The Importance of Variance Control in PV Manufacturing

UNSW Seminar, 3rd March 2016

Rhett Evans

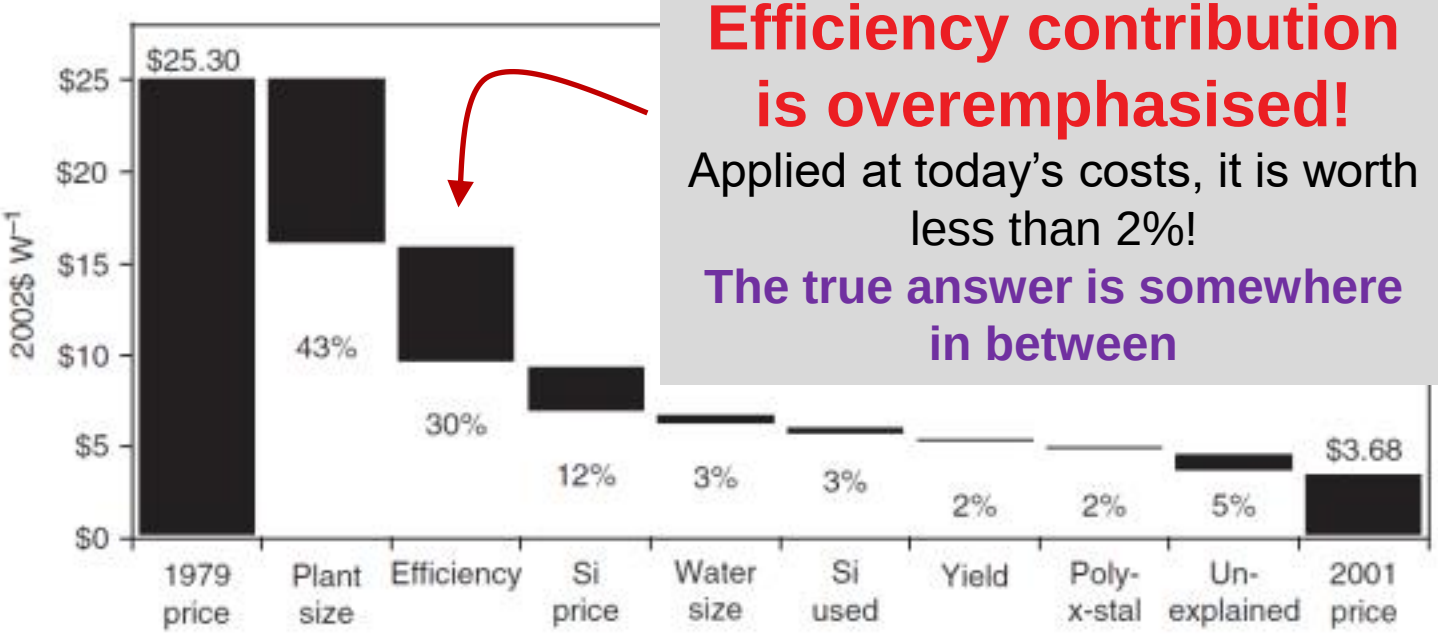
evansrhett@gmail.com

Contents

1. Context of my research
2. Why care about variance?
3. Introduction to Path Models
4. Path Model Solution for PV Manufacturing Data
5. Knowledge is Power……

1. How Important is Manufacturing Research?

- Manufacturing improvements have been crucial in lowering the price of the PV.



Efficiency contribution is overemphasised!

Applied at today's costs, it is worth less than 2%!

The true answer is somewhere in between

from Nemet, G. F., Husmann, D., 2012. Historical and future cost dynamics of photovoltaic technology

- At ~ $0.50-0.60/W, photovoltaics has a very competitive LCOE and the technology is likely to undergo significant expansion.

1. Research Context

- Photovoltaic manufacturing is an industry that can best be described as being in its “adolescence” (Verlinden 2013)
- This definition fits with the growth of other industry sectors
  - Market is turbulent
  - Technology development is turbulent
  - Approach to product is rudimentary and based on technology push. Early signs of a market pull approach is developing
  - Unlike other manufacturing sectors, there is nearly nothing in the published literature about the development and optimisation of the manufacturing from a data perspective. Why?
    - Data sensitivity
    - No work in an academic context
    - No motivation to publish in private sector
    - Little work is being done at all
    - Photovoltaic manufacturers are “spoiled”. They can directly measure the cell power anyway!

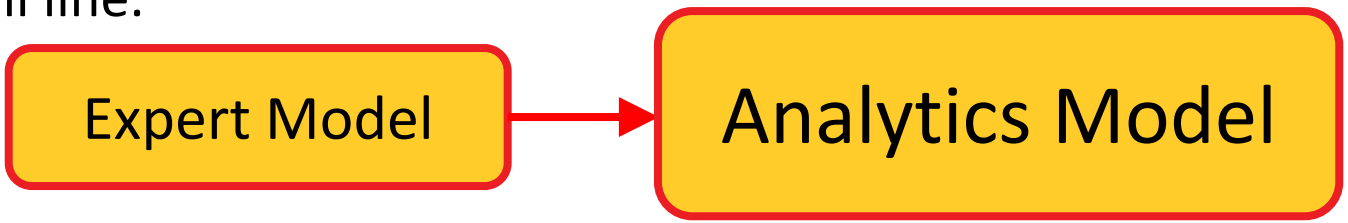
1. My PhD Research Topic

"Increasing the statistical sophistication of photovoltaic manufacturing"

What	Building multivariate statistical models to describe the manufacturing system
Why	Improve understanding of variance and its sources Optimise the utilisation and therefore collection of data Improve product quality Facilitate system level thinking around photovoltaic energy
How	….lets find out

1. Some barriers

- Discussion of statistical techniques needs to become a higher profile topic within our industry.
  - A barrier to this is an apparent embedded hatred of statistics.
- All data must be normalised to share it publically.
  - This can be disappointing or annoying to some people
  - It can also (falsely I believe) be seen as obstructionist
  - But this is standard practice in other industries, and so we need to get over this if this important area of development is going to be discussed in the literature.
- Need to think about solar cell operating theory in terms of their relationships, not just individual values.
- We shouldn't need to be semiconductor experts to debug a solar cell line.



Contents

2. Why care about variance?

- Variance is a direct indicator of product quality.
  - This is quality defined in the manufacturing sense of making something the same every time. i.e consistency
- Average efficiency of production has been on a steady path upwards for sometime, and so mean performance is usually the highest consideration.
  - Can this last forever?
  - What comes next?
- A stable, mature manufacturing industry is more concerned with quality.
  - Maybe we are a few years off this being of dominant importance, but it is important already.
  - We need to move towards developing a **genuine "quality function"** for PV cell manufacturing.

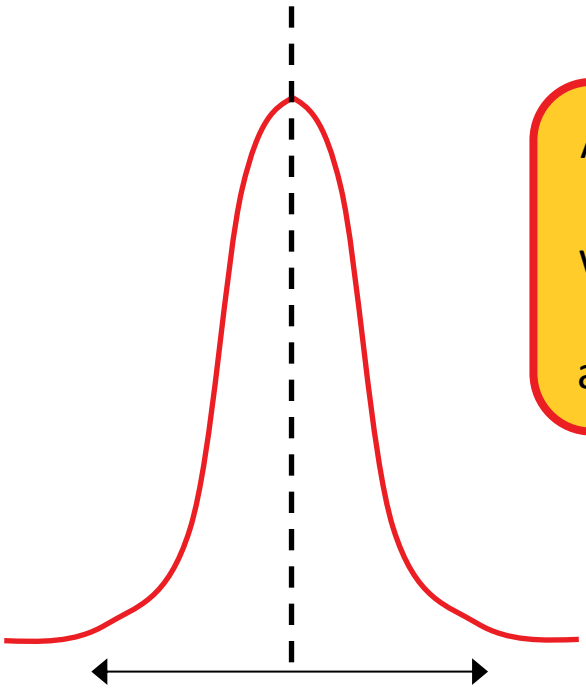
2. Why care about variance?

What would you want if you were an end use customer?

What would you want if you were a cell customer?

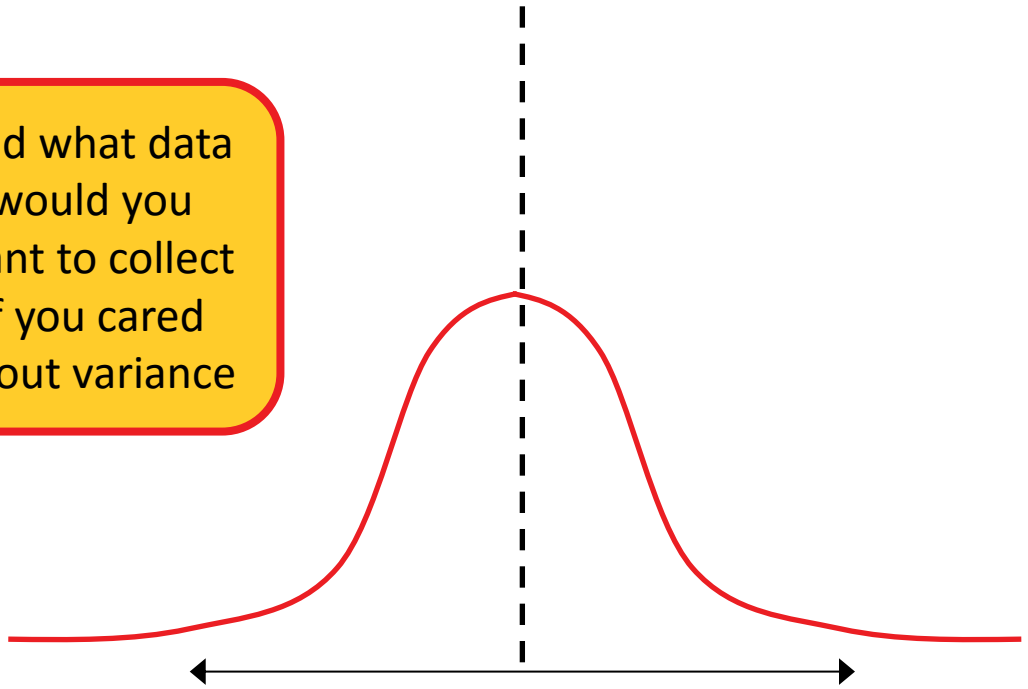
What would you want if you sold the cells?

What would you want if you were manufacturing the cells?



A 19.0 ± 0.1 % process

And what data would you want to collect if you cared about variance



A 19.5 ± 0.3 % process

2. Why care about variance?

What is the value proposition for variance control?

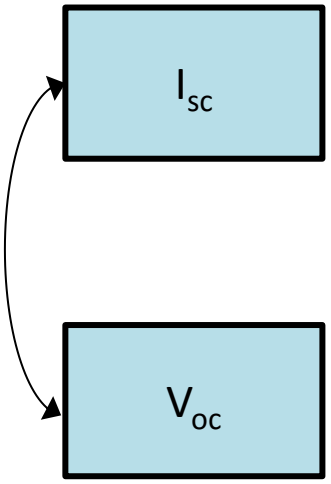
- This in itself is an interesting topic, and we should seek to use actual data and actual operational practices to examine it.
- Ways to improve margin with lower variance include -

Improvement	Value (US c/W)	Who saves?
Electrical Yield	0.5-2	Manufacturer
Experimental Yield	0.1-0.5	Manufacturer
Sales & Logistics	0.5-1	Manufacturer
Field Installation Logistics	1-5	System developer
Energy over a system life	3-5	System operator

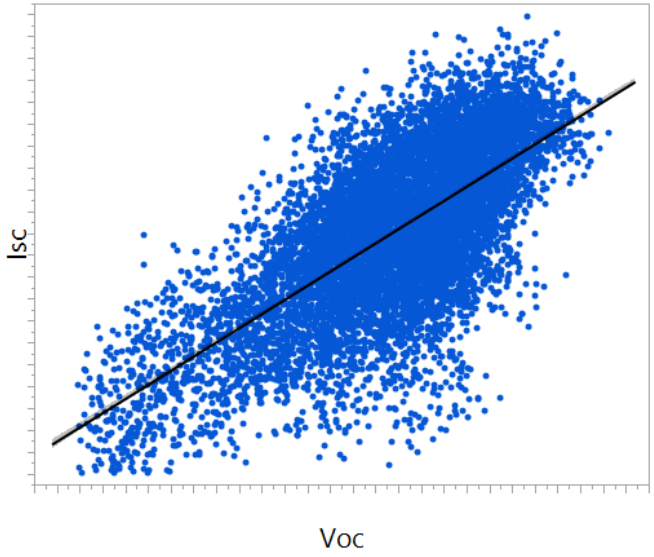
Contents

3. An Introduction to Path Models

- A path model is a way to express the root causes of the relationships between the variables we measure to describe a cells performance.
  - The path models I am using attempt to describe the correlation / covariance between the measurements.
- Start by looking at the correlation between two variables, the I_{sc} and the V_{oc}.



Path Model

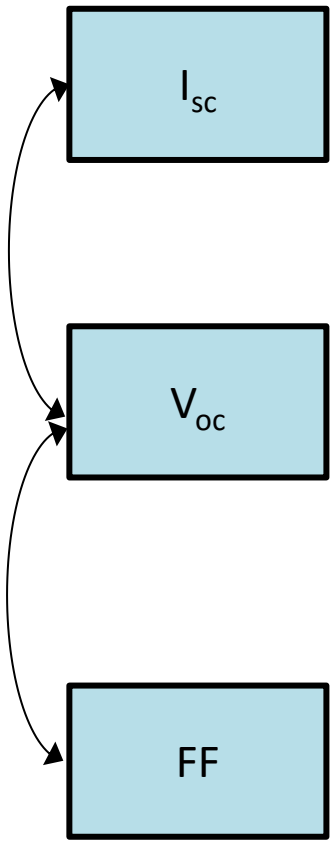


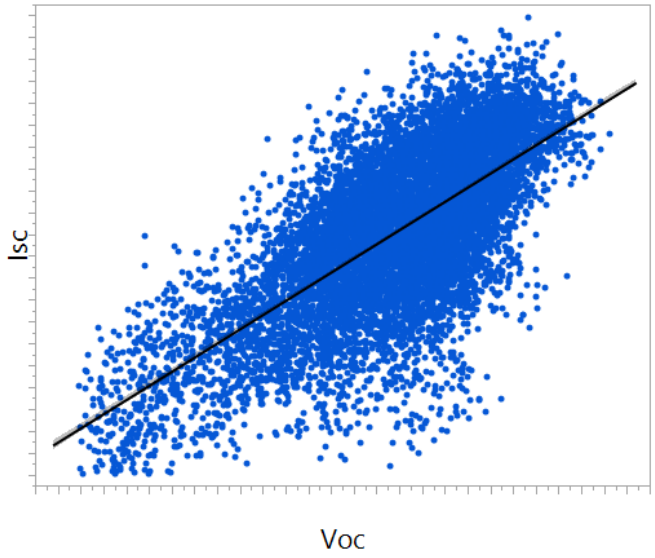
Scatterplot

	Isc	Voc
Isc	1.0000	0.6950
Voc	0.6950	1.0000

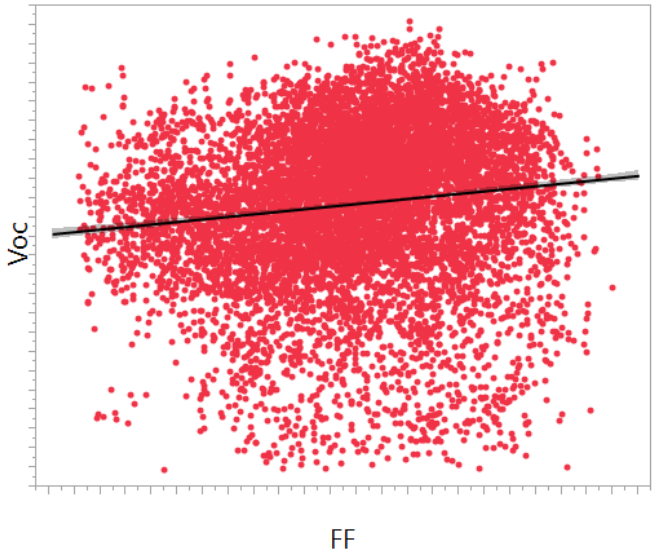
Correlation matrix

3. An Introduction to Path Models



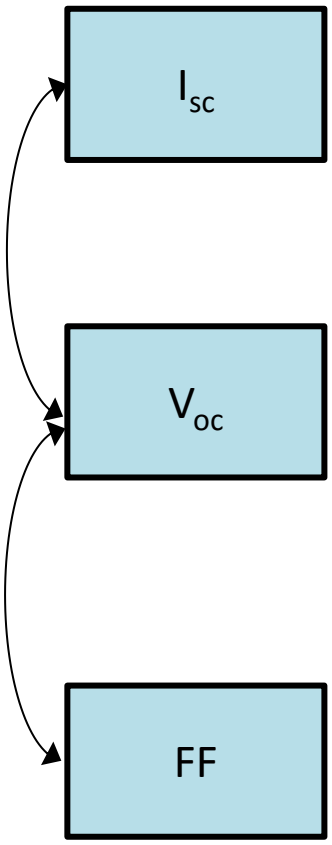


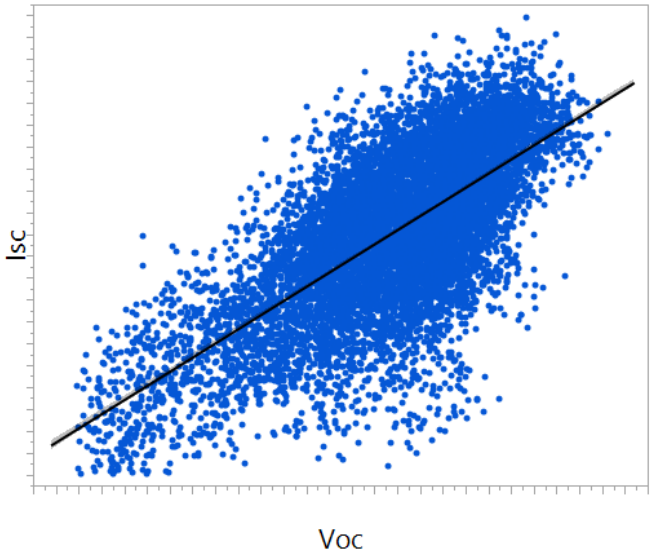
	Isc	Voc
Isc	1.0000	0.6950
Voc	0.6950	1.0000

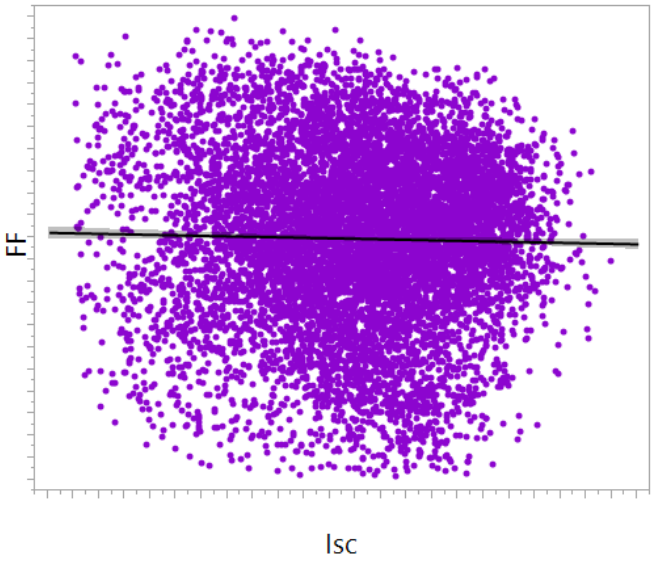


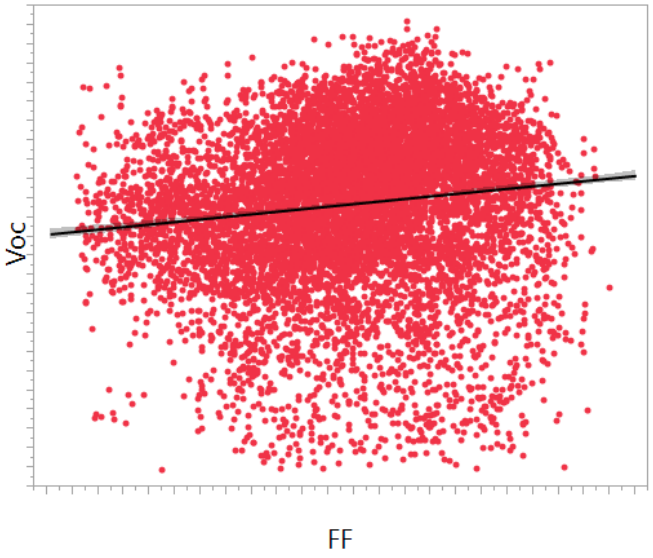
	Voc	FF
Voc	1.0000	0.1259
FF	0.1259	1.0000

3. An Introduction to Path Models



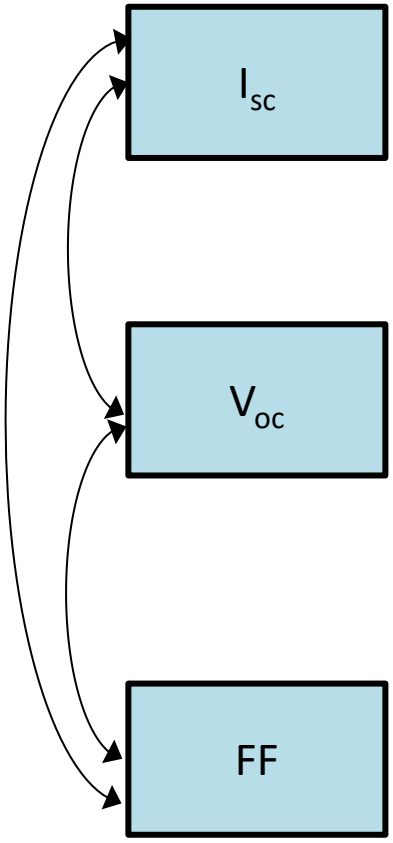






	Isc	Voc	FF
Isc	1.0000	0.6950	-0.0222
Voc	0.6950	1.0000	0.1259
FF	-0.0222	0.1259	1.0000

3. An Introduction to Path Models



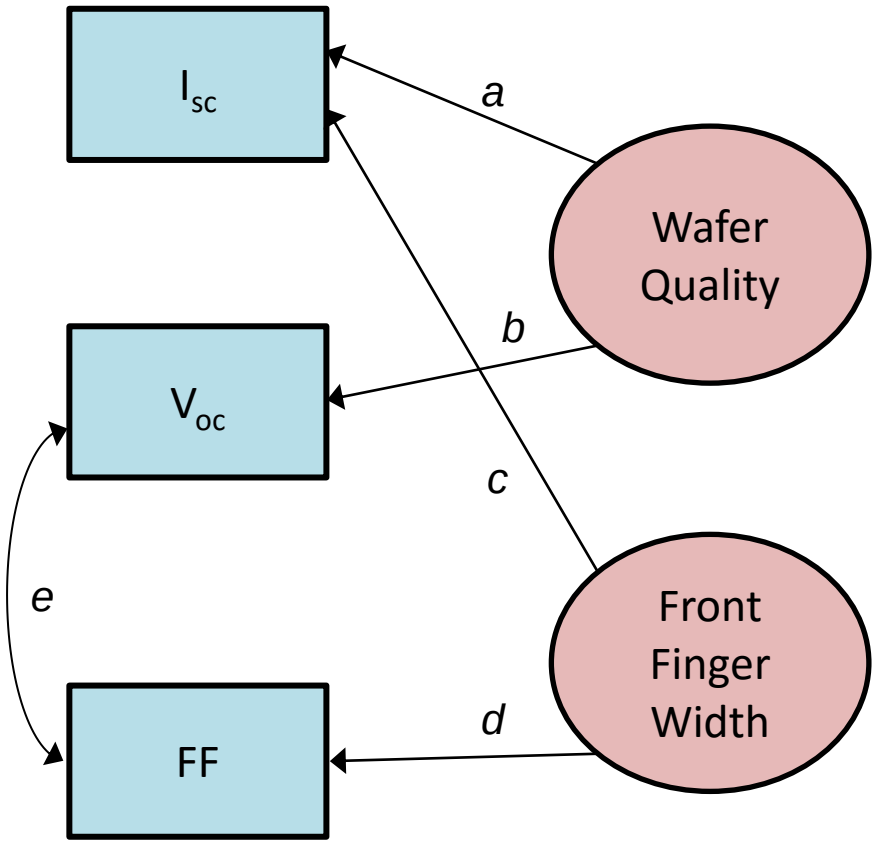
	Isc	Voc	FF
Isc	1.0000	0.6950	-0.0222
Voc	0.6950	1.0000	0.1259
FF	-0.0222	0.1259	1.0000

If there was a single root cause to these relationships, we would expect

$$r_{13} = r_{12} \times r_{23}$$

- But obviously it doesn't. The conclusion here then is there is more than one effect governing the relationship between these three variables.
  - We need another path on our diagram.

3. An Introduction to Path Models



	Isc	Voc	FF
Isc	1.0000	0.6950	-0.0222
Voc	0.6950	1.0000	0.1259
FF	-0.0222	0.1259	1.0000

- We can use the path model nomenclature to resolve this by introducing these root causes as **"latent variables"**
- A latent variable is a variable that we don't directly measure, but which is ***implied*** by the relationships between other variables

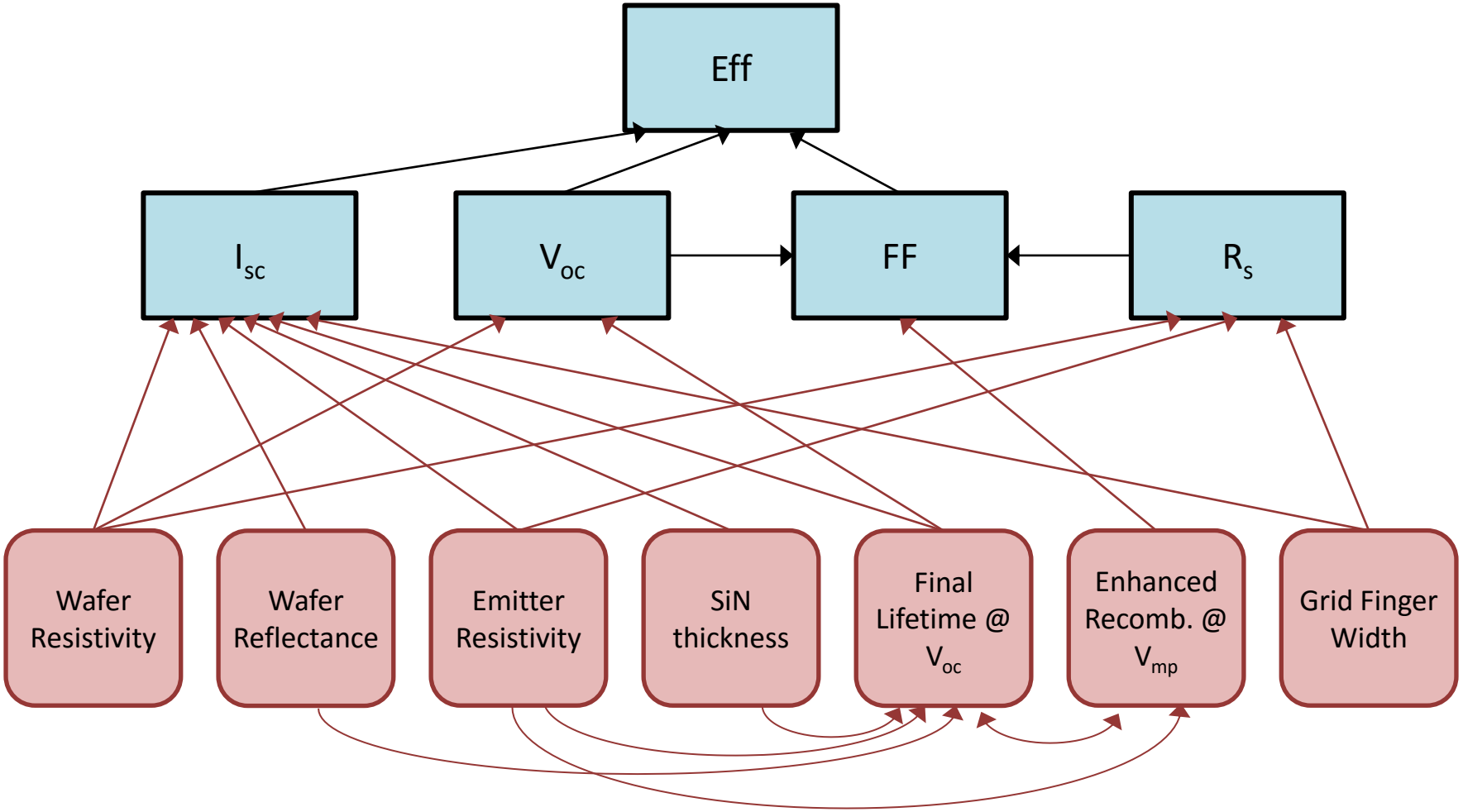
$$r_{12} = ab + ecd$$
$$r_{13} = cd + eab$$
$$r_{23} = e$$

3. An Introduction to Path Models

- Are you convinced?
- How do we actually ***know*** what the latent variables are?
- The limit to which you know is entirely determined by how well the path model captures the variance.
- There are several techniques we can use to help with this
  - Build a more complete model as a first step
  - Solve the model on multiple data sets and check how if performs
  - Use fully joined datasets to check the models
  - Improve the techniques for calculating the correlations

3. An Introduction to Path Models

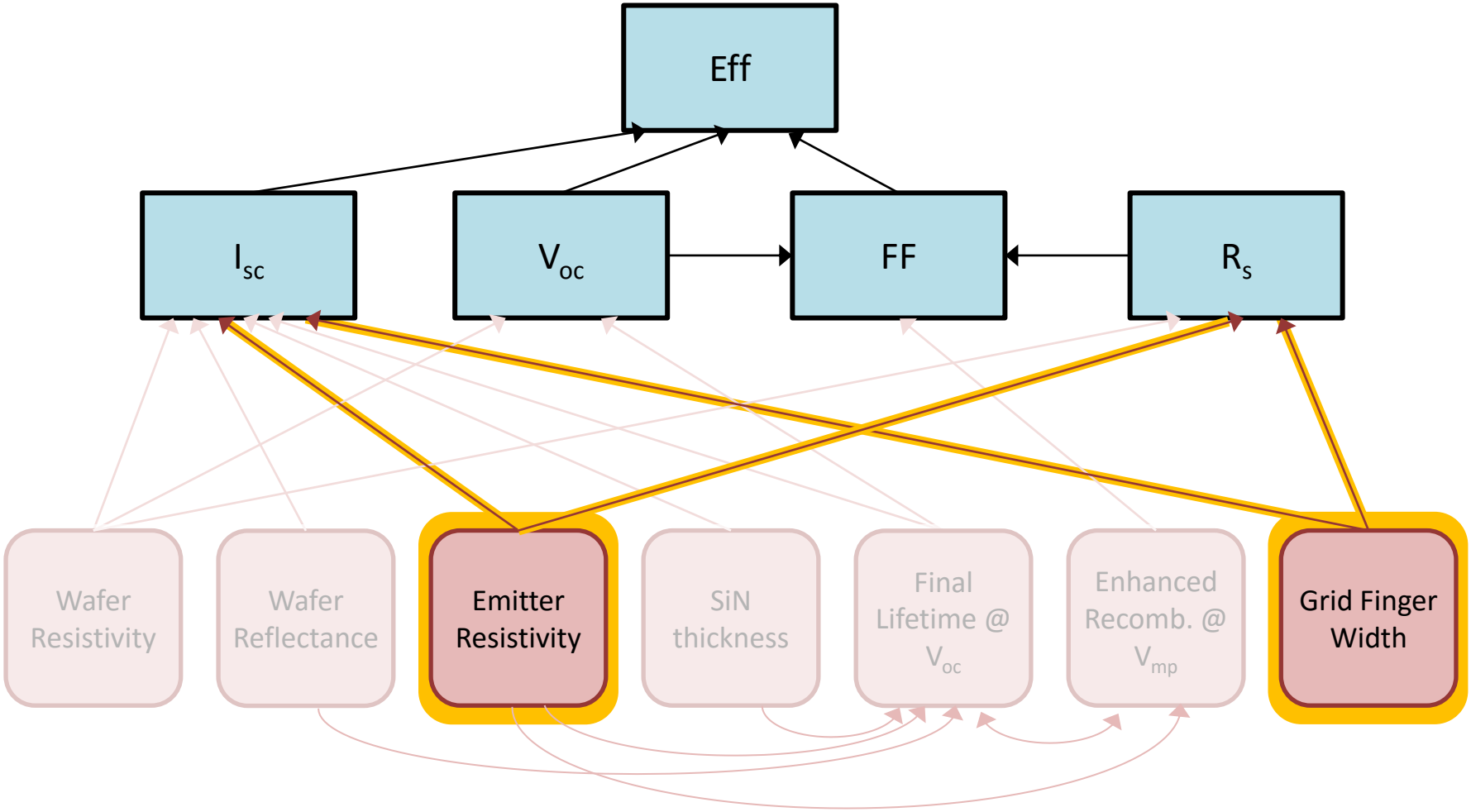
- Lets start again by building a more complete path model.



- Note the “rounded square” concept for these causal variables. These are sometimes latent and sometimes measured.

3. An Introduction to Path Models

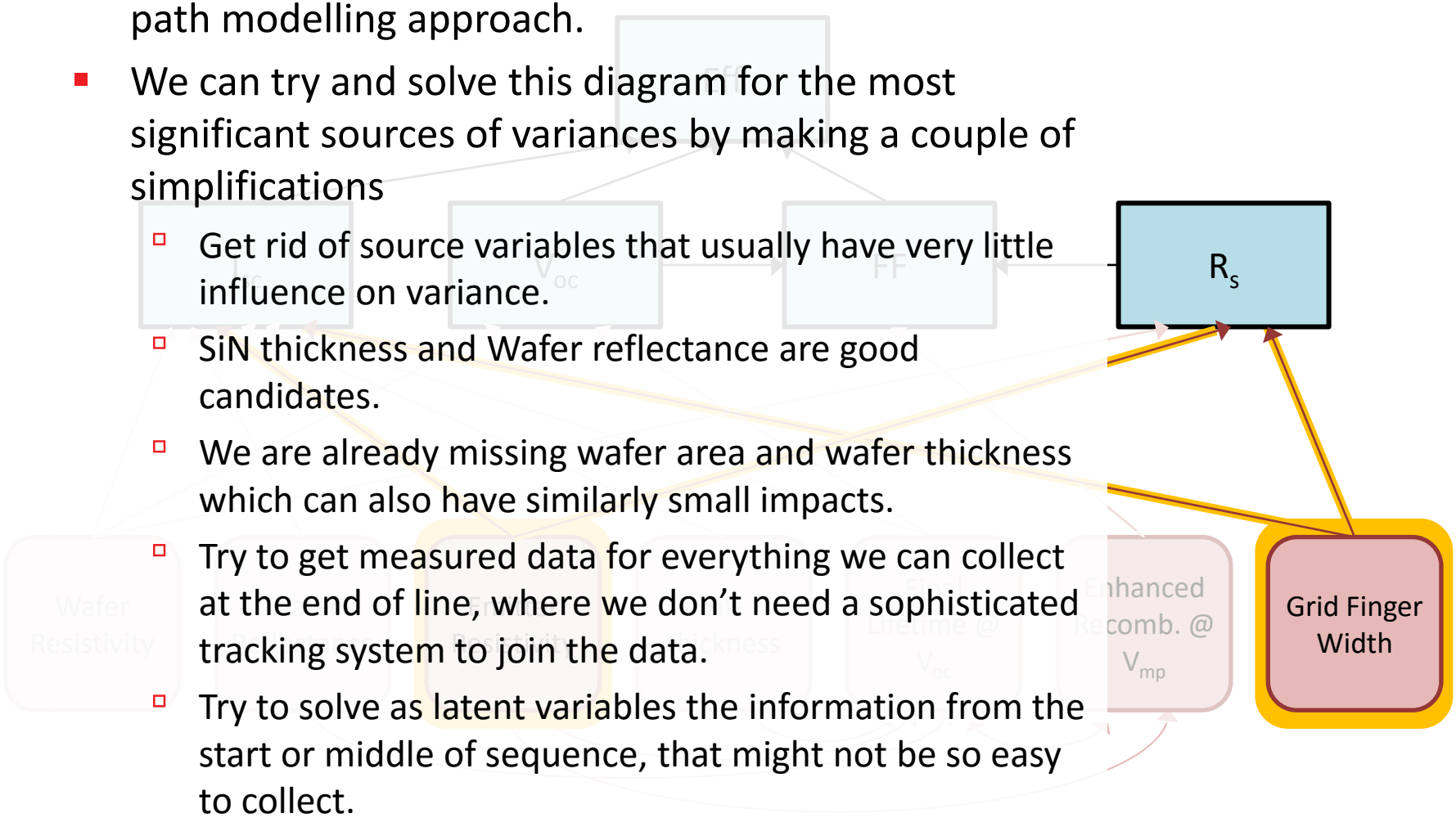
- Review the usefulness of the initial path model examined



- How do we separate these two causes that act similarly in the path model?

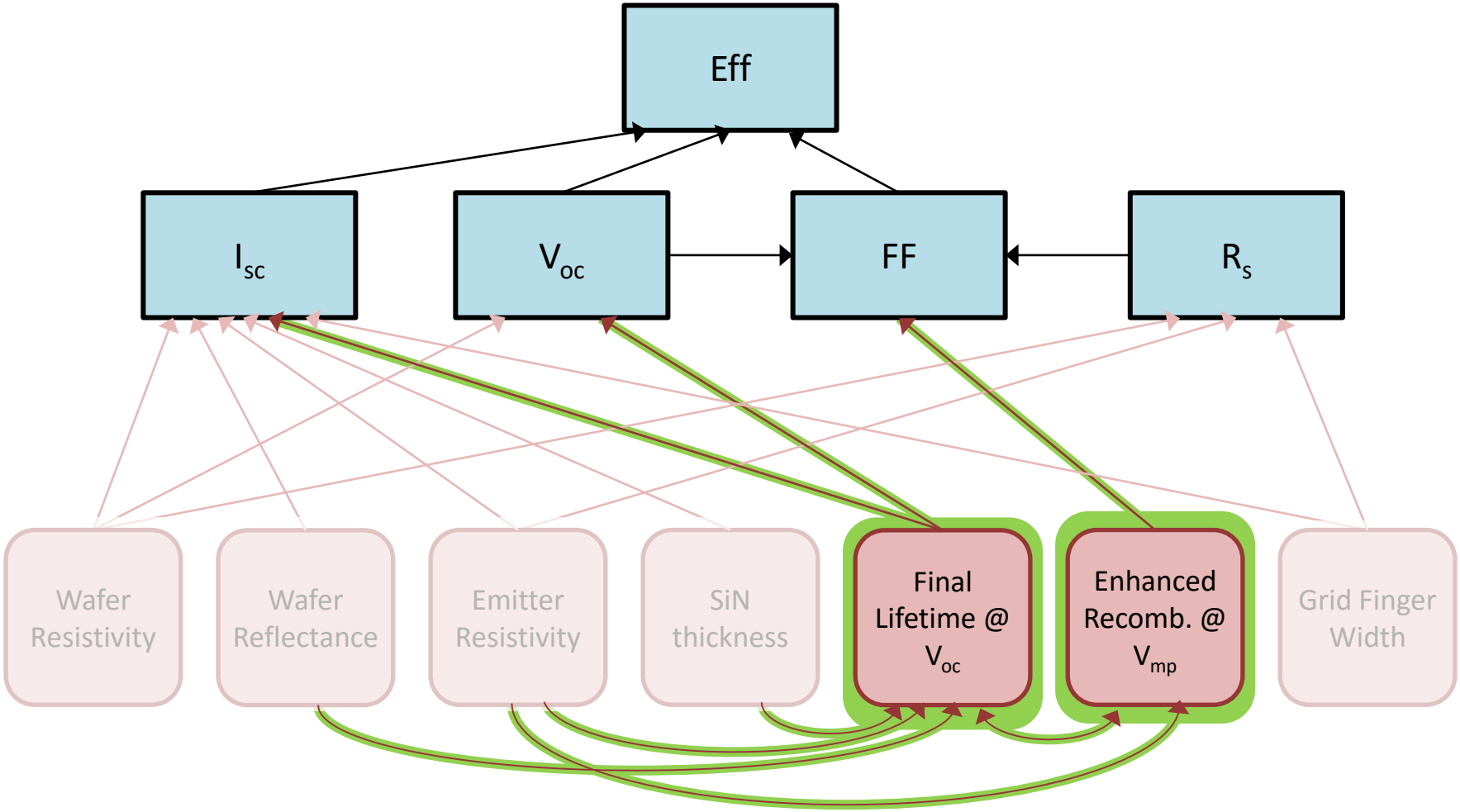
3. An Introduction to Path Models

- We can't easily make this separation using just the path modelling approach.
- We can try and solve this diagram for the most significant sources of variances by making a couple of simplifications
  - Get rid of source variables that usually have very little influence on variance.
  - SiN thickness and Wafer reflectance are good candidates.
  - We are already missing wafer area and wafer thickness which can also have similarly small impacts.
  - Try to get measured data for everything we can collect at the end of line, where we don't need a sophisticated tracking system to join the data.
  - Try to solve as latent variables the information from the start or middle of sequence, that might not be so easy to collect.



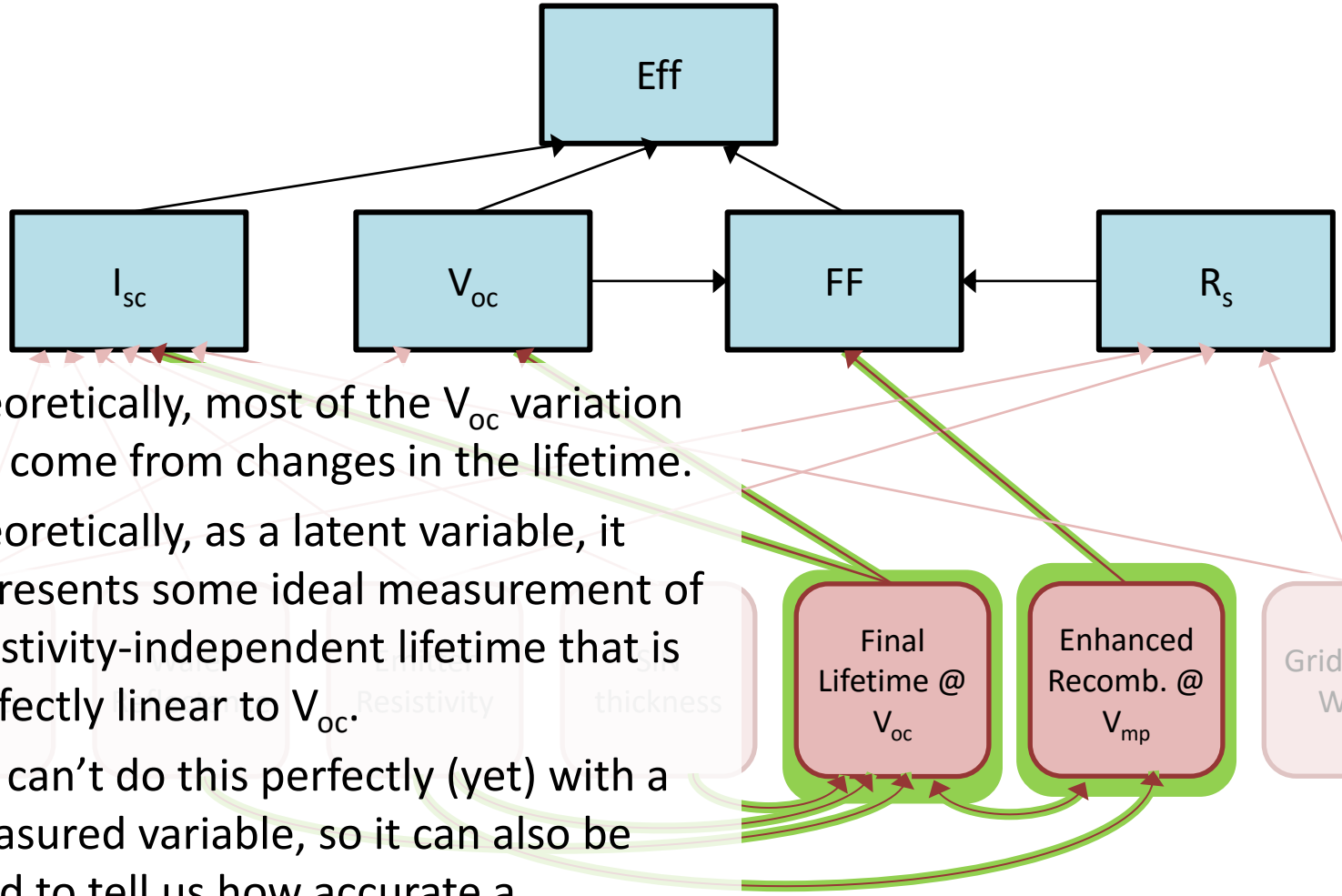
3. An Introduction to Path Models

- The “lifetime” parameters are also interesting to think about



3. An Introduction to Path Models

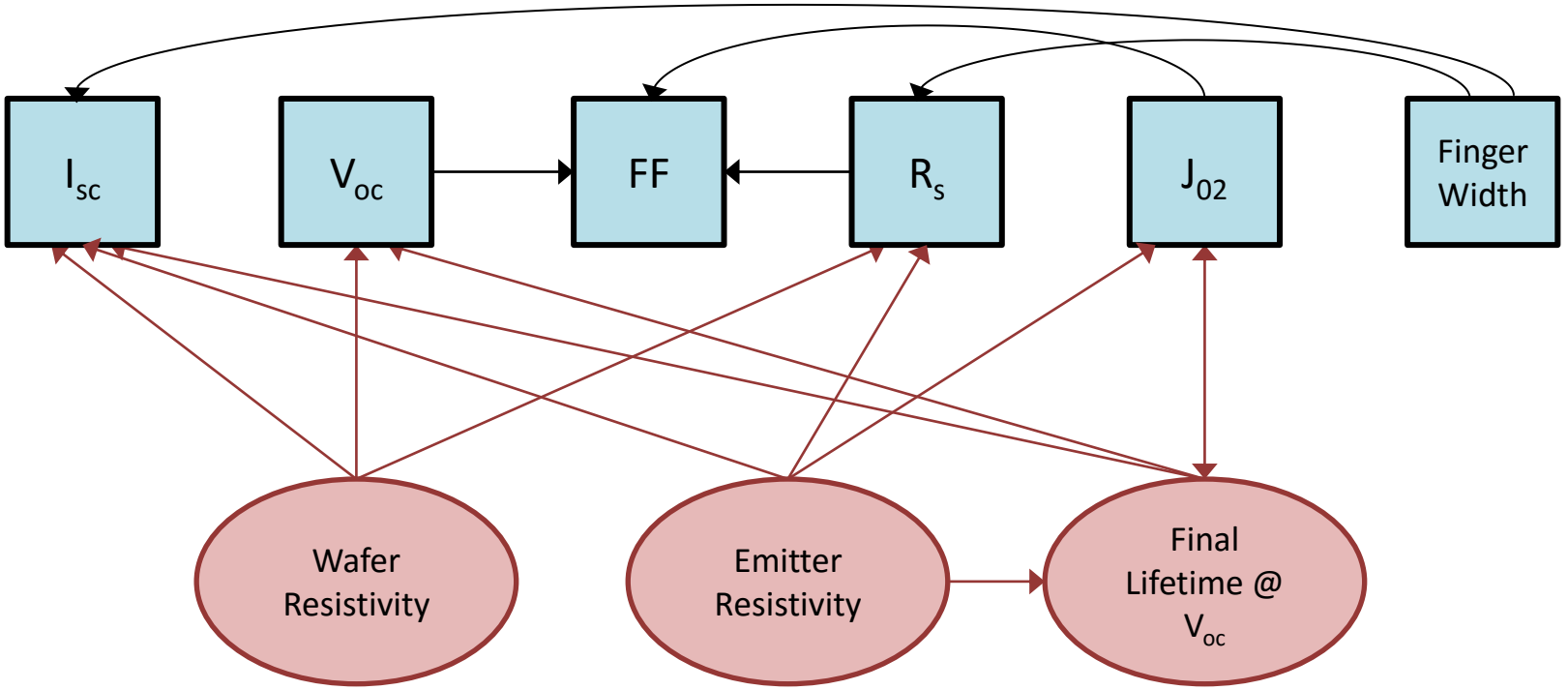
- The “lifetime” parameters are also interesting to think about



- Theoretically, most of the V_{oc} variation will come from changes in the lifetime.
- Theoretically, as a latent variable, it represents some ideal measurement of resistivity-independent lifetime that is perfectly linear to V_{oc}.
- We can’t do this perfectly (yet) with a measured variable, so it can also be used to tell us how accurate a measured variable is

3. An Introduction to Path Models

- Lets try now to build a path model that is tractable using common end-of-line parameters – to learn about the influence of parameters we don't (or can't) measure end-of-line.
  - These are shown now as latent variables (in circles)

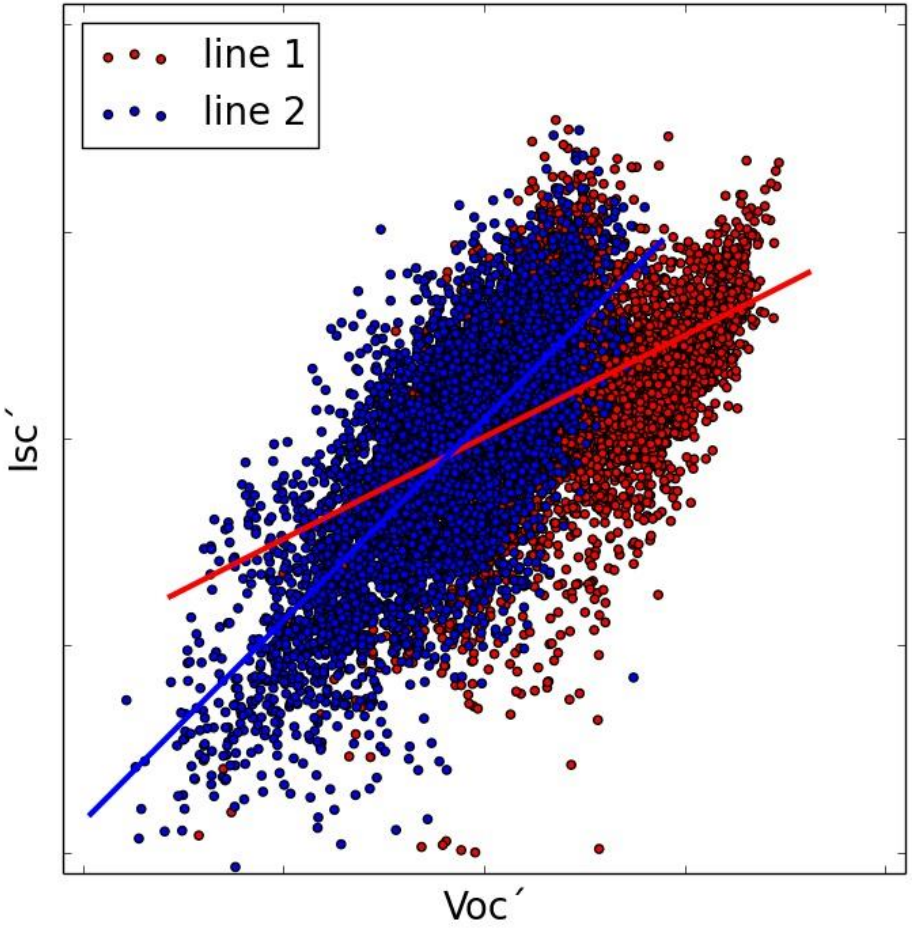


- A path model such as this is tractable, with a few assumptions, but the solutions from the correlation matrix can be unstable.

3. An Introduction to Path Models

- A path model such as this is tractable, with a few assumptions, but the solutions from the correlation matrix can be unstable.

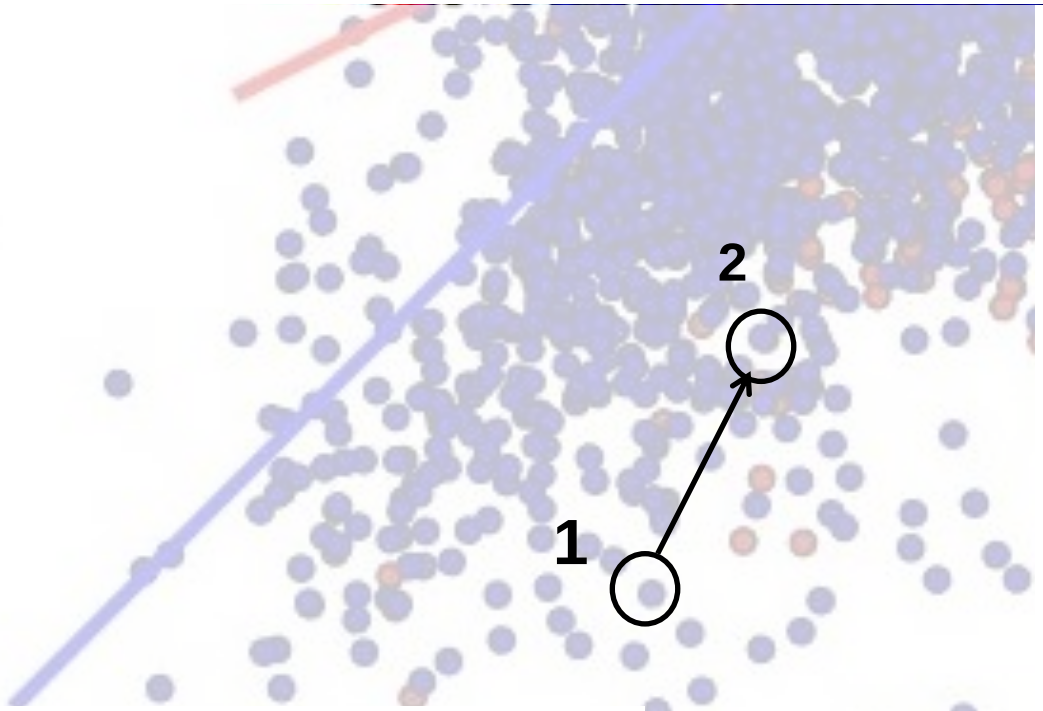
- Conventionally, to solve the path diagram, we label the paths and solve using the correlation matrix
- But due to autocorrelation effects in the data, the correlation matrix can be unreliable for representing the relationships between the variables



3. An Introduction to Path Models

- We can find some alternate ways of expressing this relationship

1. Consider some point on this relationship in time
2. Look at what changes to make the next cell



Express this change as a

vector $\overrightarrow{vi}$,

with

length $\left|\overrightarrow{vi}\right|$

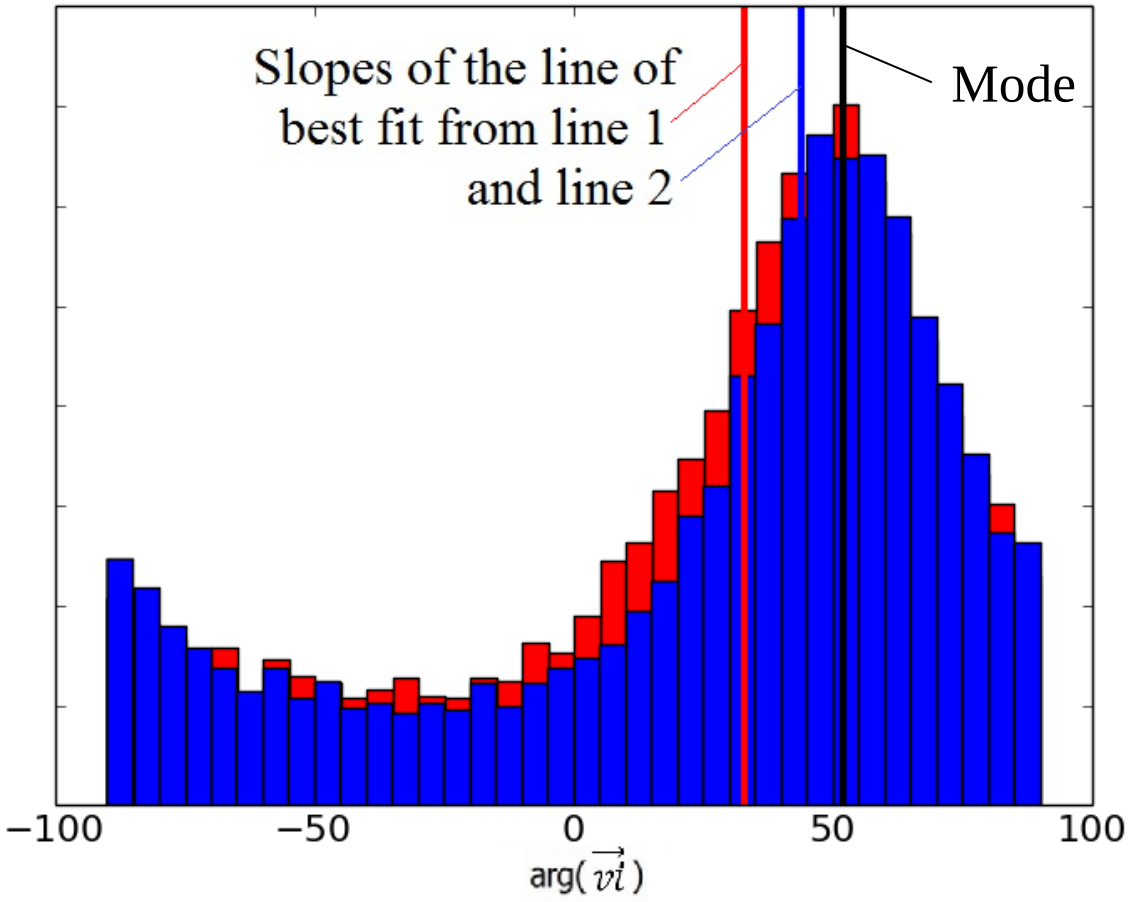
and

direction $\arg\left(\overrightarrow{vi}\right)$

Note: Angle is calculated as $\arg(vi) = tan^{-1}(\Delta I/\Delta V)$ and so exactly opposite directions are represented by the same angle.

3. An Introduction to Path Models

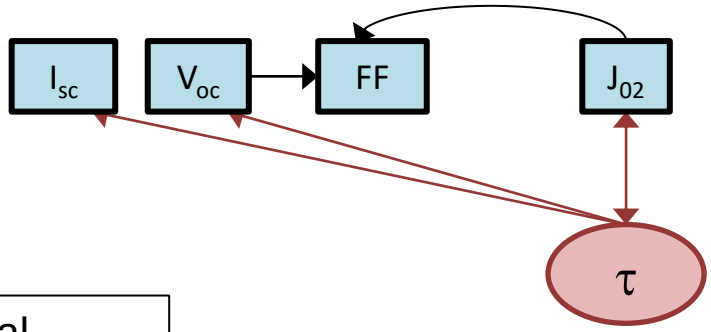
- Plot the histogram of the direction angle of this vector
- The modal response is the same in both cases. This is a more useful interpretation of the relationship between the variables.
- We can do this pairwise for all the relationships in the path model



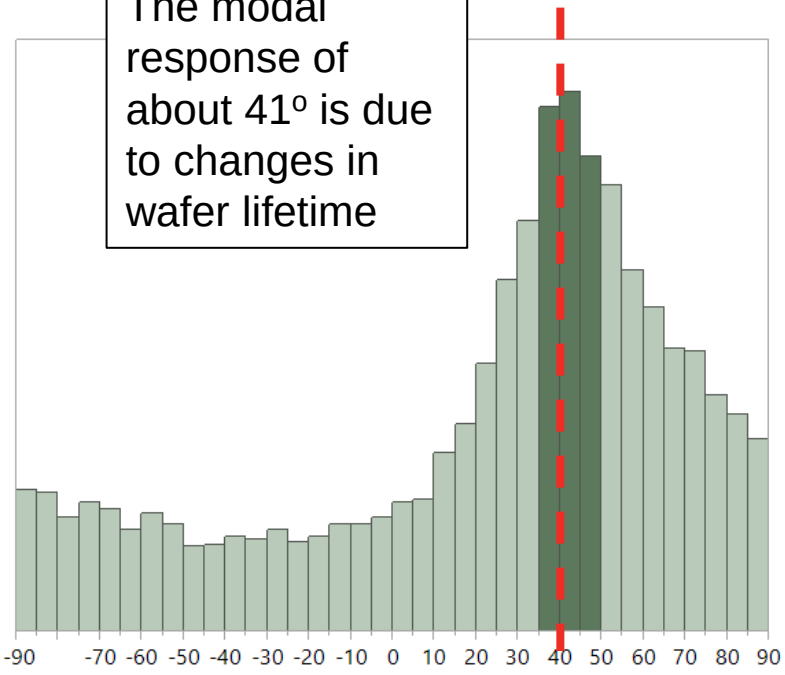
Contents

4. Path Model Solutions

- Work through the data pairwise.
- Start with **Isc and Voc**. The dominant interaction here is the **lifetime** of the wafer.

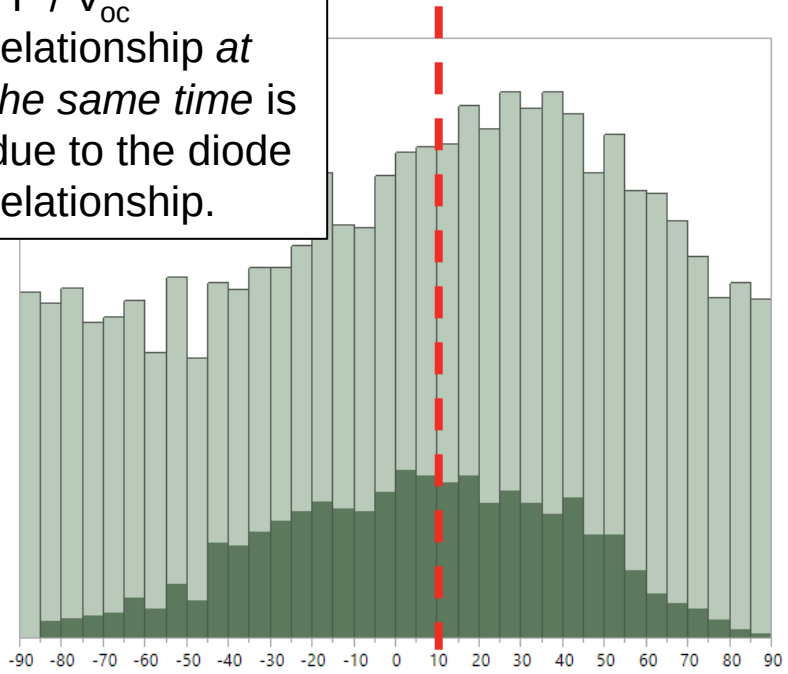


The modal response of about 41° is due to changes in wafer lifetime



$tan^{-1}(\Delta I_{sc}/\Delta V_{oc})$

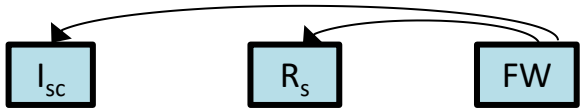
The modal response in the FF / V_{oc} relationship *at the same time* is due to the diode relationship.



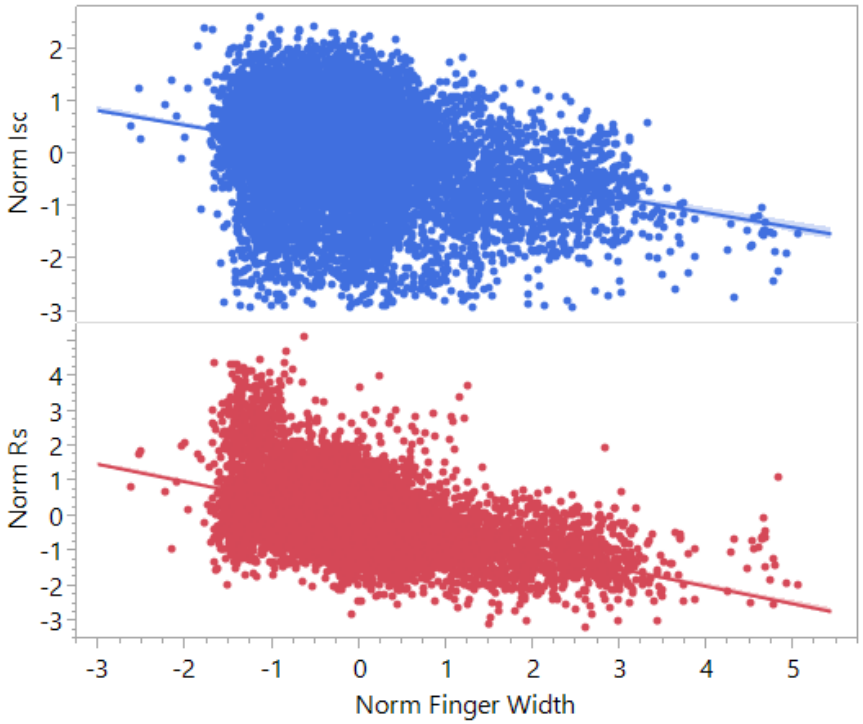
$tan^{-1}(\Delta FF/\Delta V_{oc})$

4. Path Model Solutions

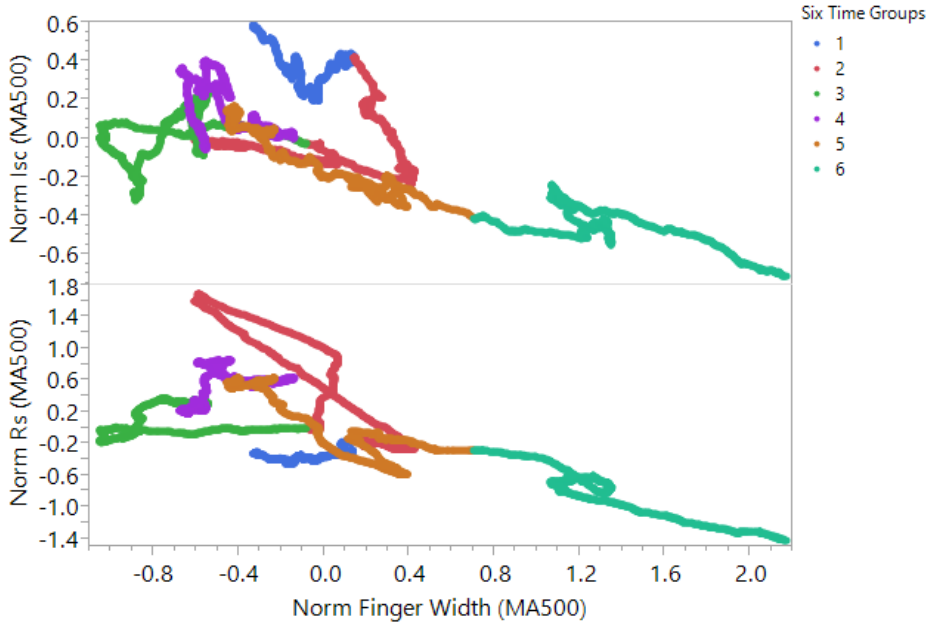
- Next consider the **grid finger width (FW)**



- The overall relationship between these three parameters is best found from a LARGE set of data due to noise in the measured metrics
  - The differencing method does not work so well.



Normalised Isc and Rs against grid finger width for 10 000 cells

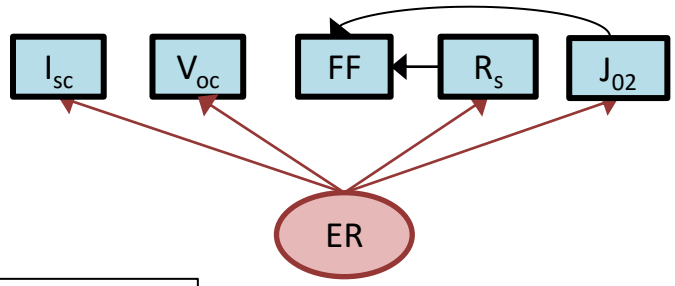


Moving Average or normalised Isc and Rs against grid finger width showing how relationship changes over time

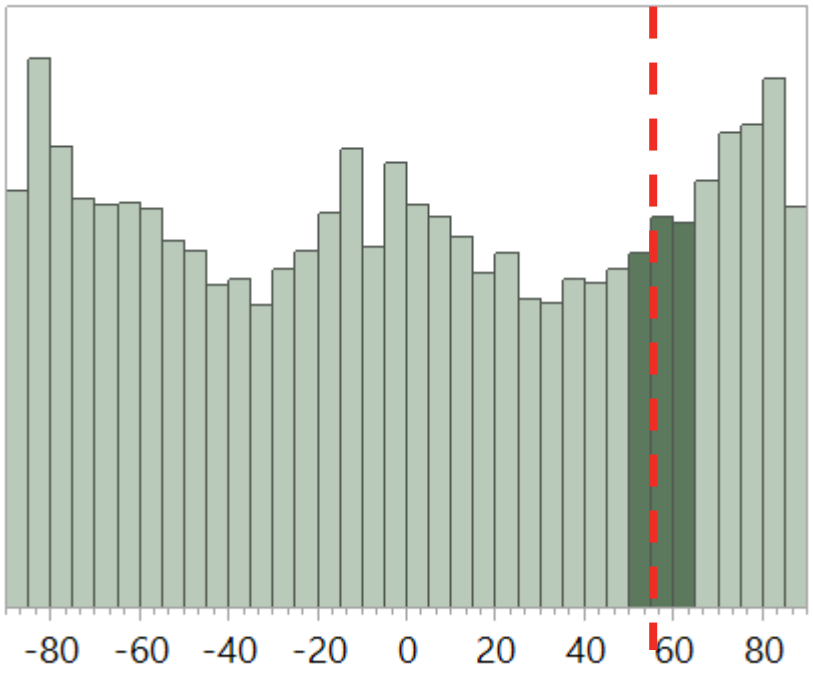
4. Path Model Solutions

- Next is the **Emitter Resistivity (ER)**
- Looking for a trends where

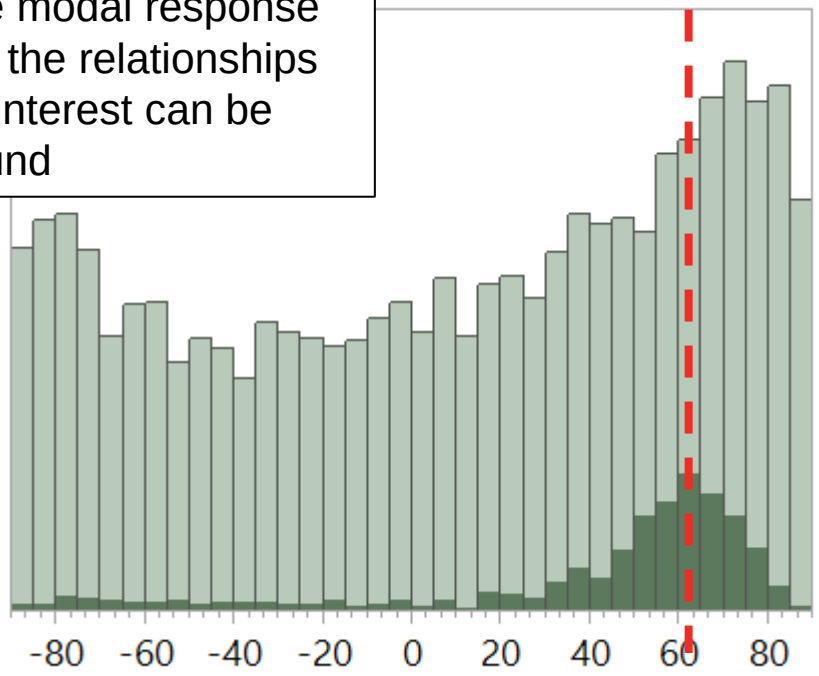
$V_{oc}\uparrow, I_{sc}\uparrow, R_s\uparrow$



With some searching, the modal response for the relationships of interest can be found



$tan^{-1}(\Delta V_{oc}/\Delta R_s)$

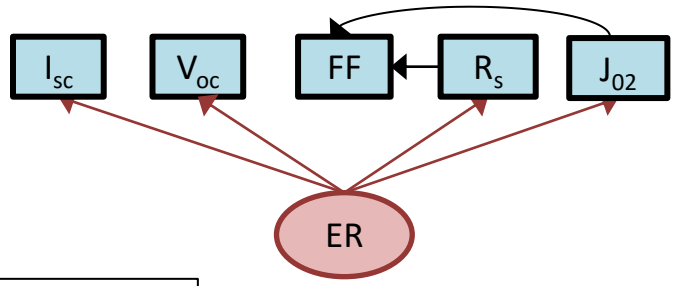


$tan^{-1}(\Delta I_{sc}/\Delta R_s)$

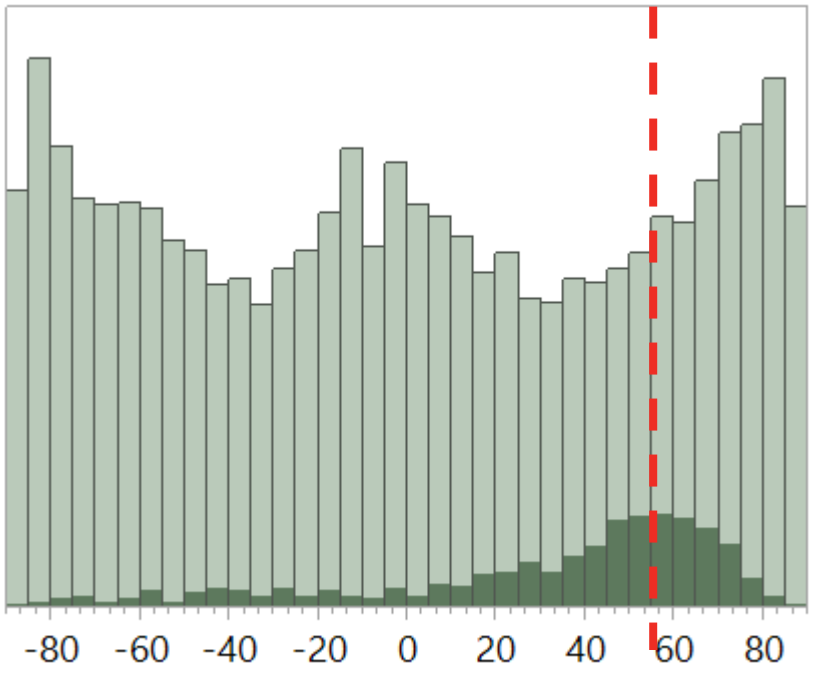
4. Path Model Solutions

- Next is the **Emitter Resistivity (ER)**
- Looking for a trends where

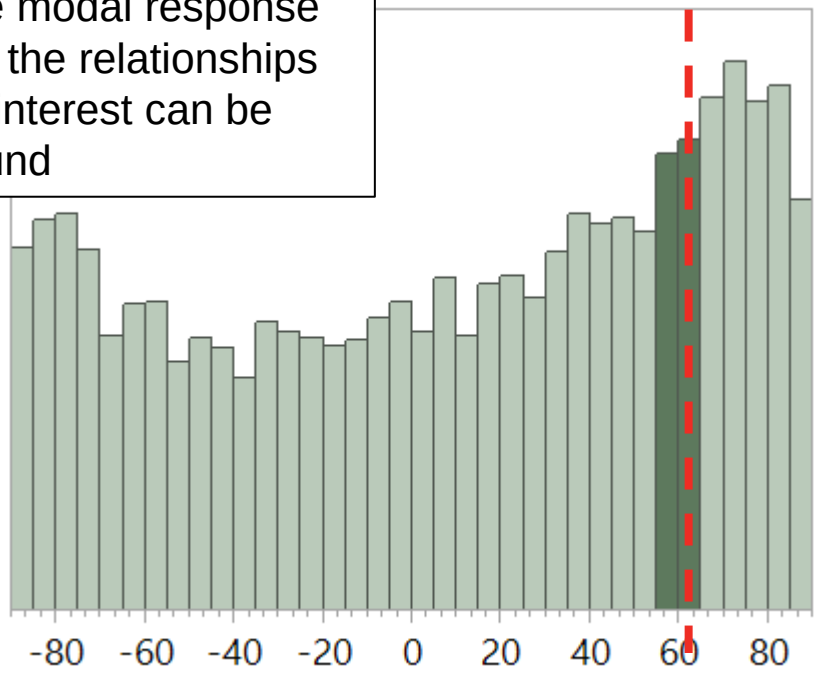
$V_{oc}\uparrow, I_{sc}\uparrow, R_s\uparrow$



With some searching, the modal response for the relationships of interest can be found



$tan^{-1}(\Delta V_{oc}/\Delta R_s)$

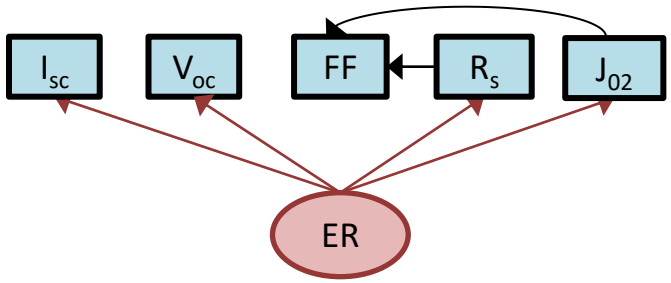


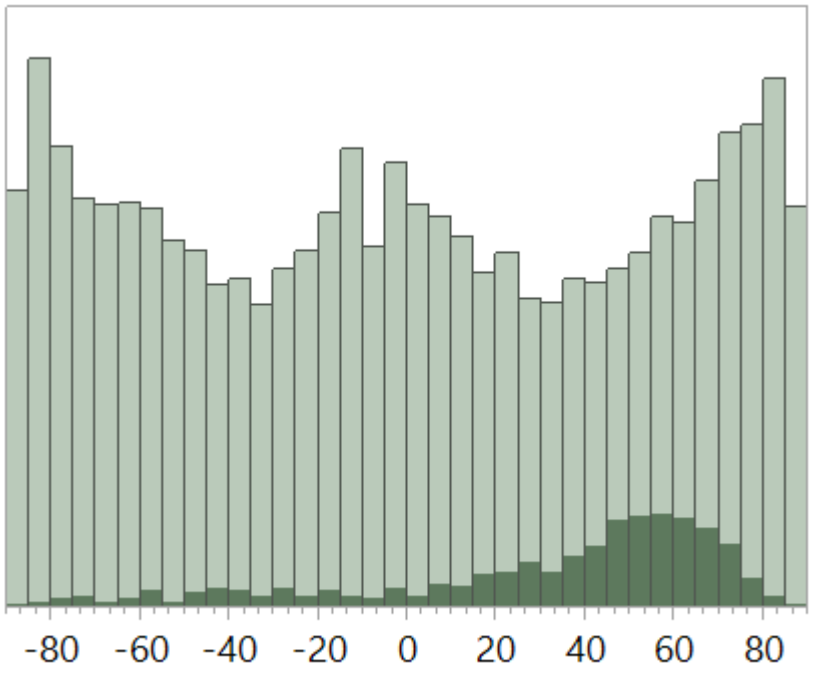
$tan^{-1}(\Delta I_{sc}/\Delta R_s)$

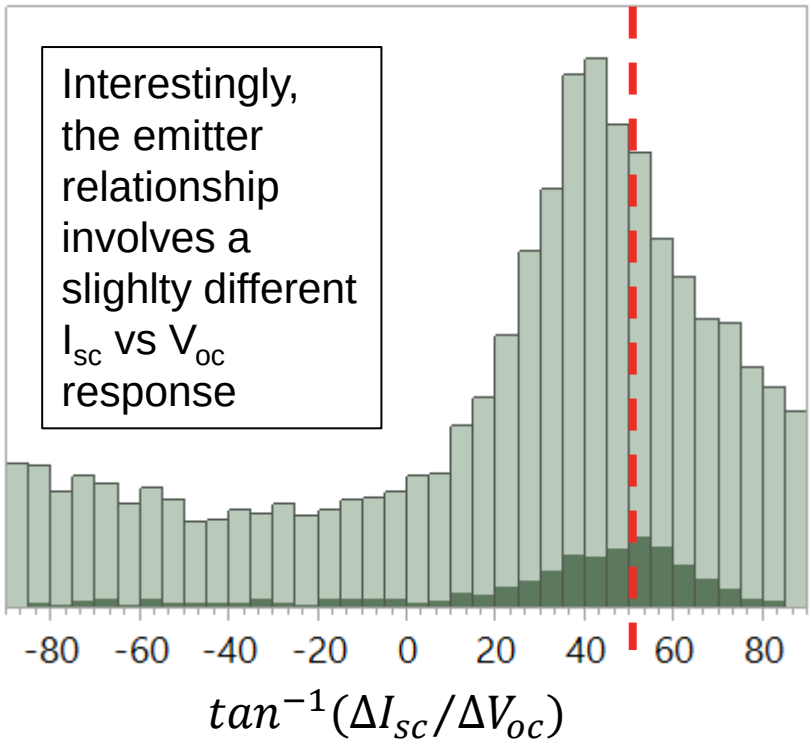
4. Path Model Solutions

- Next is the **Emitter Resistivity (ER)**
- Looking for a trends where

$V_{oc}\uparrow, I_{sc}\uparrow, R_s\uparrow$





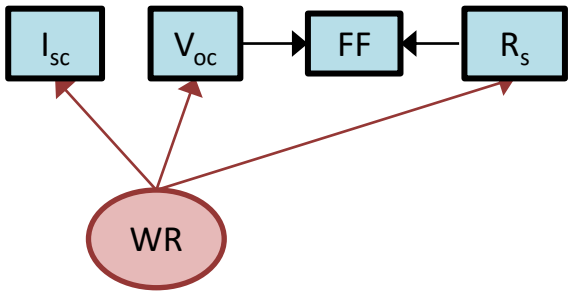


- There is no strong enhanced recombination effect detectable

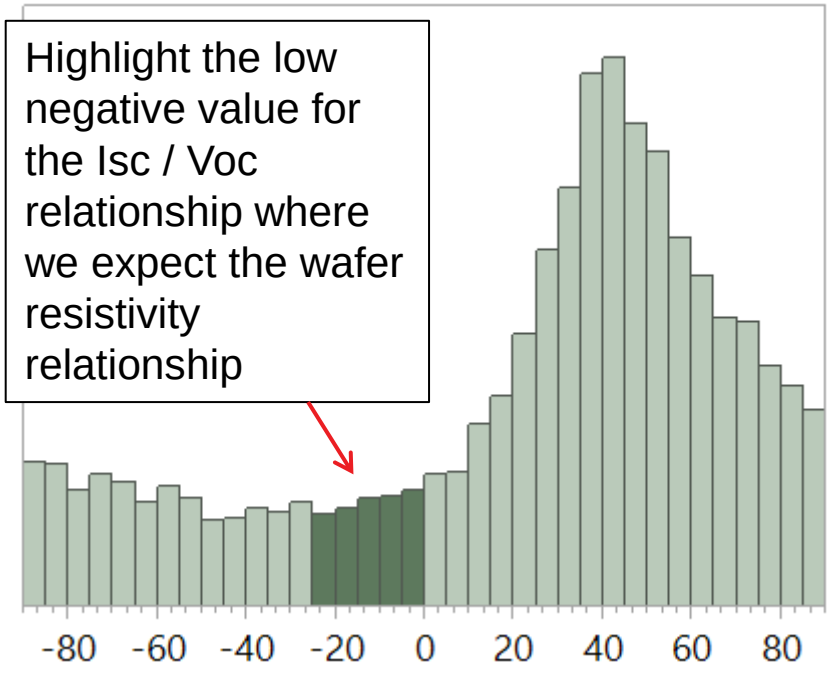
4. Path Model Solutions

- Next is the **Wafer Resistivity (WR)**
- Looking for a trends where

$V_{oc}\uparrow, I_{sc}\downarrow, R_s\downarrow$

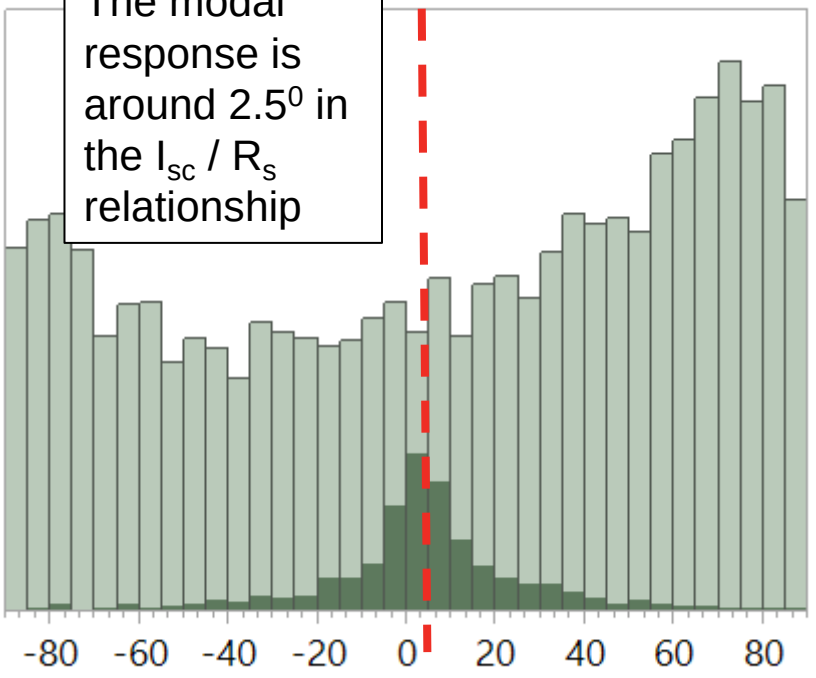


Highlight the low negative value for the Isc / Voc relationship where we expect the wafer resistivity relationship



$tan^{-1}(\Delta I_{sc}/\Delta V_{oc})$

The modal response is around 2.5^0 in the I_{sc} / R_s relationship

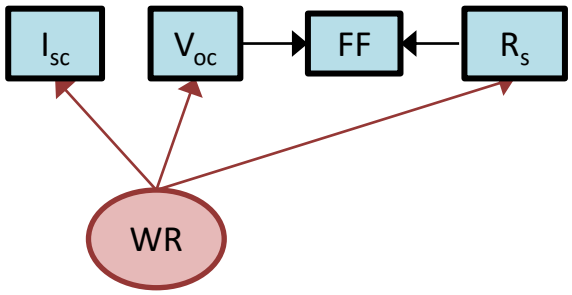


$tan^{-1}(\Delta I_{sc}/\Delta R_s)$

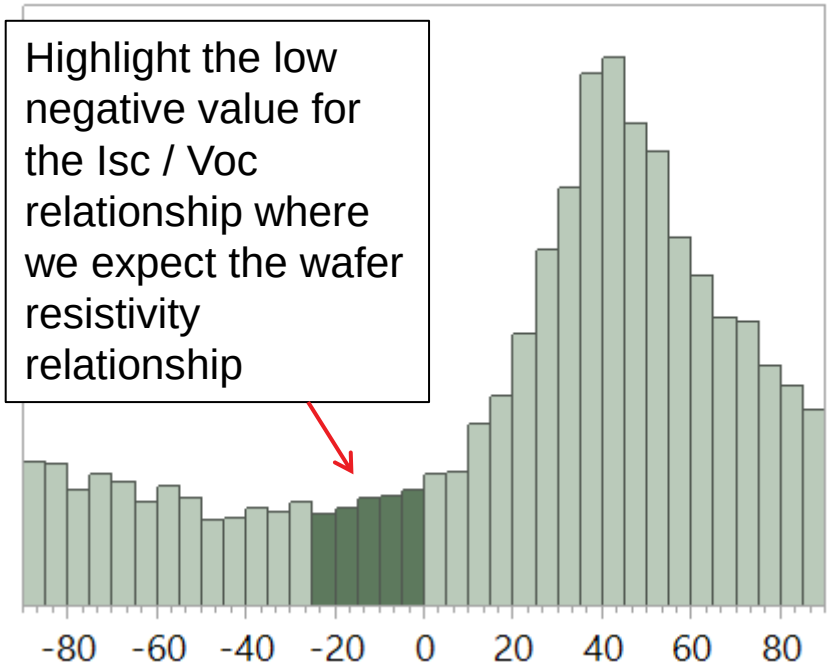
4. Path Model Solutions

- Next is the **Wafer Resistivity (WR)**
- Looking for a trends where

$V_{oc}\uparrow, I_{sc}\downarrow, R_s\downarrow$

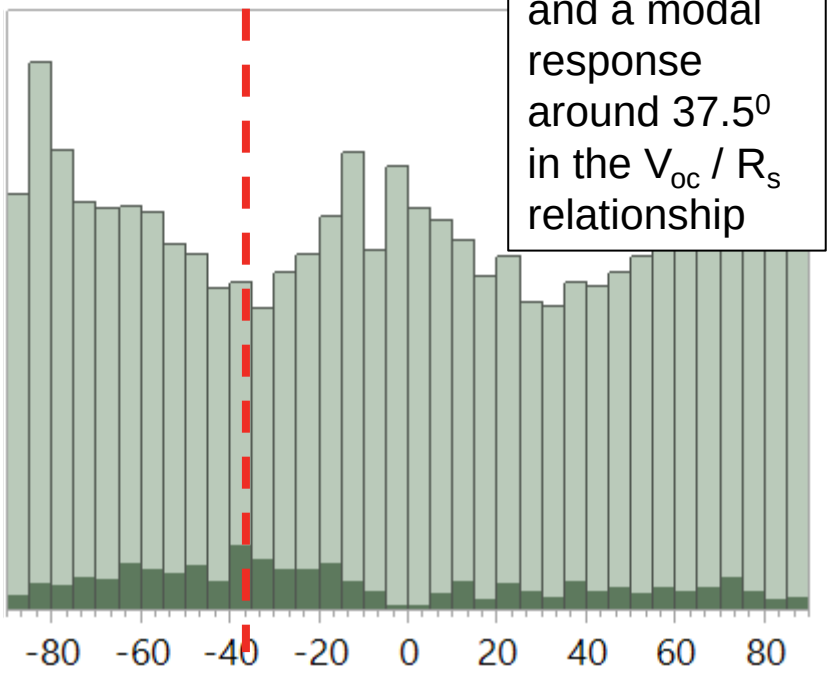


Highlight the low negative value for the Isc / Voc relationship where we expect the wafer resistivity relationship



$tan^{-1}(\Delta I_{sc}/\Delta V_{oc})$

and a modal response around 37.5^0 in the V_{oc} / R_s relationship

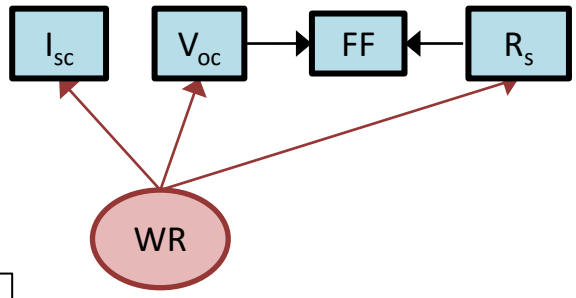


$tan^{-1}(\Delta V_{oc}/\Delta R_s)$

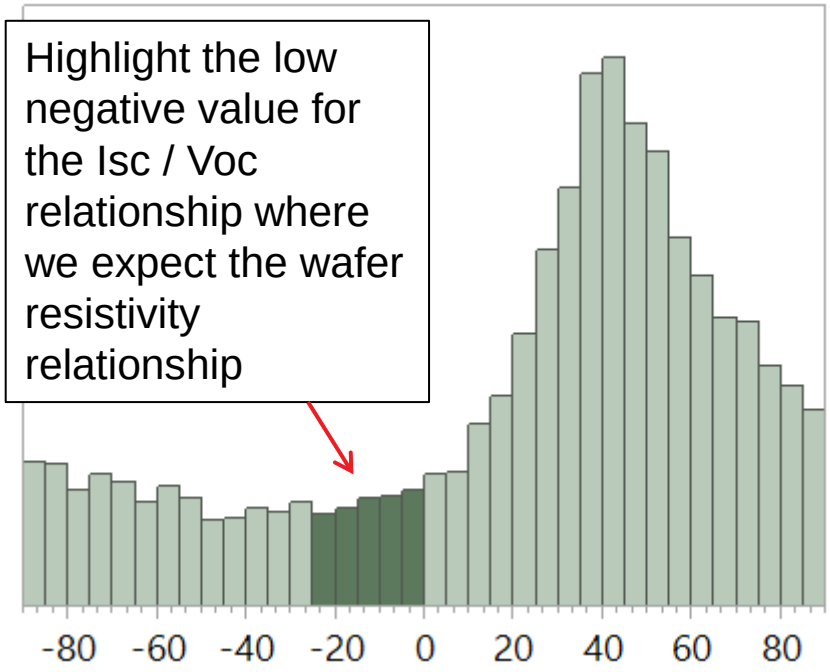
4. Path Model Solutions

- Next is the **Wafer Resistivity (WR)**
- Looking for a trends where

$V_{oc}\uparrow, I_{sc}\downarrow, R_s\downarrow$

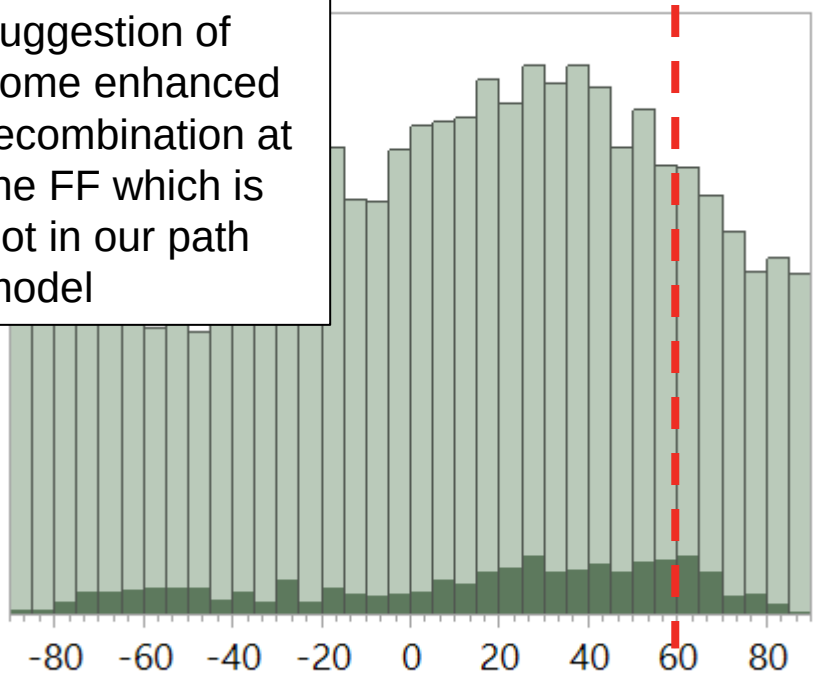


Highlight the low negative value for the Isc / Voc relationship where we expect the wafer resistivity relationship



$tan^{-1}(\Delta I_{sc}/\Delta V_{oc})$

There is also a suggestion of some enhanced recombination at the FF which is not in our path model



$tan^{-1}(\Delta FF/\Delta V_{oc})$

4. Path Model Solutions

- Once these relationships are all known, they can directly be used to calculate the components of variance.
  - I've spared you the maths, but it mostly involves data rotation and projection
- These variance components can be used as
  - A simple and very sensitive indicator of consistency and hence quality in production, across shifts or days or week or lines.
  - A targeted approach to improving variance
  - As a way to define process capability in a way that relates to overall variance targets, rather than on an ad-hoc basis.
- The vectorial dataset also contains some highly detailed information about underlying noise / variance in the measurement techniques (beyond the scope of this presentation)

- In the case of this dataset, the veracity of the techniques can even be checked, because the dataset contains actual measured data that attempts to represent these latent variables.

4. Path Model Solutions

% of Variance in -	Due to	Latent Variable Path Model	Linear Regression on Measured Data
Isc	Lifetime	62%	22%
	Emitter Resistivity	18%	3.5%
	Wafer Resistivity	1.1%	1.2%
	Finger Width	7.7%	7.4%
Voc	Lifetime	86%	36%
	Emitter Resistivity	0.8%	0.2%
	Wafer Resistivity	13%	1.5%
	Finger Width	n/a	n/a
Rs	Lifetime	n/a	n/a
	Emitter Resistivity	24%	2.6%
	Wafer Resistivity	4.4%	8.2%
	Finger Width	25%	24%

- How good are these solutions
  - They are as good as our sum total of knowledge about all the interactions.
  - We should know these as completely as possible. This often requires a detailed offline variance analysis in the way company's would do a detailed loss analysis.

4. Path Model Solutions

- Which one is correct?
  - Both have advantages and disadvantages.
  - Direct measurements require wafer level tracking and are subject to error
  - Latent variables depend on some assumptions and require thorough knowledge of interactions.
- The measured data confirms most of the underlying relationships from the path model, but the variance in this data means different results are found for the components of variance.

- The latent variable approach allows us to look for underlying relationships and this is of value regardless of how you assign cause.
  - Doing this over time on a production line will provide information on the consistency and quality of the production, and also ways to improve and tailor the models.

Contents

5. Knowledge is Power

- Latent variables don't invalidate the need for inline metrology, but
  - They can help you choose the best ones,
  - They can help you get the most out of the data you have,
  - They can help you check the accuracy / validity of measured data.

- Variance analysis is a very sensitive indicator of quality.
- Data produced during manufacturing can be used to optimise the quality of the manufacturing
  - What is the best metrology to help with this?
  - What are the most useful and cost effective measurements – what I call the "minimum data set"?
  - What set of data would constitute the best "quality function" for PV manufacturing?
  - What analytics can help us to achieve all of this?

5. Knowledge is Power

- When decisions are made in manufacturing, how do they affect power in the field?
  - There are many embedded assumptions in these relationships, not all of them are correct and not all of them are significant.
  - Can we join data sets to try to make these decisions clearer.

Next Steps

- Finish developing the multivariate approach to the relational analysis.
- Develop techniques to extract error / noise estimates from the directional data.
- Road test the algorithms.
- Finishing coding the algorithms with some sort of attractive front end dashboard and try to get manufacturers interested in using them.

- Further work needs to be done on how to interface field performance data into manufacturing decision making
  - The impact of variance on mismatch loss as a field ages
  - The impact of bankability criteria on field development