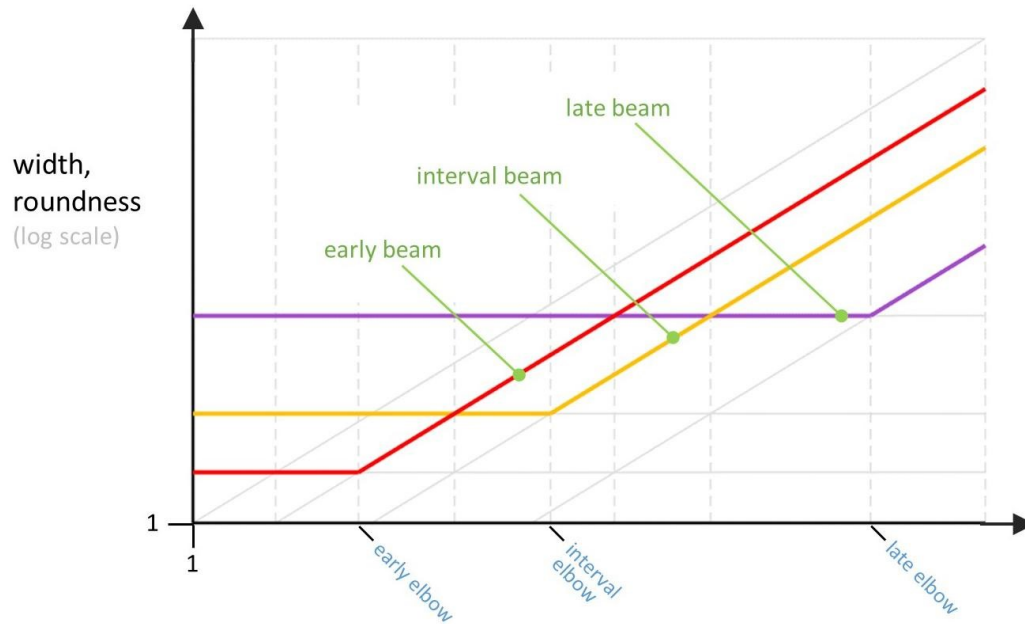
(late-breaking.avi)

Now we'll see the late-breaking case and note the differences. Of the three stages, the first and third are basically the same. At first, the period increases. But it stops at the interval elbow. The next stage is different from early breaking, because the screen pattern remains constant. The coincidences in this stage correspond to the Talbot Effect, in which images of the grating seems to reappear at regular intervals. The third stage is similar to the early-breaking case, and terminates at the core end.

(balanced-breaking.avi)

In the balanced-breaking case, the interval elbow and breakout occur in the same plane, so there are only two stages. The center stage is missing, which is the very stage where you see the difference between early and late breaking. In the first stage, the period grows while the stripe and array stay constant, and in the second stage, the stripe and array grow while the period remains constant.

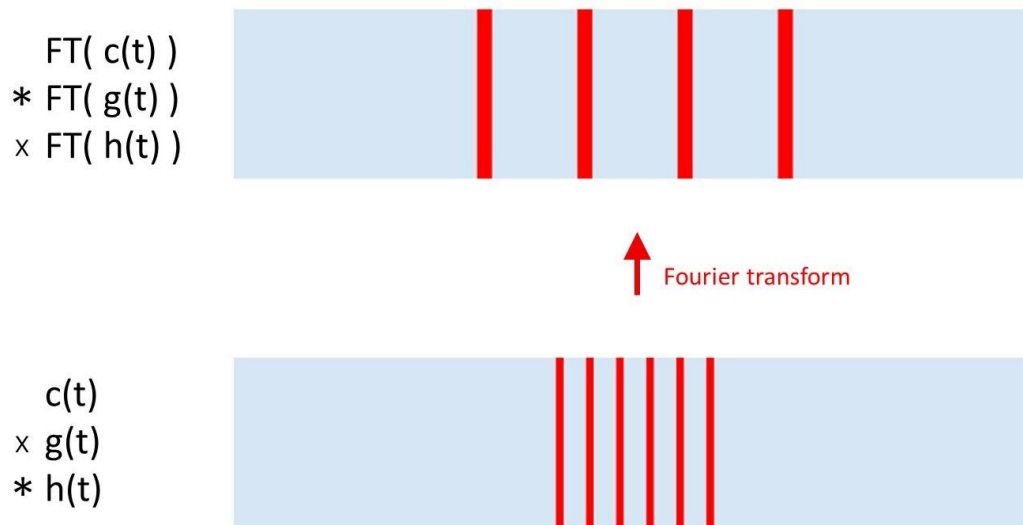
4.6 Conceptual beams



Let's move on to the next topic. We can understand many-slit interference in terms of 3 *conceptual beams*. These are: the *early beam*, the *interval beam*, and the *late beam*.

Each beam starts as one of the static features, and ends as one of the live features. Each beam also has an elbow. As you might guess, these are located at the early elbow, the interval elbow, and the late elbow respectively.

4.7 Fourier transform rules



The conceptual beams don't explicitly exist in conventional optics, but they do roughly correspond to the conventional topic of Fourier transform rules. Let's review these briefly.

Here is a discrete comb function, which we call $c(t)$. A comb function means, a series of very narrow peaks spaced at regular intervals. The Fourier transform of a comb function is another comb function with a different interval.

Now we clip $c(t)$ to some finite width. Mathematically, we are *multiplying* $c(t)$ times a second function, $g(t)$. $g(t)$ is a rectangle function which is zero to the left, 1 in the middle, and 0 to the right. The effect of multiplying by $g(t)$ is to *convolve* the wide comb pattern with the Fourier transform of $g(t)$. So, the stripe profile is the Fourier transform of $g(t)$, and it is applied to each peak of the comb. The stripe profile in this case is actually a more complicated function sinc, but we're just drawing it approximately.

It also works the other way around. In other words, we can convolve $c(t)$ with another function $h(t)$, which defines this stripe profile. And the effect on the Fourier transform is to multiply the whole function by another wide rectangle, the Fourier transform of $h(t)$.

These rules are a very well-known aspect of the Fourier transform, but in symmetry optics we use the conceptual beams to describe the same phenomena.

4.8 Conceptual beams, video

Let's watch some simulated videos of these conceptual beams. Note that for these, both windows use the same magnification but they show different objects. The top window shows many-slit interference, and the bottom window shows the early beam.

(early beam.avi)

At the flat, the early beam is the stripe. In Fourier-optical terms, you would say that it's convolved with the comb pattern. The beam stays constant until the early elbow, and then starts to grow wider, but it remains the stripe up through the core start. For a short while, the beam doesn't correspond to any feature. But at breakout, the early beam reappears. This time, it's the array, not the stripe, of the screen pattern. In Fourier-optical terms, it's multiplied by the comb function. This continues all the way to the far limit.

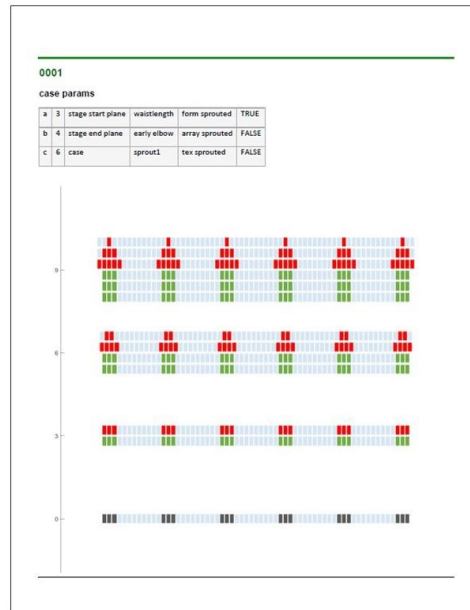
(late beam.avi)

The late beam is the other way around. It begins as the array of the screen pattern. This continues until breakout. Then it's briefly invisible until the core end, when it reappears as the stripe. Finally, after the late elbow, it expands.

(interval beam.avi)

The interval beam is the period of the screen pattern, from the flat up to the core start. Then it goes invisible. At the interval elbow it begins to expand, and it becomes visible again at the core end, where once again it is the period.

4.9 Confirmation calculations



One final topic for this lecture. The continuous-factor model isn't truly derived from the discrete-factor model though any formal mathematical process. Rather, it has simply been discovered by guesswork, and found to yield good results.

To verify and to demonstrate that the two models agree, I'm providing a set of confirmation calculations, which you can download as a pdf. The red distributions were calculated using the discrete-factor model, and the green distributions were calculated using the continuous-factor model.

These calculations modeled a wide variety of different conditions, and overall, the agreement is very good.

5 Conclusions

5.1 Reviewing key points

That's all for this lecture, so let's review the key points:

- The continuous-factor model is an alternative method for computing the same distributions we've been computing all along.
- To calculate distributions, split static and live chains to get common ranks. Then, use the quasi-OR rule to get the screen pattern, and the quasi-AND rule to get the roundness.
- A log-scale plot of feature widths vs distance (Z) gives a very elegant representation of many-slit interference.
- Depending on the relative sizes of parameters B and C , interference can be early-breaking or late-breaking, or balanced between the two.
- Interference can also be understood in terms of conceptual beams: early, interval, and late.

5.2 Outro

Wherever you found this video, you'll also find links to more resources for learning symmetry optics.

Depending on your learning style, the lecture notes may be more effective than watching the video.

And I recommend trying the problem set, because you can only learn so much by watching someone else do the calculations; you need to walk through every step on your own to really master the concepts.

And finally, get the companion code so that you can try your own experiments and explore your own questions.

In the next lecture, we'll see the results of the Symmetry Optics Validation Study. We'll compare symmetry-optical theory against real empirical data to demonstrate that symmetry optics does indeed work. So please join me.

I'm Paul Mirsky; thanks for listening.