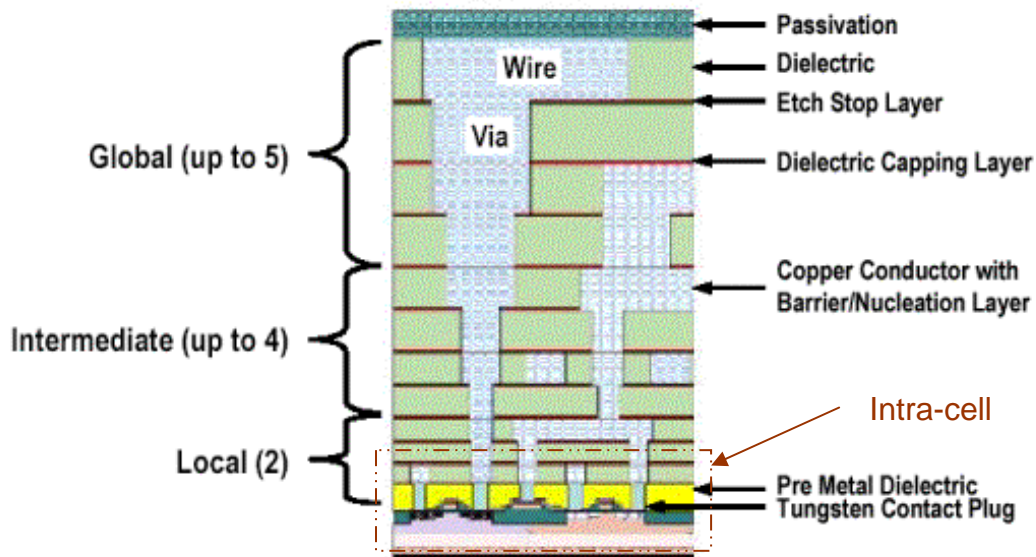

On-chip Interconnect Variation

Lou Scheffer

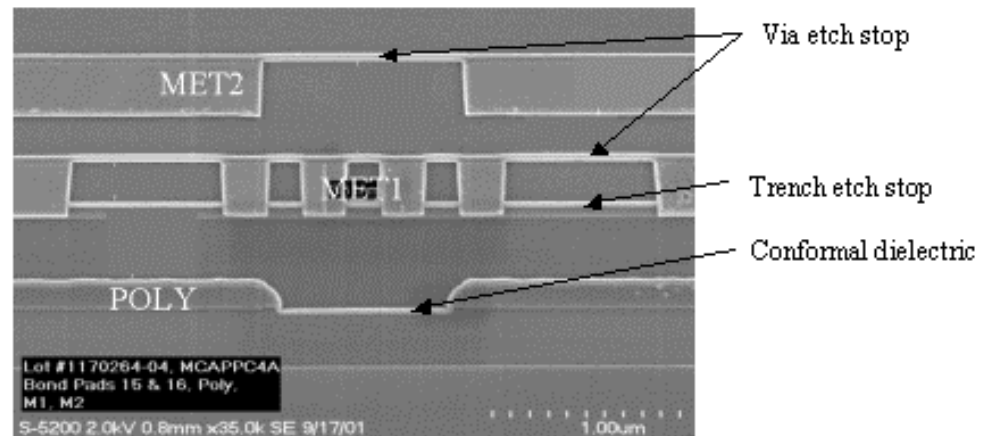
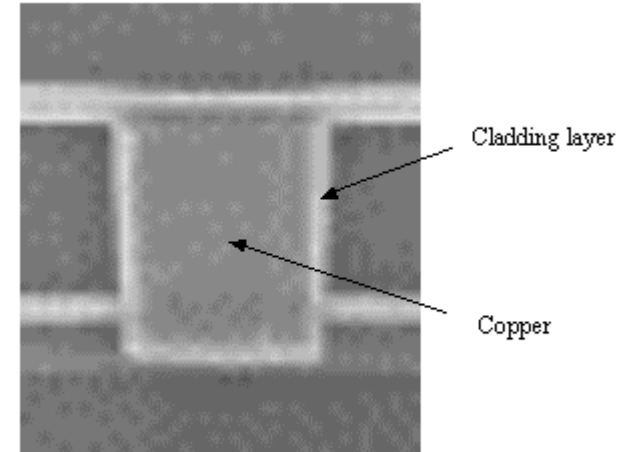
Outline

- How is interconnect constructed?
- What makes it vary from place to place?
- How do fabs currently spec what variation is possible?
- How do designers allow for on-chip variation?
- What are some possible improvements in this process?
- What are some of the obstacles to such improvement?
- Conclusions

Interconnect Stack in Copper/Low-k Technologies



Source : ITRS Roadmap



- Accurate modeling needed for
 - conductors
 - dielectrics
 - VIAs and contacts

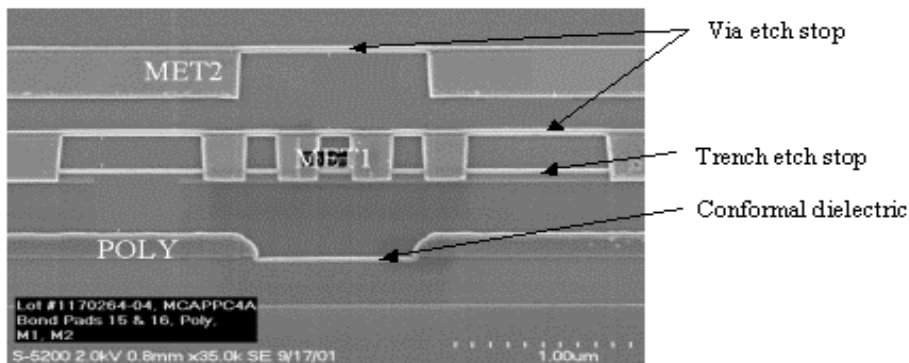
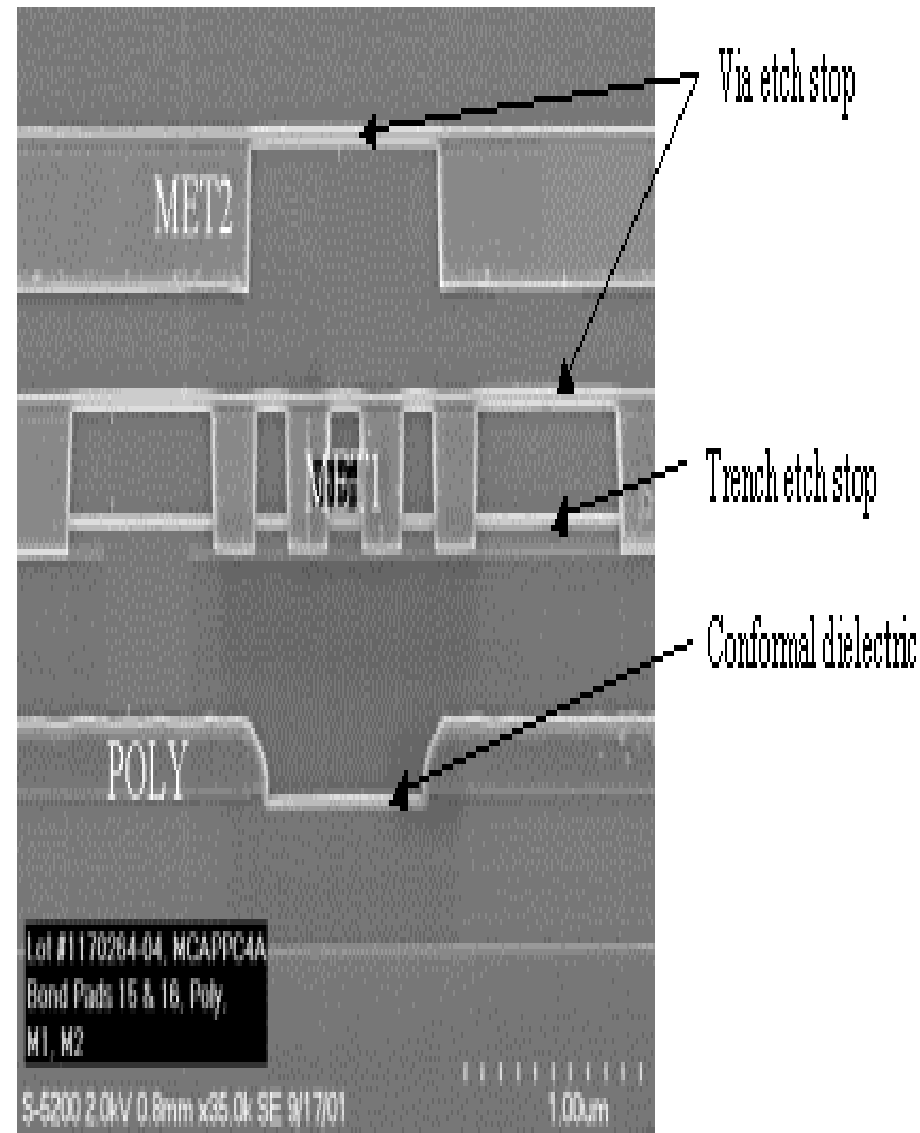
From 2005 DAC tutorial of NS Nagaraj

Interconnect Stack in Copper/Low-k Technologies

- Looks very nice, but check your design rule manual.

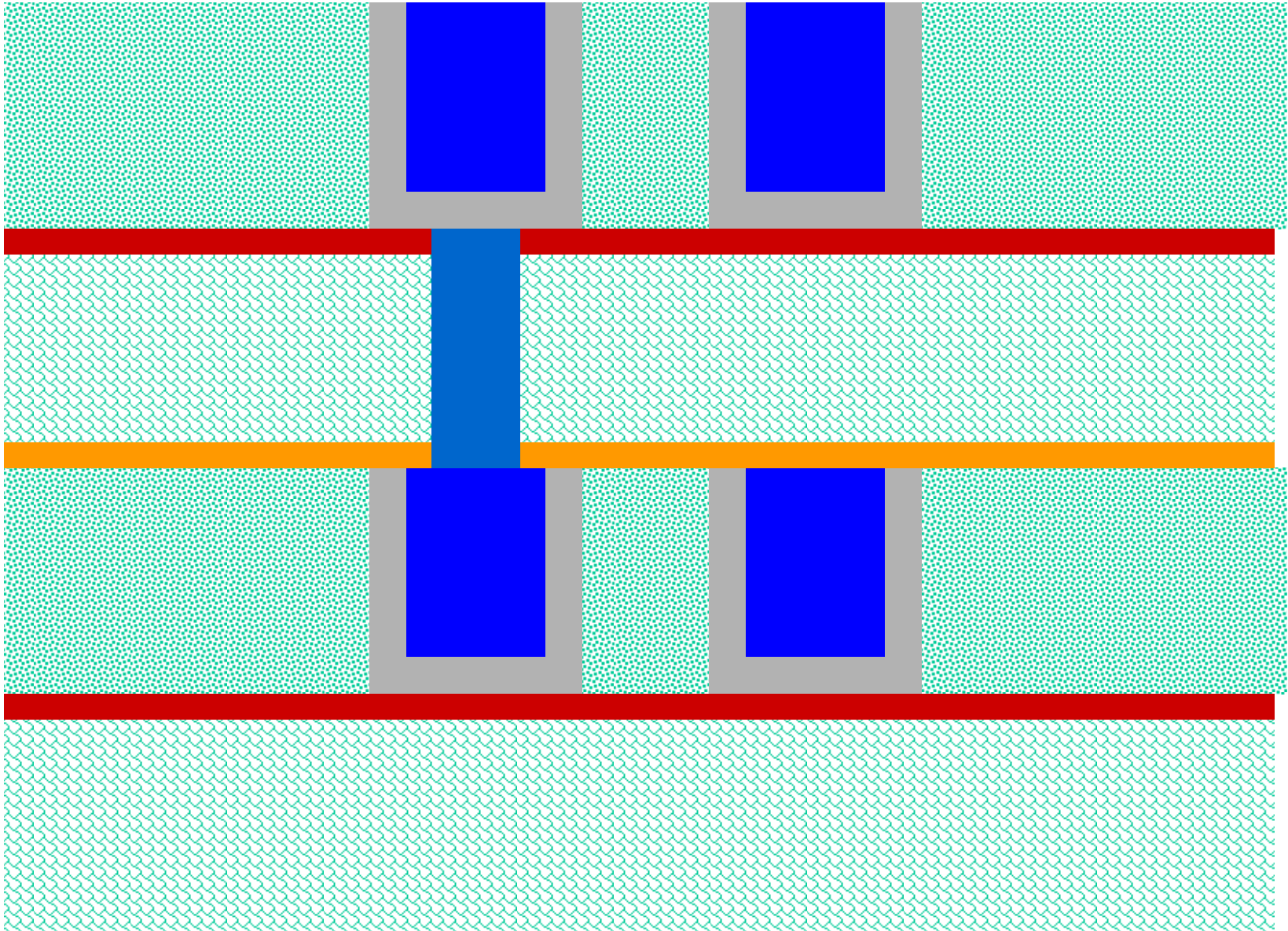
- You might get this, or this →

- According to specs, might get both, on the same chip, on adjacent wires



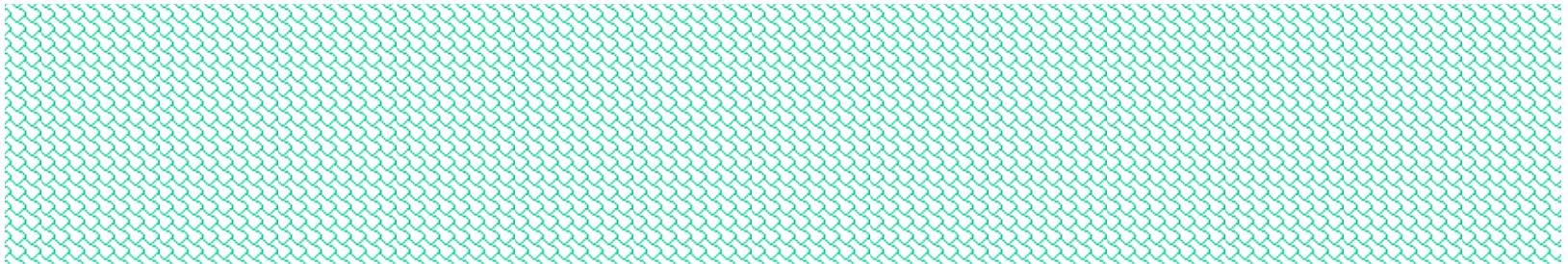
A quick model of interconnect construction

How is this multi-layer structure built?



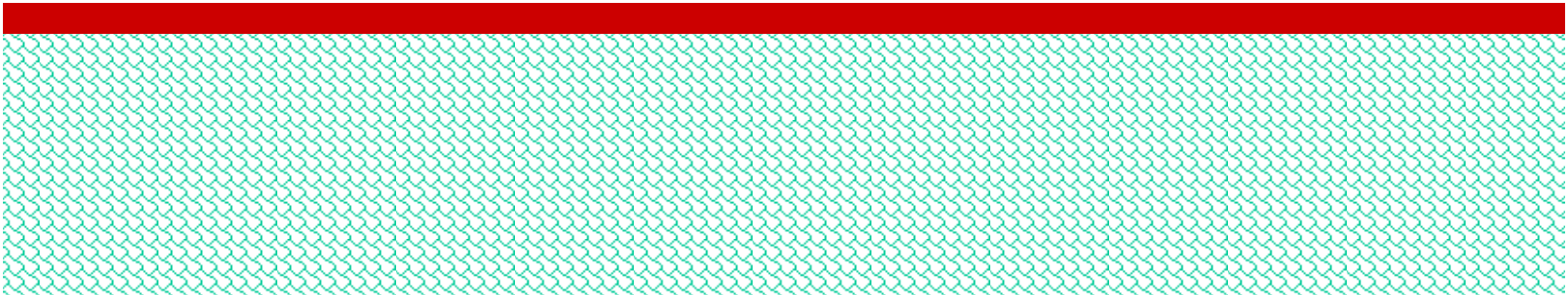
A quick model of interconnect construction

Start with the previous interlayer dielectric



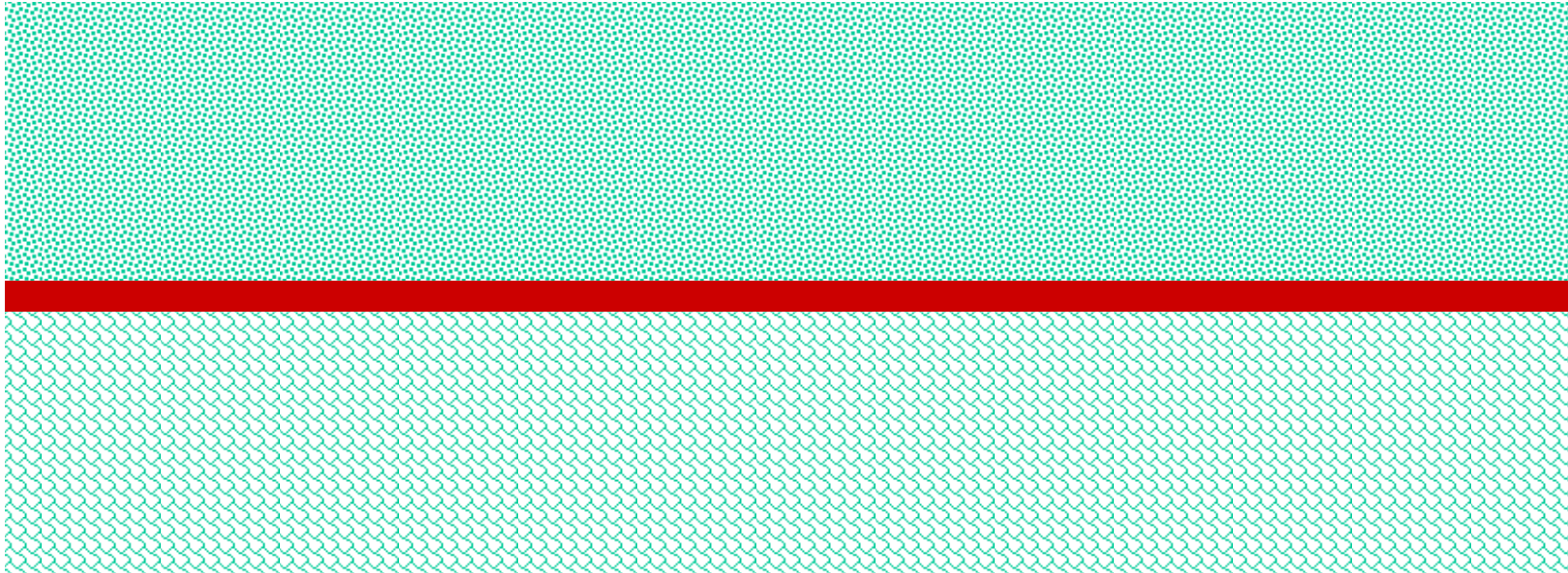
A quick model of interconnect construction

Add an 'etch stop' layer



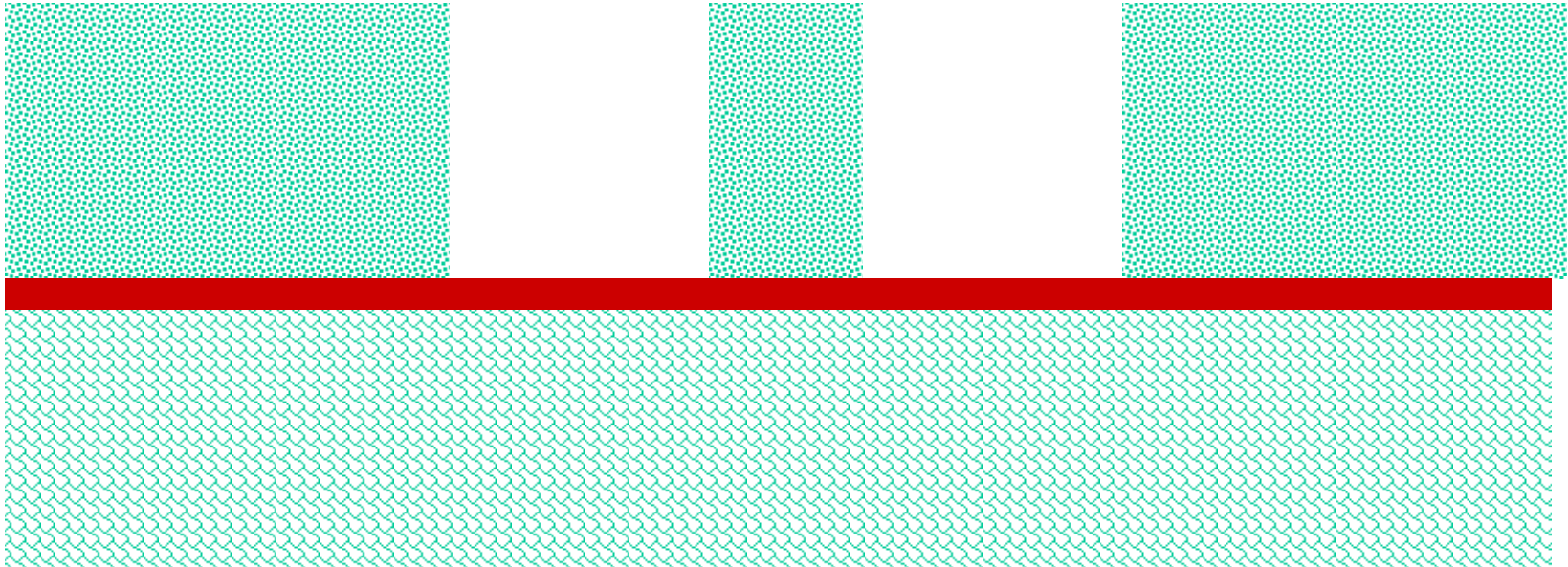
A quick model of interconnect construction

Add a new layer of intra-layer dielectric



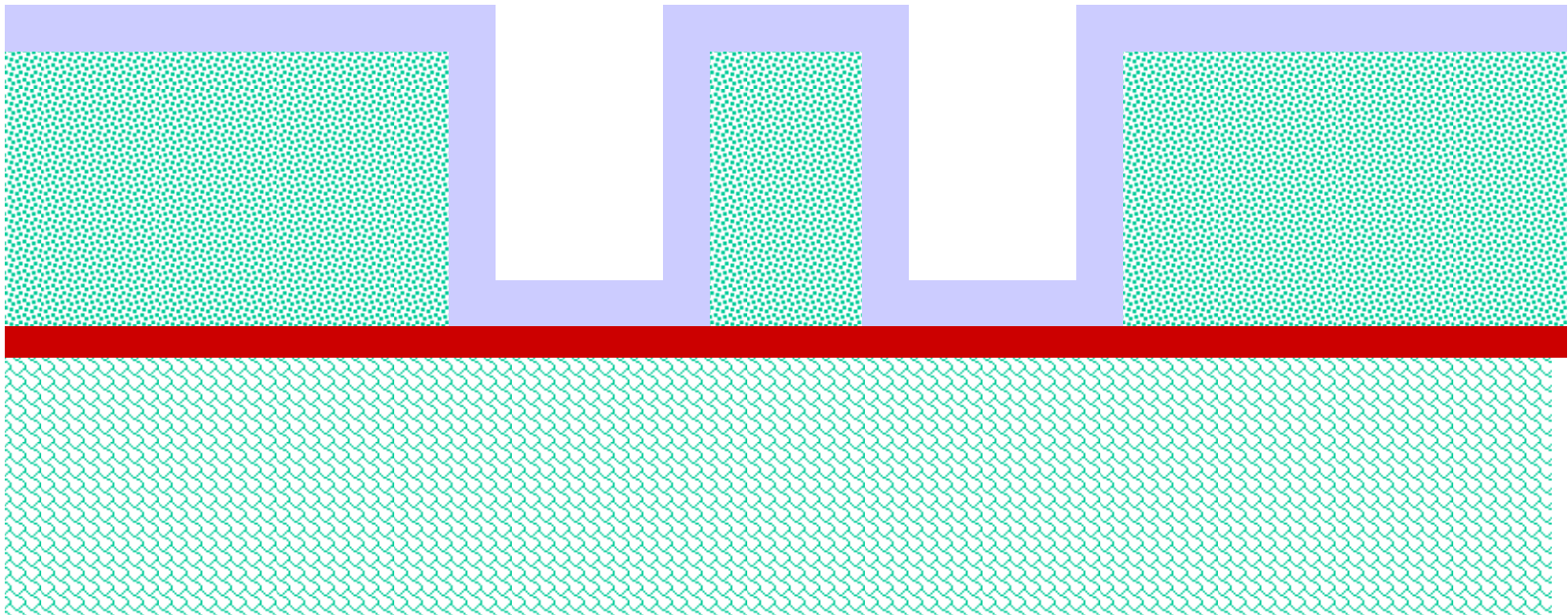
A quick model of interconnect construction

Etch spots in the dielectric for wires



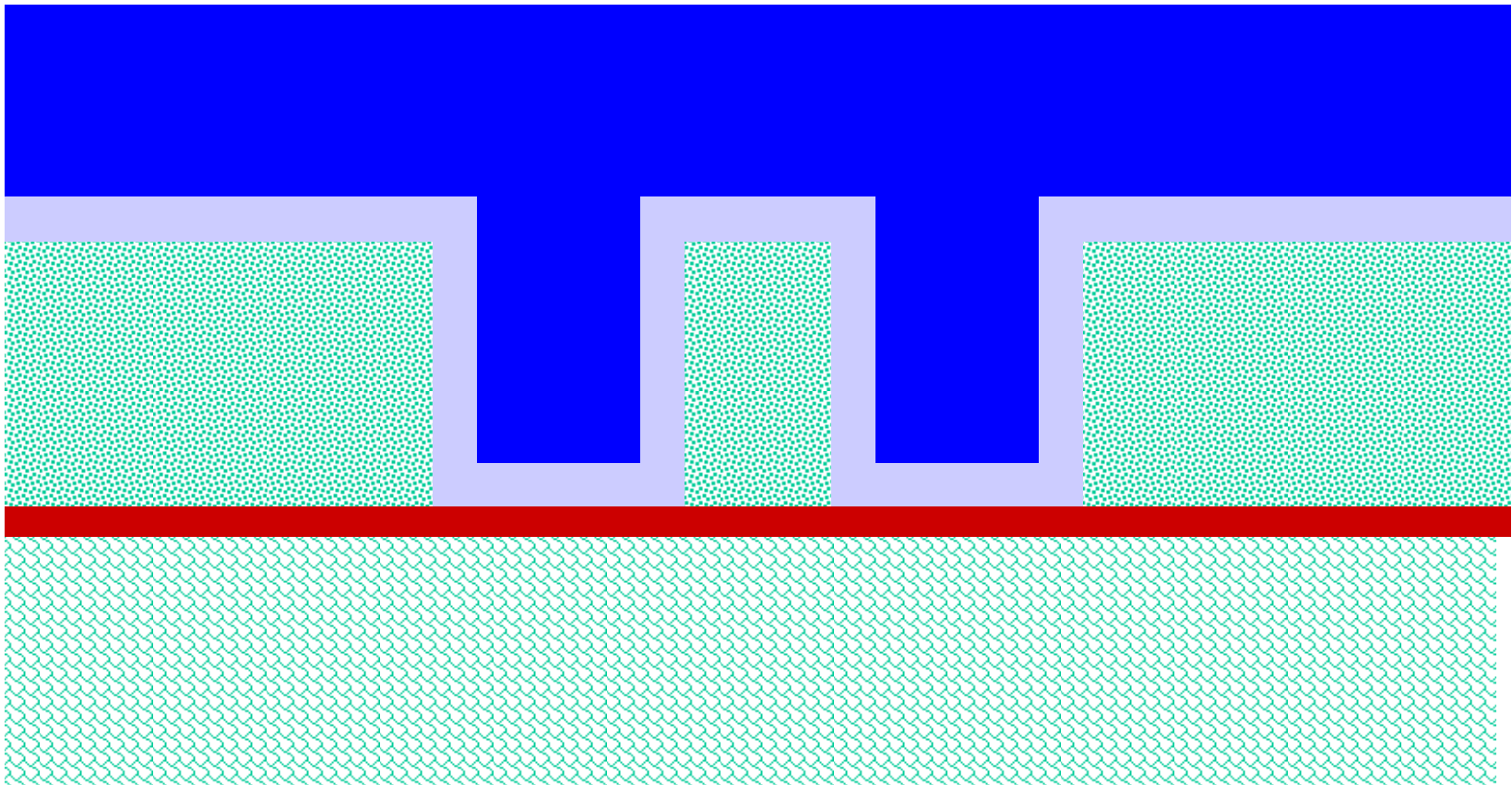
A quick model of interconnect construction

Deposit the barrier metal



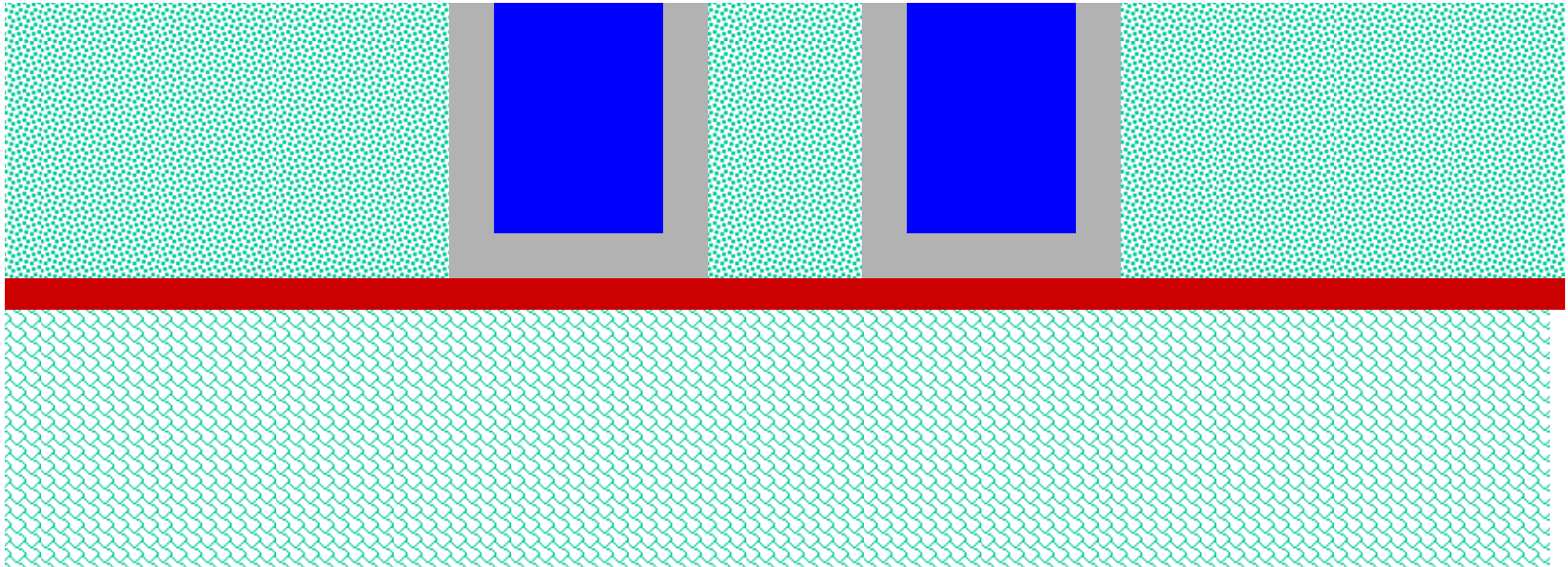
A quick model of interconnect construction

Deposit the copper



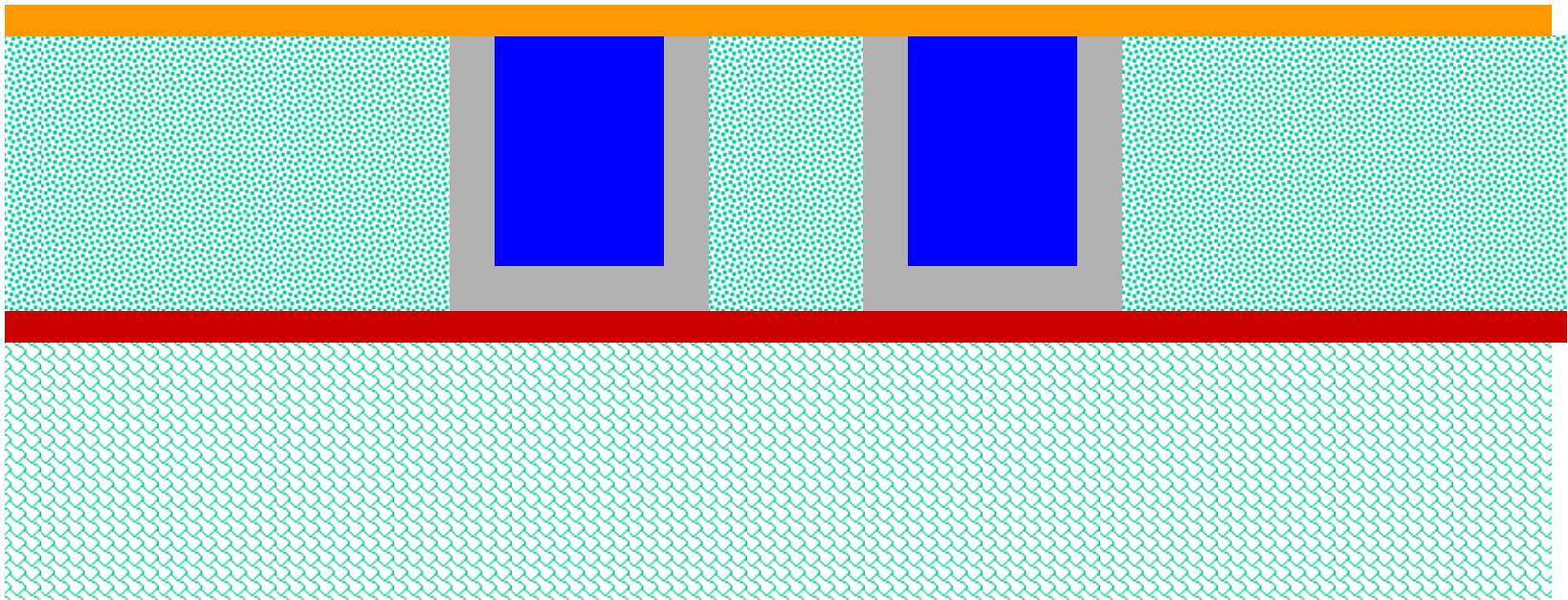
A quick model of interconnect construction

Grind it flat



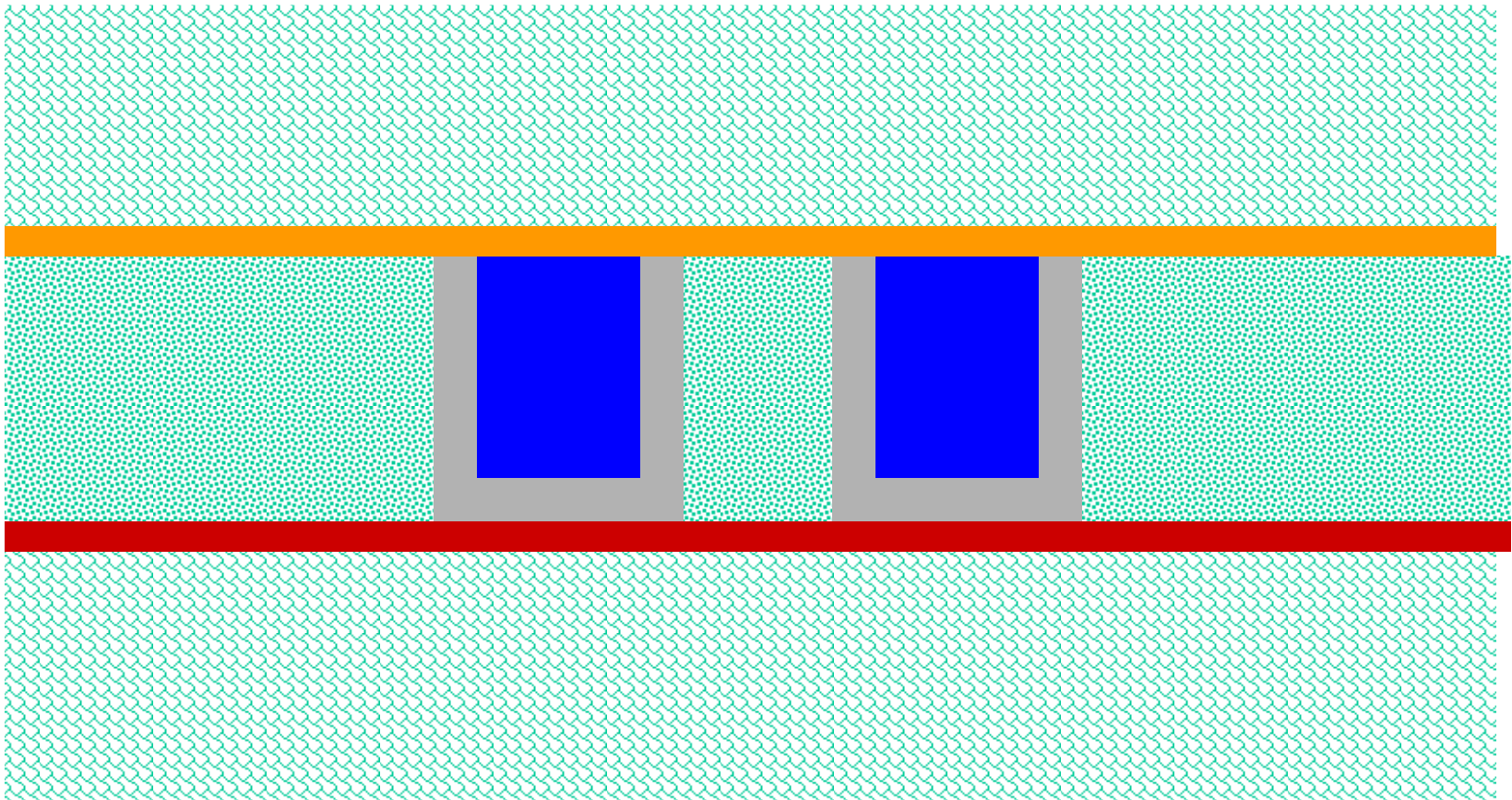
A quick model of interconnect construction

Add a dielectric cap/via etch stop



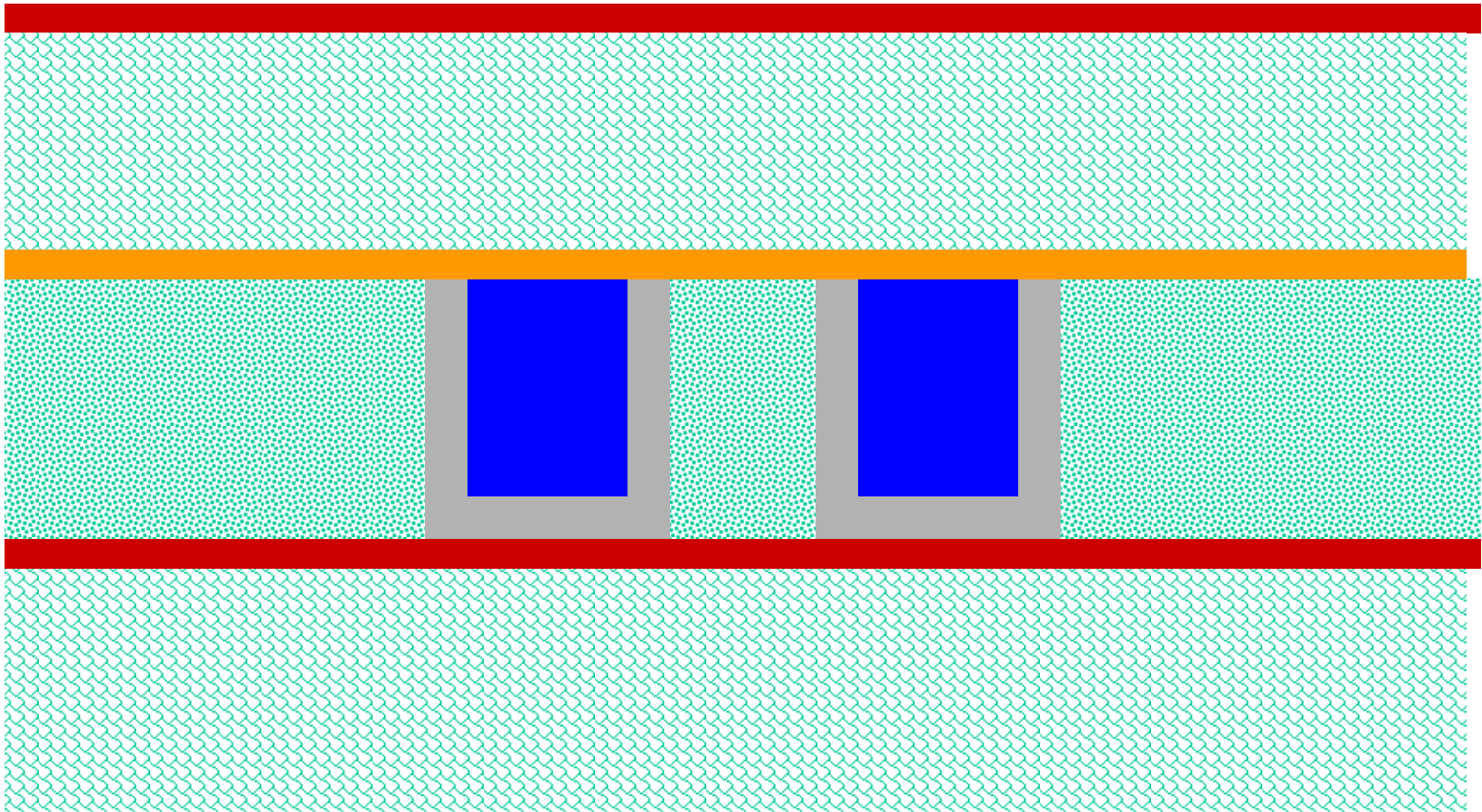
A quick model of interconnect construction

Add more inter-layer insulator



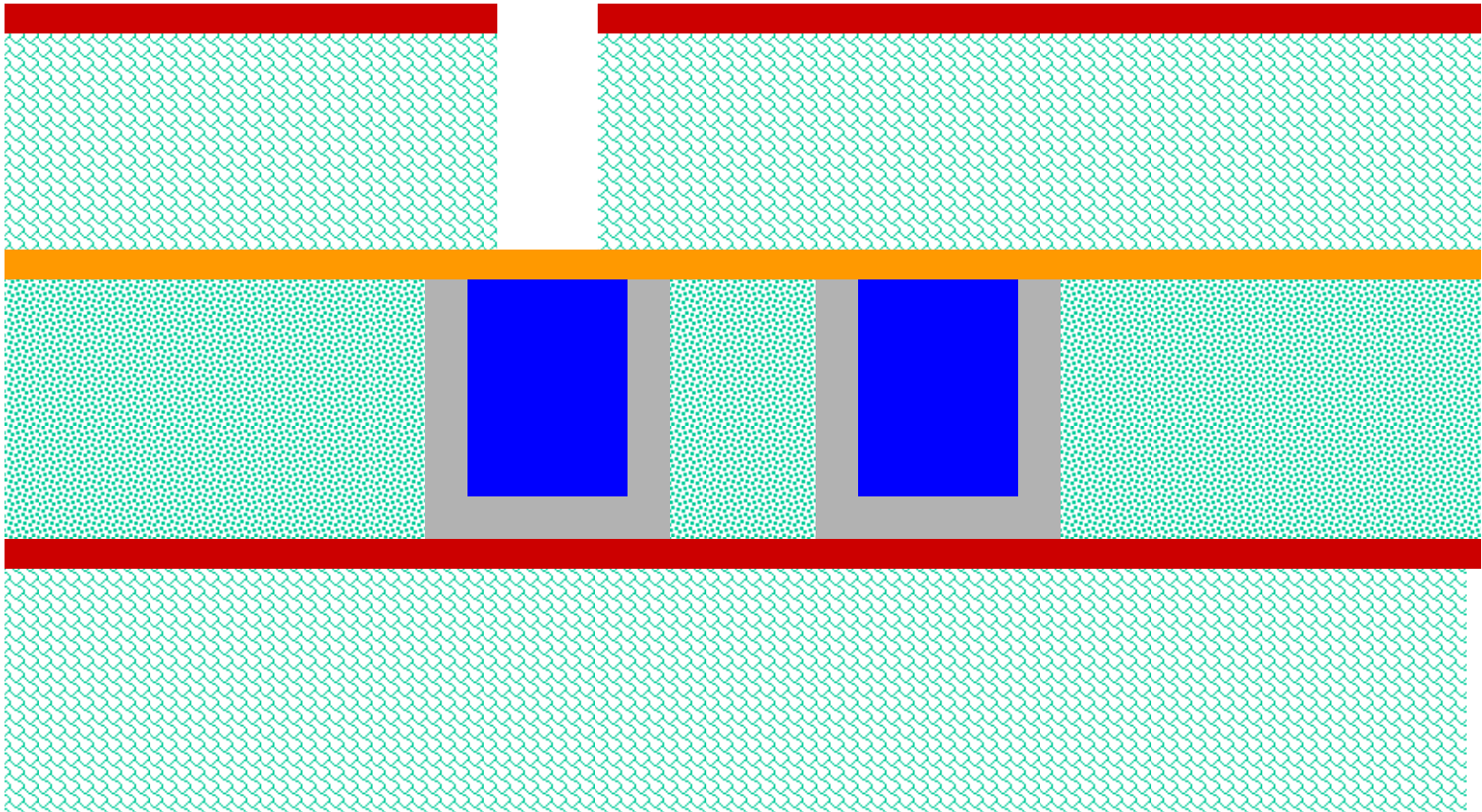
A quick model of interconnect construction

Add another etch stop



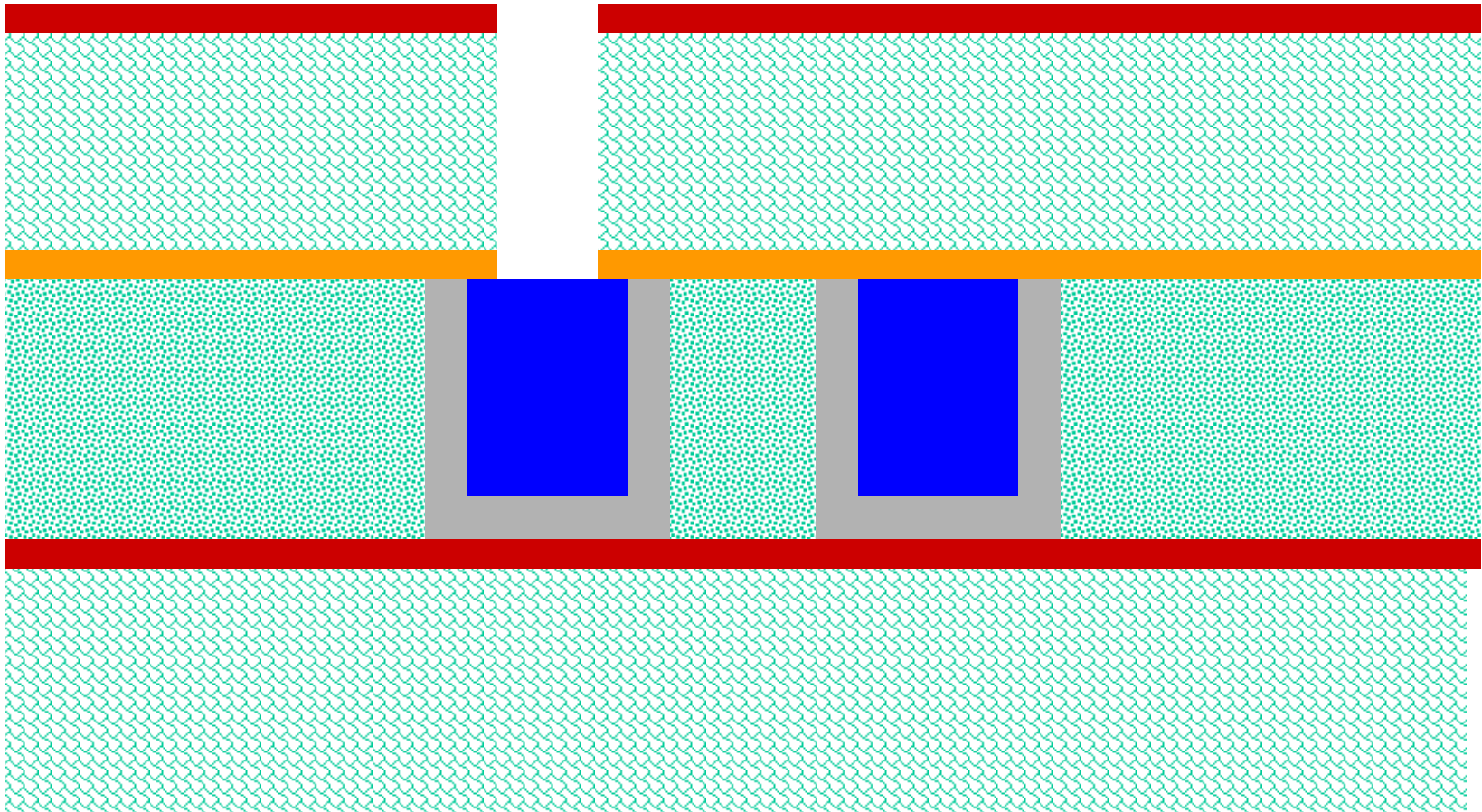
A quick model of interconnect construction

Cut holes for the vias (part 1)



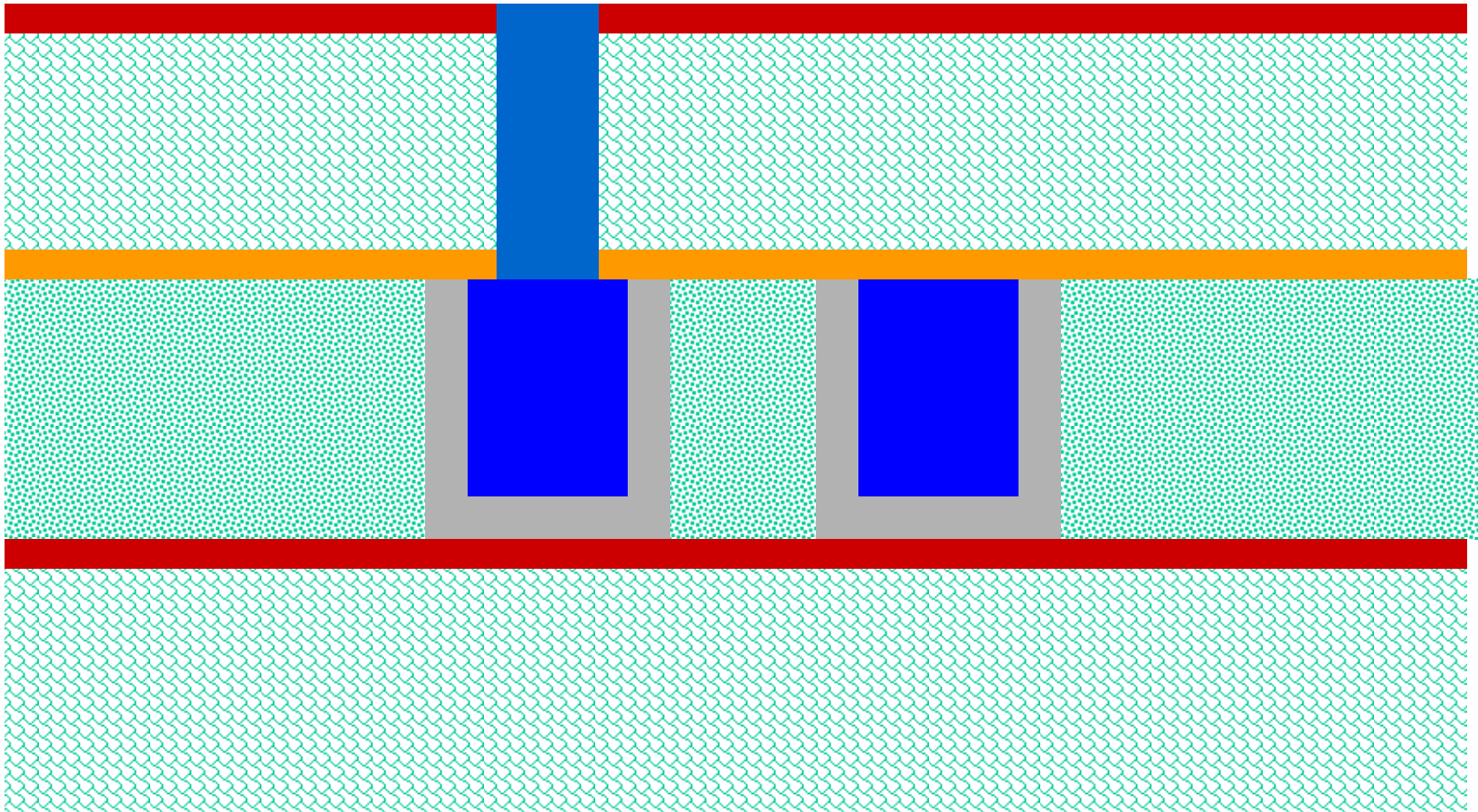
A quick model of interconnect construction

Cut holes for the vias (part 2)



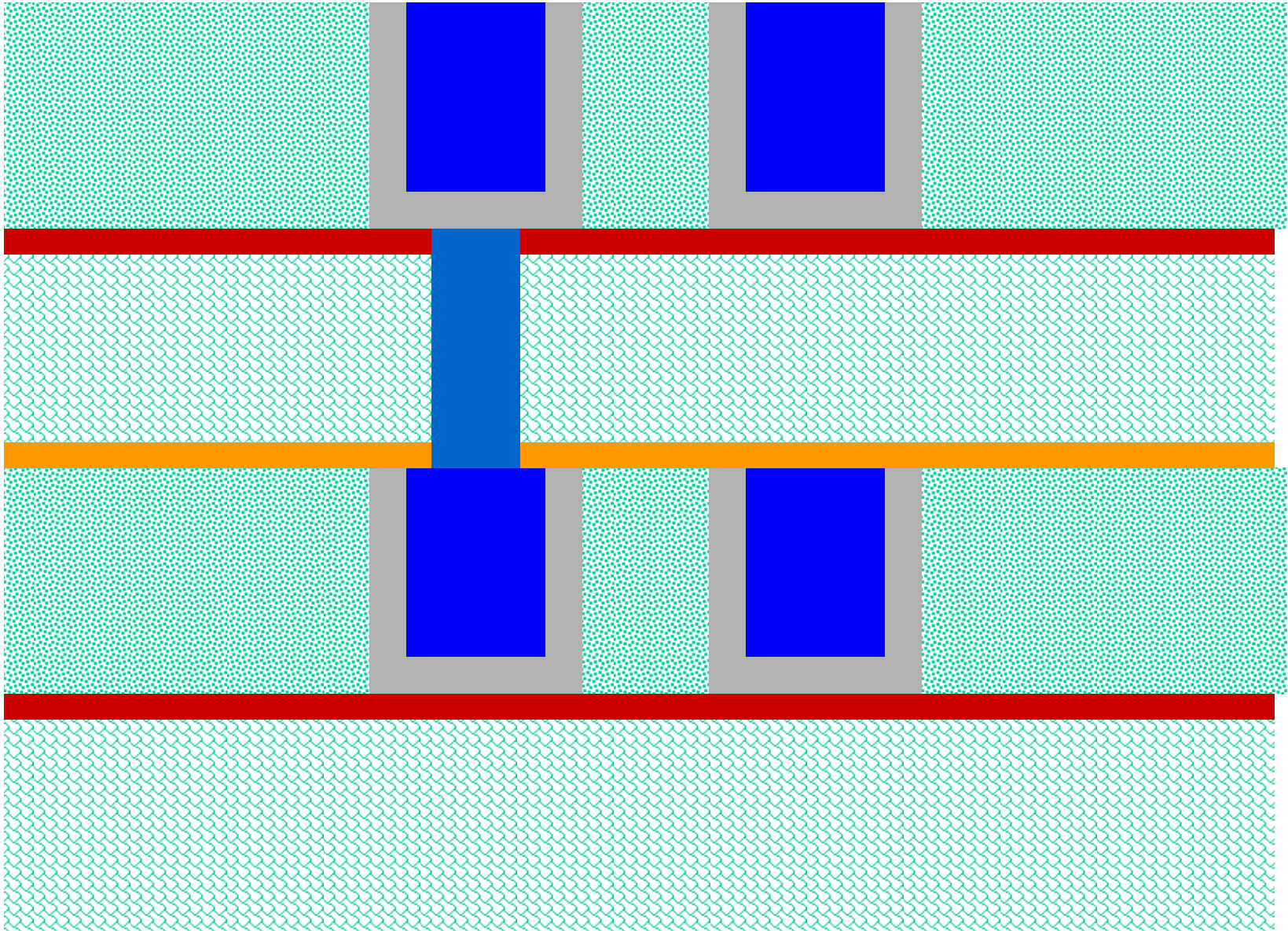
A quick model of interconnect construction

Deposit via metal (or do as part of next layer)



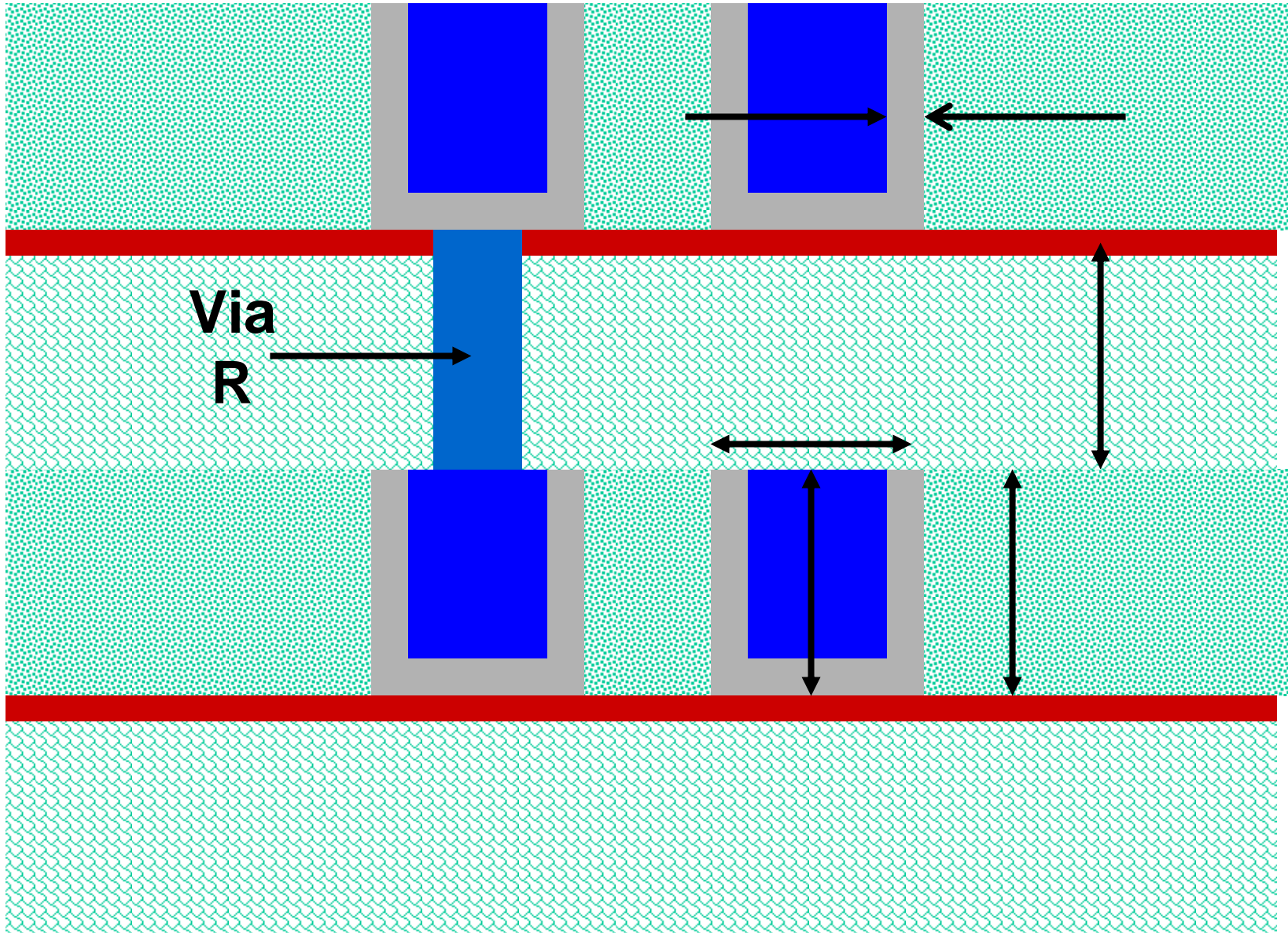
A quick model of interconnect construction

Repeat, repeat, repeat



A quick model of interconnect construction

Many variables per layer



Why are two nominally identical wires different?

- Litho: exact size and shape depend on environment
 - Also OPC treatment, which may be known or unknown
- Etching – also depends on environment
- Wires may be horizontal or vertical
- Thickness of metal depends on CMP environment
- Sizes and placement accuracy depend on where in the optical field wire is located
- Deposition and polishing steps depend on chip location on the wafer
- Alignment between layers can affect vias

On chip variation (electrical)

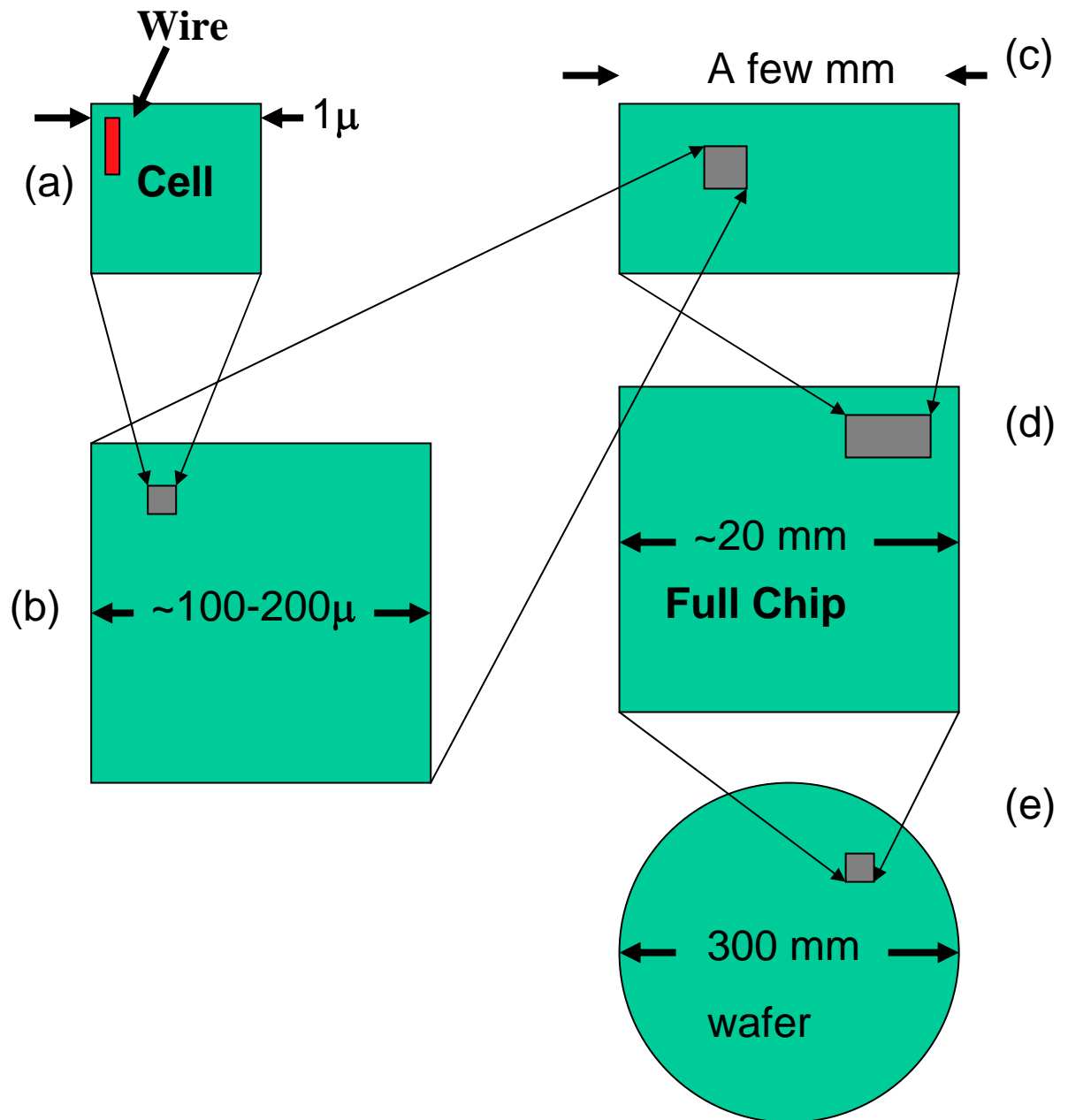
- Even if the wires are physically identical, delays may differ
 - Switching of signals on adjacent nets
- **On-chip variation (other)**
- Fab may not be known (or even built yet)
- Finally, statistical variations not predicted above
- Each source of variation behaves differently

Examples of deterministic uncertainty

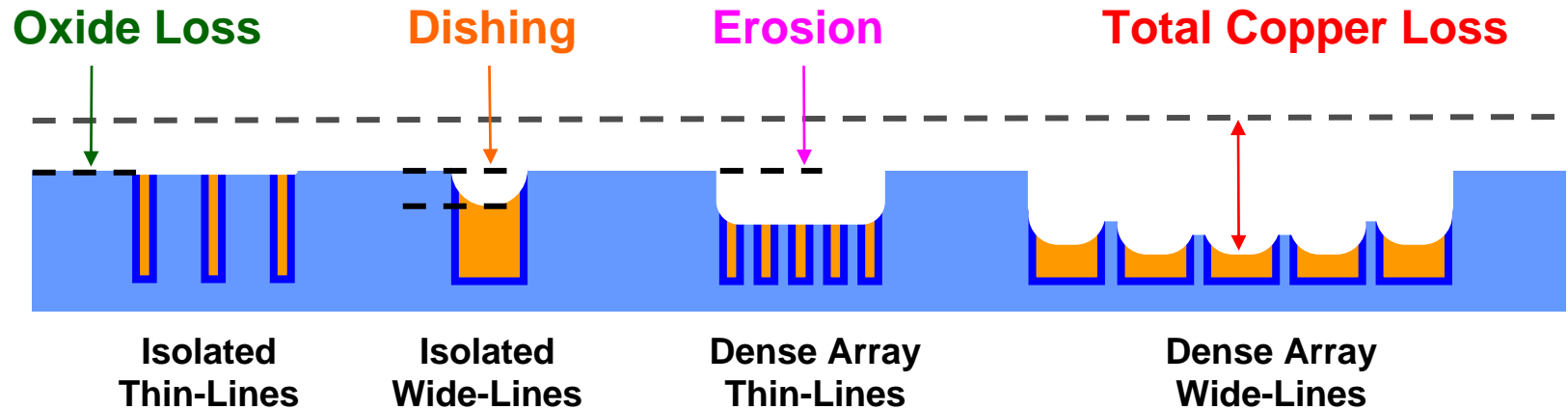
- Effects we could model, in theory
- Remain uncertain for two separate reasons
 - Early in (bottom up) flow, environment is not known
 - Later environment is known, but not modelled
- Examples
 - Optical and etching environment
 - CMP etching as function of location on a wafer
 - Feature size as a function of position in optical field.

Environment not known

In a bottom up flow, the environment is poorly defined to start with. It becomes better defined as the design progresses, until the chip is finally built.



An Illustration of Layout Dependent Variation

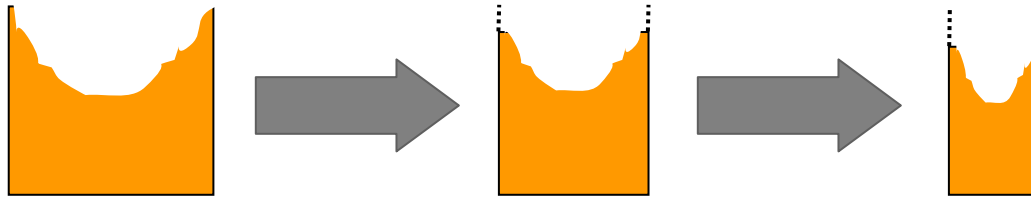


- Dishing
 - Pad bending into large line features
- Erosion
 - Oxide and metal removal in dense features

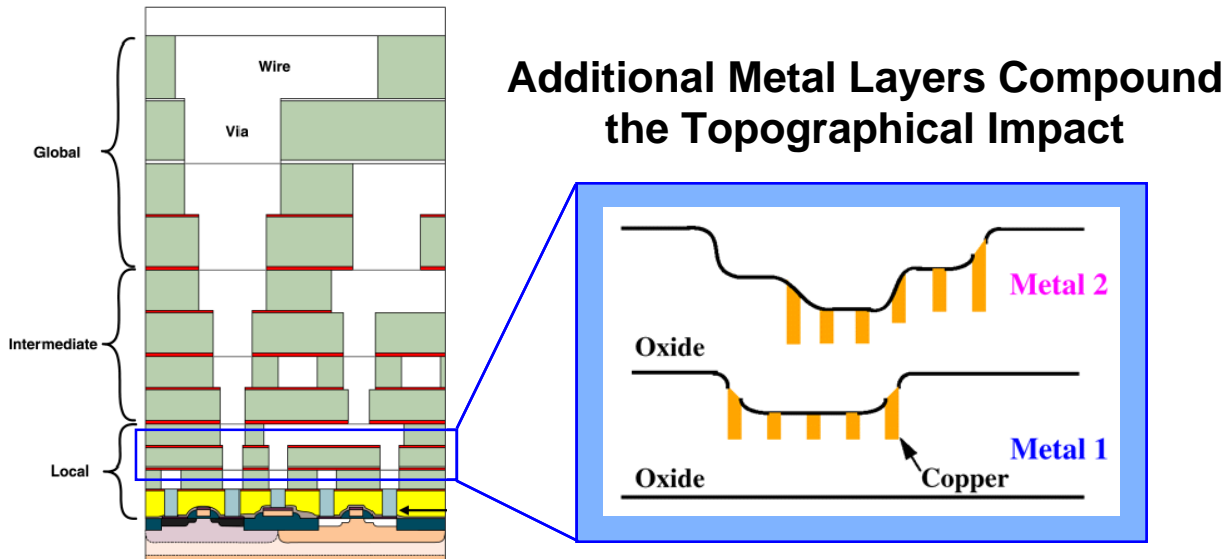
Picture from 2005 DAC
tutorial of NS Nagaraj

Scaling and Multi level Issues

Scaling: 130nm < 90nm < 65nm



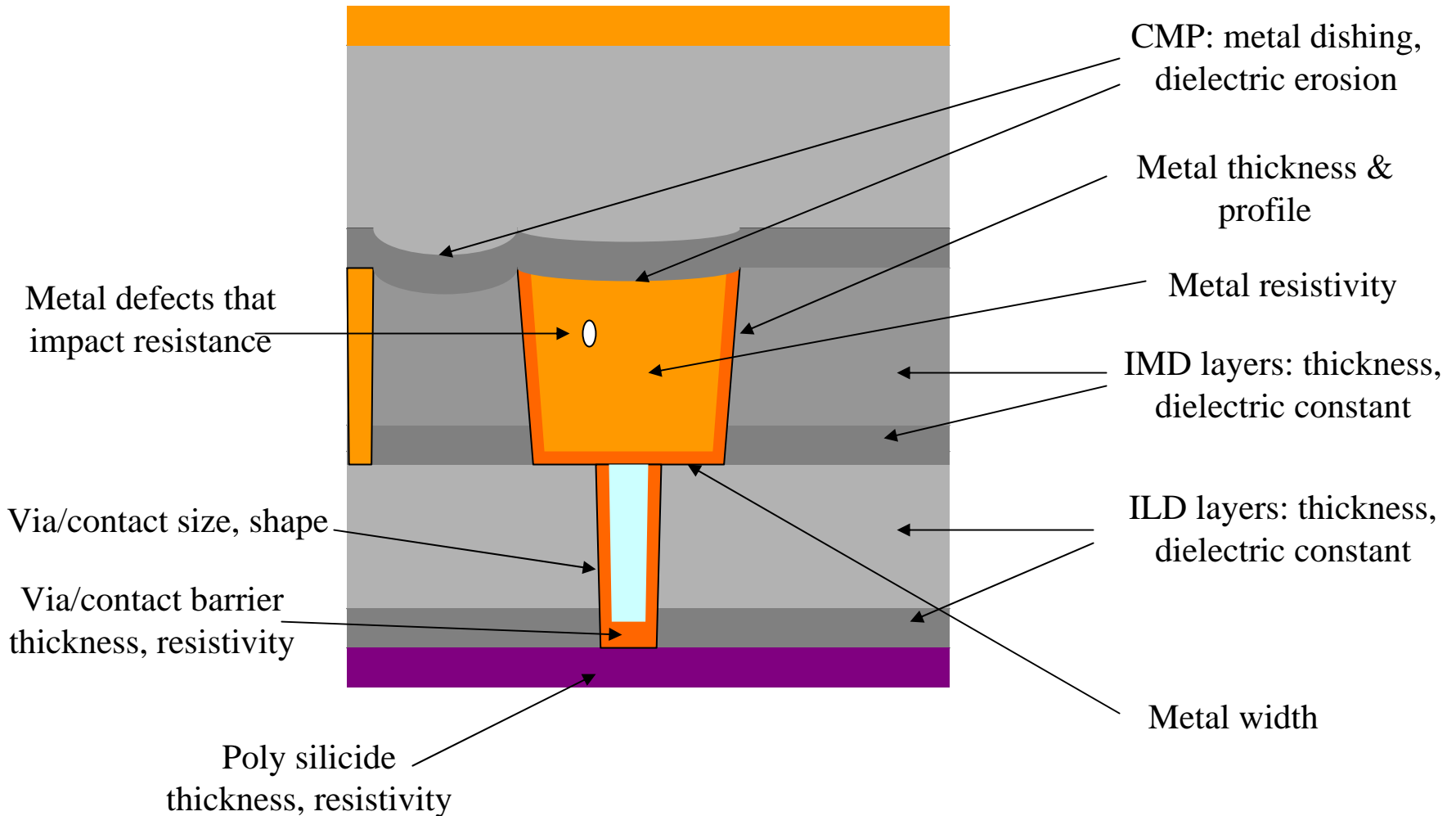
Multi Level: variability amplified at higher metal levels



Picture from 2005 DAC tutorial of NS Nagaraj

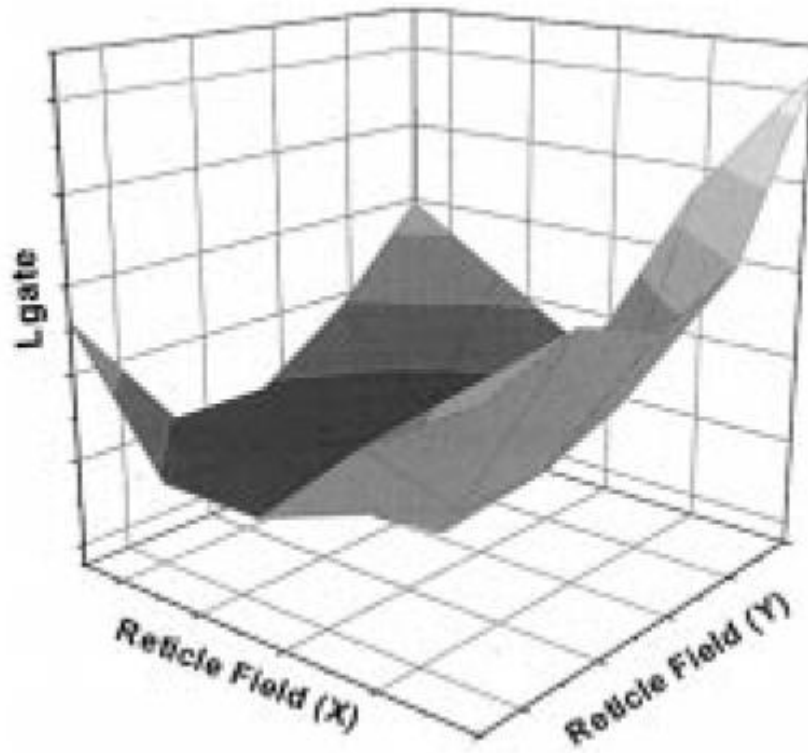
Types of Variations: Metal, VIA, contacts

Process parameters that cause variations in interconnect parasitics:



And some effects are not modelled

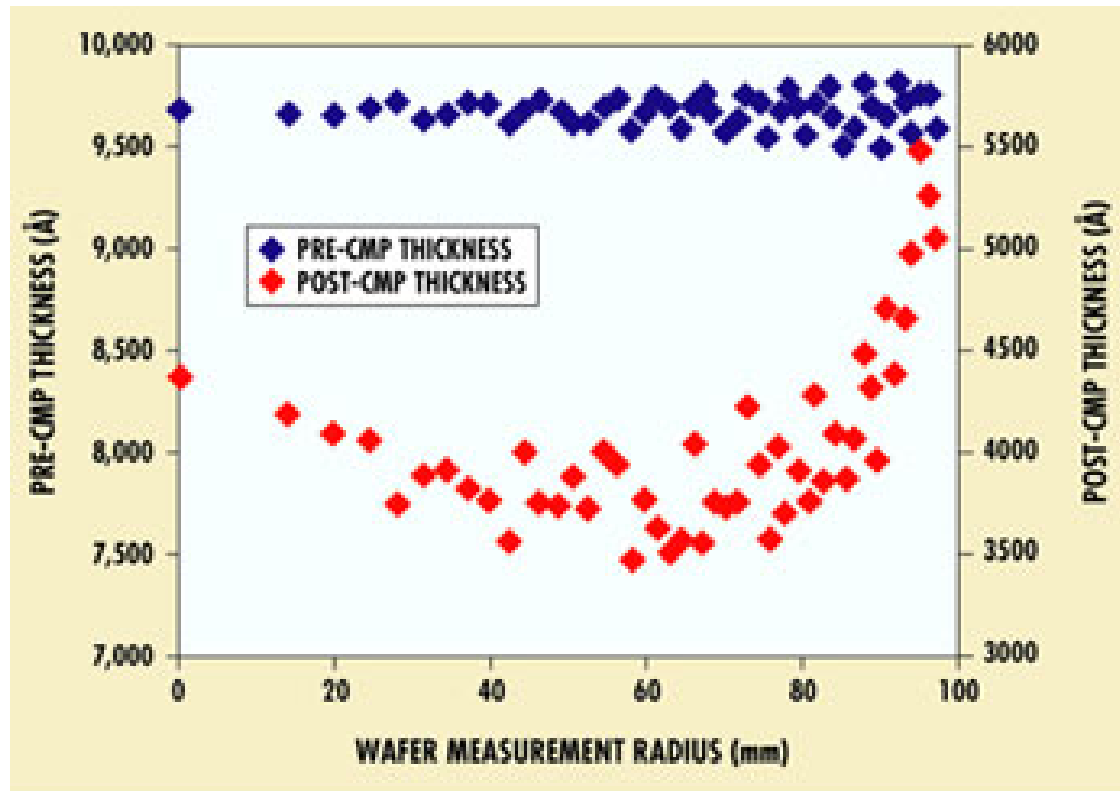
- Effective width as a function of position in field
- This plot is for gate length, but same applies for wire W .



**From Orshansky, 2002,
“Impact of Spatial
Intrachip Gate Length
Variability on the
Performance of High
Speed Digital Circuits”**

Metal thickness variation from wafer location

- Metal thickness as a function of position on wafer is very systematic

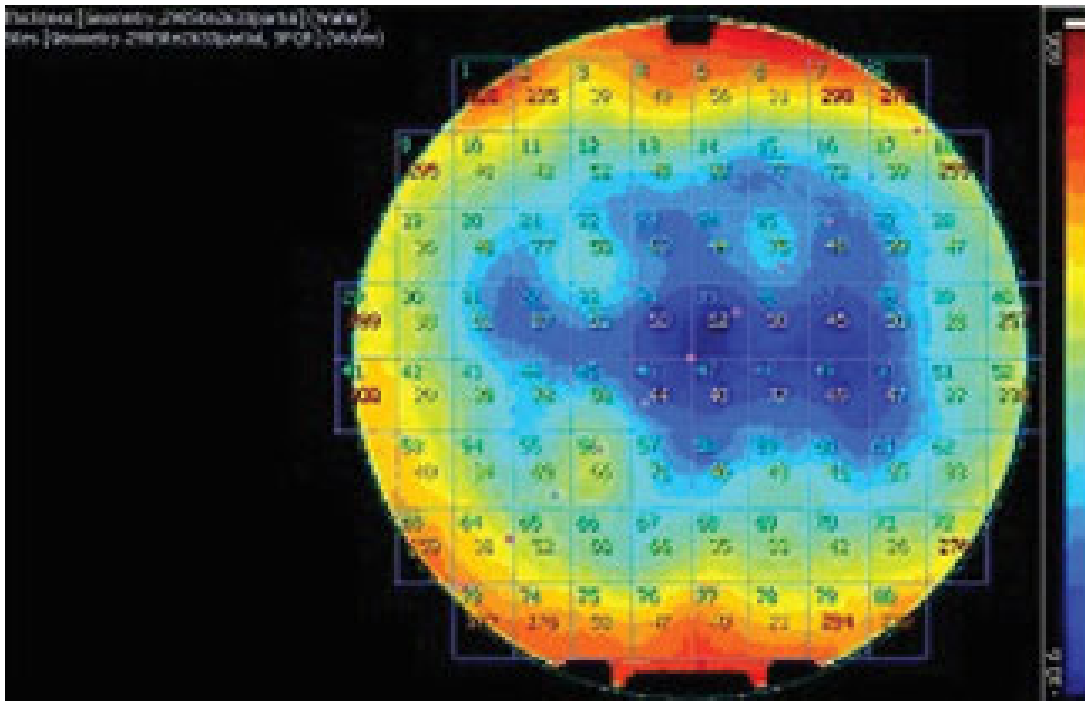


From
<http://www.micro-magazine.com/archive/02/01/Lawing.html>

- Correlation between uncertainties – for example, a chip with thin metal will have small gradients

Non-deterministic effects

- Example: Deposition and etching
- Most deposition and etching steps show slow variation across the wafer
- Each chip sees a value + gradient + small curves



From
<http://www.micromagazine.com/archive/03/10/maleville.html>

What do current design rules say?

- Limits are very loose
- Little distinction between on-chip and inter-chip
- Wire R has a 3:1 range, for example
 - 0.134 +- 0.67 ohms/sq
- Vias have a huge range, for example

Via type	Min R	Max R	Ratio
Metal 1-4	0.5	12	1:24
Metal 5-6	0.2	6	1:30
Metal 7-8	0.1	5	1:50

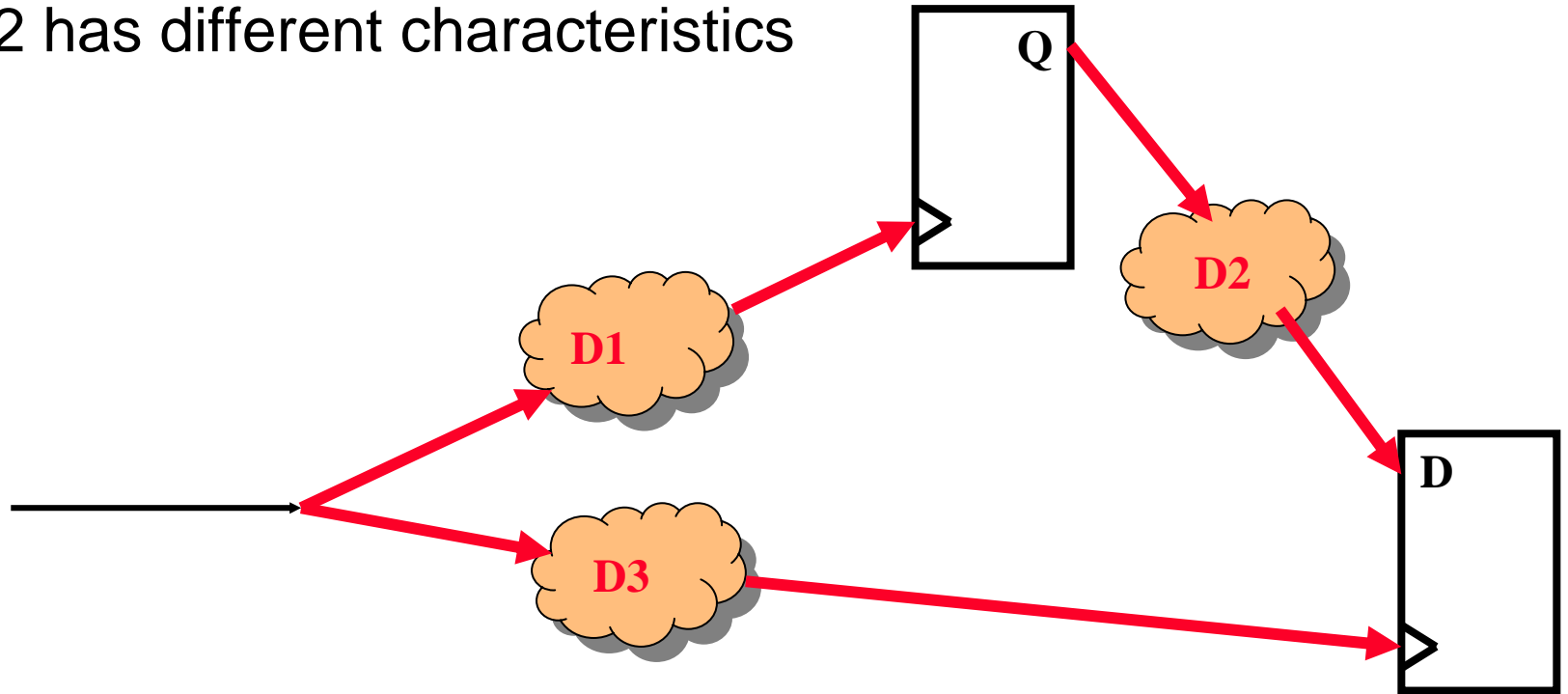
- These could happen at adjacent lines, according to the rules

What little on-chip info exists is extremely specific

- A major manufacturer will only guarantee matching if you have:
 - Identical wires geometries
 - AND they are on the same layer,
 - AND with an identical geometrical environment,
 - AND Length $< X$ (much less than chip size)
 - AND Separation $< Y$ (an even smaller length)
 - AND Same stuff above and below
- Then, the wires will match within 5-10%
- Another major manufacturer does not even specify this much!
 - Only worst case numbers, no distinction between on-chip and inter-chip variation

Designers need more info

- This matching might be enough to compute clock skew
- But only for a specific style
- Designers need matching for more cases
- Traditional example is setup/hold analysis
- D2 has different characteristics



Another example:

- Assume metal is thick, then
 - C is high,
 - So dynamic power is high.
 - But if metal is thick, the power supply grid is lower R.
 - Need correlation between narrow and wide lines, in very different environments

How do designers cope with uncertainty?

1. Compensate for it later
 2. Reduce the uncertainties themselves
 3. Reduce the impact of uncertainties
 4. Guardband against uncertainties
 5. Statistical analysis
- Almost all real designs use a combination of these

Compensate for it later

- Idea is that uncertainty can be removed later
- Post-process metal fill an example
- Designer ignores the uncertainty, and assumes what is drawn will eventually be produced
- Limitations
 - Cannot optimize for cost vs. completeness
 - Correction may not be fully possible (i.e. corners in litho)
 - Destroys hierarchy: must be used at the end of design
 - No tools for some effects (etch, optical field position)
 - 2nd order (eg. Focus and dose latitude) depend on correction
 - Unknown post-process behavior is itself an uncertainty!
 - Hard to do optimization in this model

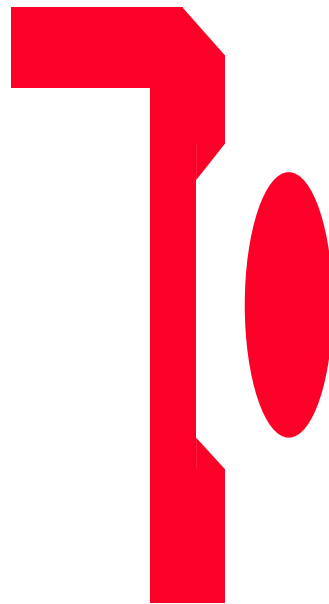
Limitation to fix later: Corrections not complete

- If OPC worked perfectly, designers could ignore it
- But as we scale down in K , complete correction is not possible (i.e. corners), or too expensive

Drawn



No OPC



OPC correction



as fabbed

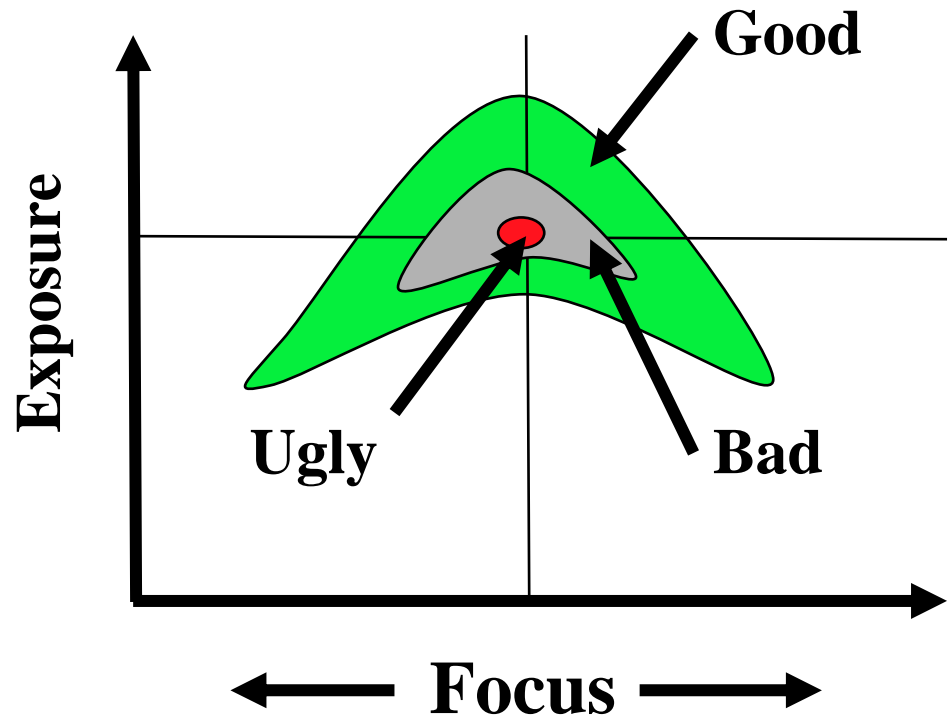


Limitation 2, example : Process window

- What does 'process window' really mean?
- Neither exposure or focus can be perfectly controlled

How far you can get from nominal, and still meet tolerances, is your process window

OPC determines this, but designer does not control OPC

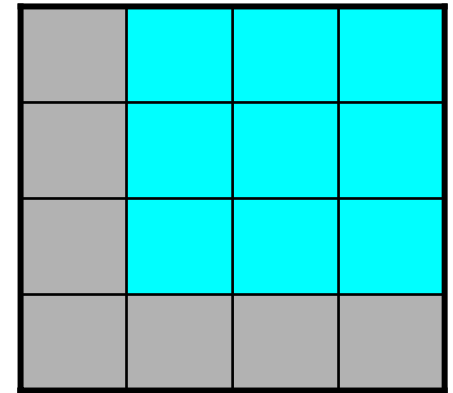
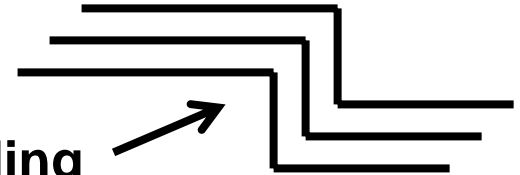


Limitation 3: cannot optimize correction

- Rules allow for (say) 25-75% metal density
- Should you do min cost correction to get within this range?
- Or try real hard to get as close to 50% as possible?
 - Better matching
 - Maybe better yield
 - More mask cost
- May depend on how much matching you need
- But if it's done as a post process, designer does not control

Reduce the uncertainties directly

- Best example for interconnect is clock shielding
- SRAM and DRAM folks also do this
 - Force litho environment to be known
 - Main array is tiled with identical cells
 - Add dummy rows and columns
 - Only a few different parent cells allowed
- Shielding clock nets helps in two ways
 - Reduces uncertainty by coupling
 - But also reduced lithographic and etching uncertainty
- Limitations of this approach
 - Required detailed designer understanding
 - Amount of reduction in variation is not clear



Reduce the impact of uncertainties

- Build design so it works despite uncertainties
 - Or at least works in as many cases as possible
- Analog designers call this ‘design centering’.
 - Traditional design centering has limitations
 - Only practical on small circuits
 - Digital designers have only a few tradeoffs to work with
- Not simple for interconnect design – almost only way to use this approach is to use asynchronous designs
 - Inherently robust to many variations
 - Performance tracks process variations
 - But infrastructure (tools, trained designers, market acceptance, even thought process) is less well developed.

Guardband or worst case

- Fallback approach when all else fails
- By far the most commonly used option
- Limitations
 - May not work at all for analog
 - Too pessimistic – leaves too much performance on the table.
 - Worst case of process generation N+1 may be worse than generation N!

Statistical design

- An entire topic by itself
- Mostly used for gate analysis so far, but can be done for interconnect, too.
- Random component of each layer is fairly small
- But layers are very uncorrelated
 - Good for statistical timing
 - Extraction must keep track of the source of parasitics

These help, but offer no guarantee

- So what do designers actually do?
- Build structures such as H-trees that are symmetrical
- Use mesh structures to even out variation
- Design each clock branch with similar characteristics
 - Similar metal usage by layer
 - Similar via counts
- But even these are not close to enabling the design to work according to the official worst case
- So designers pick a number based on experience and use that
 - Say, for example, that 30% of the total variation possible can occur across a chip
- But this will be overkill in most cases, and not enough in others!

Two suggestions

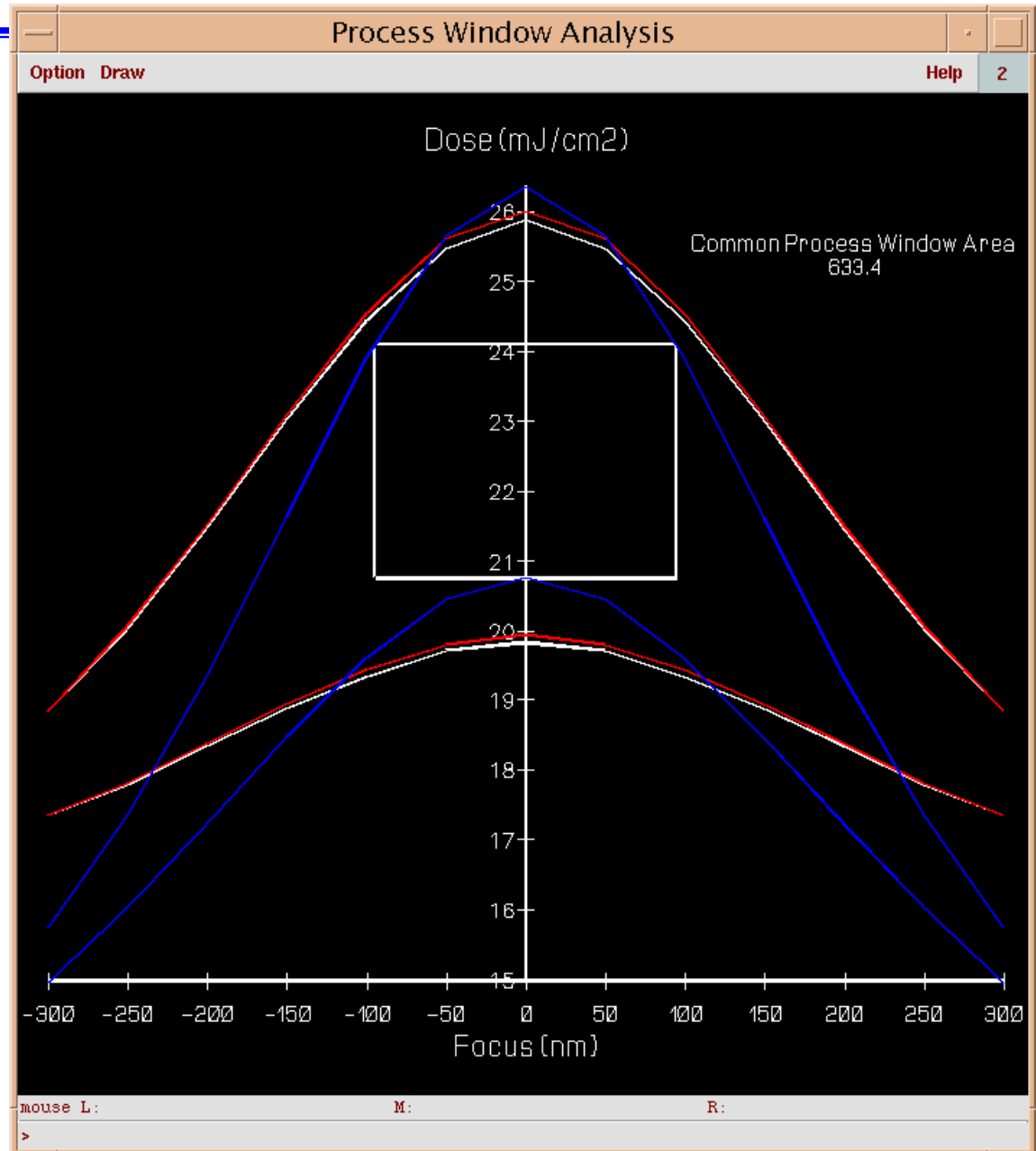
- So what can we do about this?
 - Don't include any uncertainties you don't need
 - Understand the uncertainties and their impact better
 - Combine the remaining uncertainties more realistically
- How can we do this? What would a physical model look like?
 - Need models of the things we can control
 - Litho, etch, CMP environment, horizontal/vertical, location in field
 - Need models of the spatial variation of uncontrolled effects
 - Focus, dose, deposition, etching
 - Systematically use sensitivities to evaluate effect of uncertainties

What might these models look like?

- Litho is most complex
 - Depends on neighboring image
 - Highly non-linear since it must incorporate effects of OPC
 - Or designer does OPC, then much more well behaved
 - Lookup table of patterns?
- Etch and CMP depend mostly on width and local density
 - Radius of 10 microns for etch, 100s for CMP
 - Local density changes also have an effect
- Horizontal vs. vertical adds additional uncertainty
 - Constant may suffice
- Effects of location in field is a polynomial (this is how lenses are designed and characterized)
- Focus is partly predictable, from bottom layer CMP results

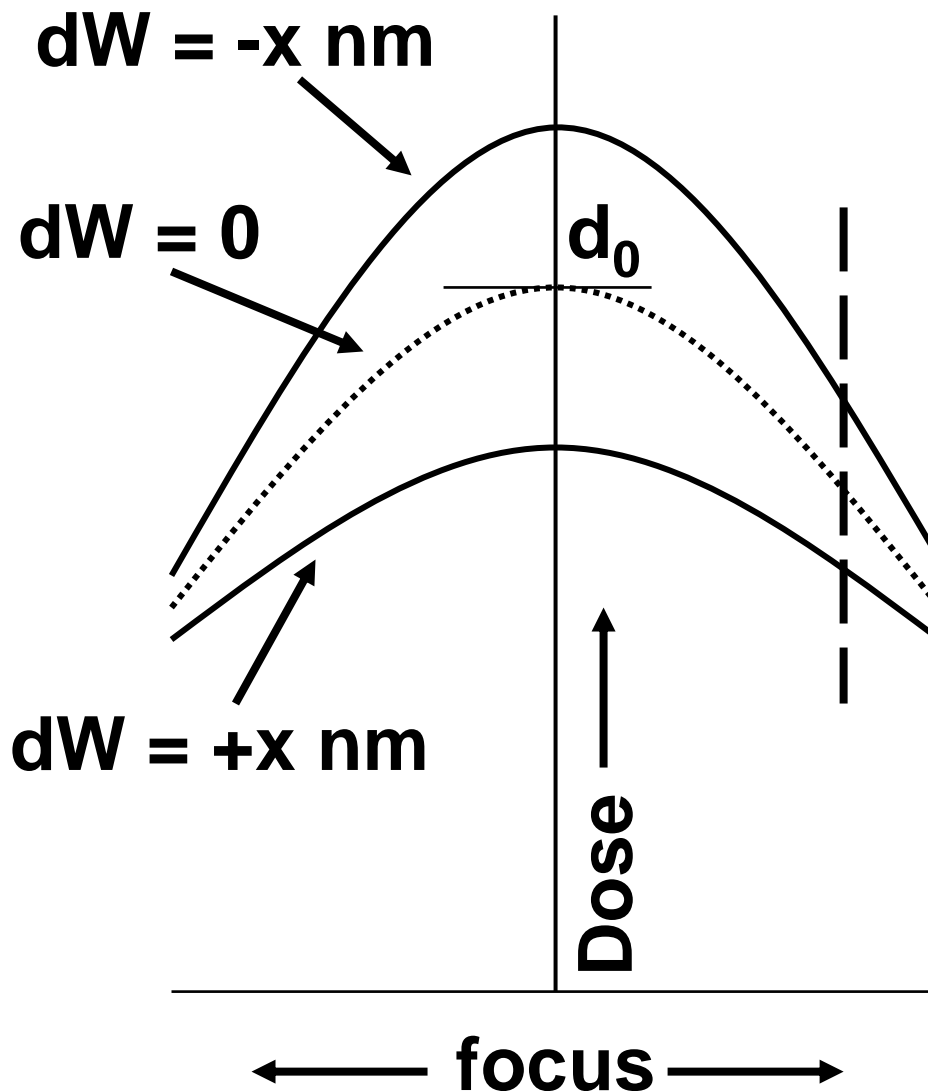
Litho model

**Litho model
takes local
environment,
computes a
function of 2
variables, focus
and dose**



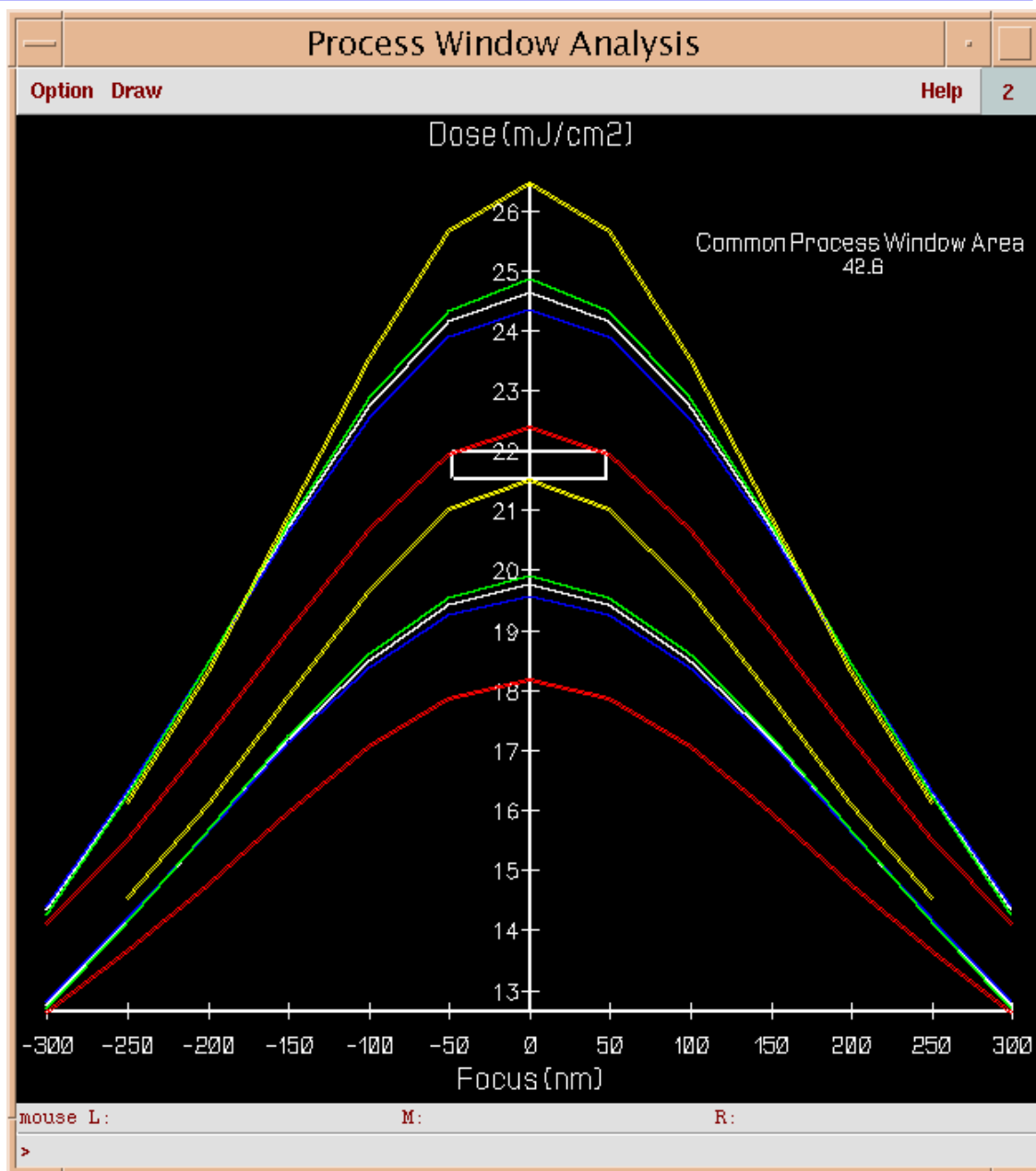
Litho analysis results give $w(f,d)$

- Relationship is not linear, but is smooth
- Low order model might look like this
- $D_{mid} = d_0 - k_1 f^2$
(equation of the dotted line)
- $dW = k^*(d - D_{mid})$
where $k = (S_0 + k_2 f^2)$
where S_0 is the sensitivity at $f=0$
- End up with terms of type: const, d , df^2 , f^2 , f^4



Focus and dose are highly correlated locally

- Correlations MUST be taken into account
- Red and yellow curves OK by themselves, but very little common process window



How do effects vary with distance?

- Dose, deposition, and etching vary slowly across the wafer
 - Value + gradient + a few higher terms probably OK
 - For some, value and gradient may be correlated
- Focus is similar
 - Global term from auto-focus in field
 - Slow terms from wafer flatness
 - Some contribution from underlying layers
 - Cell level analysis assumes completely correlated

So how many variables do we need?

- Surely need at least 5 variables per layer
- Focus, dose, metal thickness, ILD thickness, via R
- Plus maybe 2 more (from NS Nagaraj)
 - Conformal dielectric thickness
 - Barrier layer thickness
- Each variable needs a base value, gradients (at least) and perhaps low order curvatures for.
- Interconnect characteristics for the layers are largely independent and uncorrelated
- So we end up with (rough guess) 40 variables per layer times 10 layers, or about 400 variables.

What about vias?

- Quite frankly, I have no idea of what a via correlation model looks like
 - Must encompass 50:1 variations
- Cross section presumably correlates with an etching model
 - How does the local via density affect this
- Presumably terms that correlate with ILD thickness
- Probably terms whose root cause is mis-alignment
- ?? Stress related terms from local via density, out to many microns
 - Known to affect yield
 - Does it affect R?
- It would be great if someone would study (and publish) this!

How can these be combined?

- Affine (power series) seems like only solution to me
 - Keeps correlation
 - Allows worst case, corners, or statistical combination
 - And combinations of these
- Handles correlation between R and C
- Known techniques for propagating this through delay calculation, then timing graphs
- Good representation for optimization
- Matching can be addressed straightforwardly by subtraction
- Representation is (or can be) closed under usual operations – see a huge number of timing papers

$$A = a_0 + a_1 X_1 + a_2 X_2 + \cdots + a_n X_n + a_r X_{ar}$$

Technical Difficulties/Opportunities

- Uncertainty is a flow problem, not a tool problem
 - Will require flow centric development
 - Fitting a new point tool could be very hard
- New code and algorithms are needed
 - Extract in context
 - Optical models in extraction tools
- New test chips may be required to characterize effects
 - Via R correlation
- New data formats are required, or at least agreed on
 - Optical and CMP models, from fabs to customers

Business Difficulties/Opportunities

- Uncertainty is a flow problem, not a tool problem
- Data is sensitive
 - Litho and etch models
 - Characteristics of neighboring chips on multi-chip reticles
- Hard IP is now more difficult
 - Not just GDS-II, but a lot more
- Uncertainties are time varying
 - Inadvertent – process drift
 - Deliberate – process improvement
 - New steppers/scanners/etch, etc.
- Second sourcing is more difficult
- Data hard to predict for a new process
- Harder to assign blame in case of failure

Conclusions

- Lots of causes for on-chip variations
- Existing models expressed in foundry rules are basically non-existent
- Designers are using ad-hoc techniques and experience to cope now
- We can guess what better models might look like
 - Many could be derived from existing characterization data and models
 - Some need additional research
- Then new tools similar to those investigated for statistical timing might could make better predictions
- Lots of technical and business practical issues remain