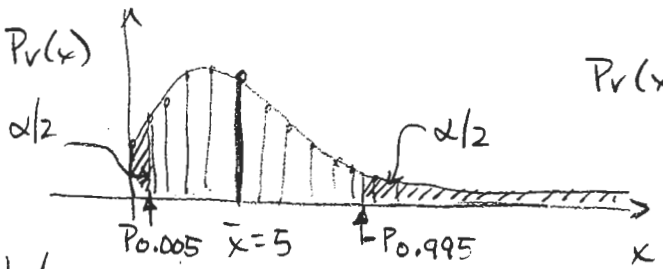
6.780 SEMICONDUCTOR MANUFACTURING | PS #3 SOLUTIONS | SPRING 2003

Problem 1 ISPM defect control chart design

- The foremost decision in choosing a control chart strategy is to identify what statistical distribution applies. In the main etch step we are told that an average of 5 particles/min is detected. Assuming an i.i.d. sampling from one minute to the next when the process is in control, a poisson distribution appears to be appropriate ("events" per unit opportunity).

  So in the main etch step, the distribution for x, the # defects per minute for each minute of the main etch run might be:



$$Pr(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

$\mu = \lambda$

$\sigma^2 = \lambda$

$\hat{\lambda} = \bar{x} = 5$ ptc/min.

- To pick control limits we want $\alpha = 0.01$. To be accurate, since $\bar{x} = 5$ is relatively small we might directly use the poisson distribution to set UCL and LCL; using $\alpha/2$ for the lower and upper percentage points in App. I.

  $\lambda = 5$, $P_{0.005} \doteq 0$

  $P_{0.995} \doteq 11.2$ (between 11 & 12)

So

UCL =	$P_{0.995}$ (for $\lambda = 5$)	= 11.2	particles
CL =	$\hat{\lambda}$	= 5	particles
LCL =	$P_{0.005}$ (for $\lambda = 5$)	= 0	

As a comparison, if we considered the Poisson distribution to be approximated by a normal distribution with

$$\hat{\mu} = 5$$
$$\hat{\sigma} = \sqrt{5}$$

we would have $z_{0.995} = 2.58$ so

$$UCL = \hat{\mu} + z_{0.995}\hat{\sigma} = 5 + 2.58\sqrt{5} = 10.8 \text{ ptc}$$
$$CL = \hat{\mu} = 5 \text{ ptc}$$
$$LCL = \hat{\mu} - z_{0.995}\hat{\sigma} = 5 - 2.58\sqrt{5} = 0 \text{ ptc}$$

So even for a relatively small mean particle count, the normal approximation does a pretty good job - we would trigger on 11 or 12 particles, where we plot one point on our control chart each minute.

- In the purge step we can make a similar assumption, and consider the whole purge step as the unit of opportunity for a Poisson distributed number of defects:

$$\hat{\mu} = 35 \quad \text{and} \quad \hat{\sigma} = \sqrt{35}$$

$$UCL = \hat{\mu} + z_{0.995}\hat{\sigma} = 35 + 2.58\sqrt{35} = 50.3 \text{ ptc/min}$$
$$CL = \hat{\mu} = 35 = 35 \text{ ptc/min}$$
$$LCL = \hat{\mu} - z_{0.995}\hat{\sigma} = 35 - 2.58\sqrt{35} = 19.7 \text{ ptc/min}$$

- A comment on other strategies:

(1) A common desire is to use an $\bar{x}$ and R or S chart. This misses or does not take advantage of Poisson distribution knowledge. However, this is not a fatal mistake. Why? Historical or trial data would be used to estimate $\hat{\sigma}$ separately from $\hat{\mu}$, so ultimately one would arrive at the correct limits given enough data. Since both μ and σ (i.e. $\lambda = \mu = \sigma^2$) would change at the same time, we would likely see alarms on both charts.

(2) An even better approach (suggested by a student), is to note that we may have process runs with varying time. We could then treat the run length as a "sample" of the main etch for n minutes, and use a u chart, where each minute is the "unit of opportunity."

This has the advantage of putting all runs on a similar control chart, but with varying control limits depending on the length of the run.

$$CL = \bar{u} = 5 \text{ particles per minute}$$

$$\begin{matrix} UCL \\ LCL \end{matrix} = \bar{u} \pm z_{\alpha/2} \frac{\sigma_u}{\sqrt{n}}$$

where n is the number of minutes, and we plot the average # particles detected during the total main etch step

$$= \bar{u} \pm 2.58 \frac{\sqrt{5}}{\sqrt{n}}$$

The key difference between this and the approach on page 1, is that this correctly accounts for the averaging resulting from multi-minute long runs.

This is the best way to handle the particles resulting from the main etch step as a whole.

Problem 2 $\bar{x}$, R control chart design

- Sample size $n=7$
- Trial data, $m = 35$ samples (each of size 7)

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{35} \bar{x}_i}{35} = \frac{7805}{35} = 223$$

$$\bar{R} = \frac{\sum_{i=1}^{35} R_i}{35} = \frac{1200}{35} = 34.28$$

Parts a Per $\bar{x}$ chart we use the correction factor to estimate the process standard deviation as

$$\hat{\sigma} = \bar{R}/d_2 \quad \text{where } d_2 (n=7 \text{ samples}) = 2.704 \text{ from charts (App. VI).}$$

$$\hat{\sigma} = \frac{34.28}{2.704} = 12.7$$

Since we're taking samples of size $n=7$, we construct our $\bar{x}$ control chart for sample averages ($\bar{x}_i$) with $\alpha = 0.0027$ or 3 σ_{sample} limits:

$$UCL = \bar{\bar{x}} + 3\sigma_{sample} = \bar{\bar{x}} + 3\hat{\sigma}/\sqrt{n} = 223 + 3(12.7)/\sqrt{7} = 237.4$$

$$CL = 223 = 223$$

$$LCL = \bar{\bar{x}} - 3\sigma_{sample} = 223 - 3(12.7)/\sqrt{7} = 208.6$$

For the R chart, and assuming the within-sample values are identically and independently distributed as the across-sample values (not typically the case if our samples come from one lot for one set of 7, then another lot for the next run of 7), then we can estimate the std. deviation for R as

$$\hat{\sigma}_R = d_3 \frac{\bar{R}}{d_2} \quad \text{and} \quad \hat{\mu}_R = \bar{R}$$

Again, for sample size $n=7$ we have

$d_3 = 0.833$ or $\hat{\sigma}_R = 10.56$

$1/d_2 = 0.3698$

$\hat{\mu}_R = 34.28$

Thus the limits for our R chart to monitor within run variance are:

$$UCL = \hat{\mu}_R + 3\hat{\sigma}_R = 34.28 + 3(10.56) = 66.0$$
$$CL = \hat{\mu}_R = 34.3$$
$$LCL = \hat{\mu}_R - 3\hat{\sigma}_R = 34.28 - 3(10.56) = 2.6$$

//

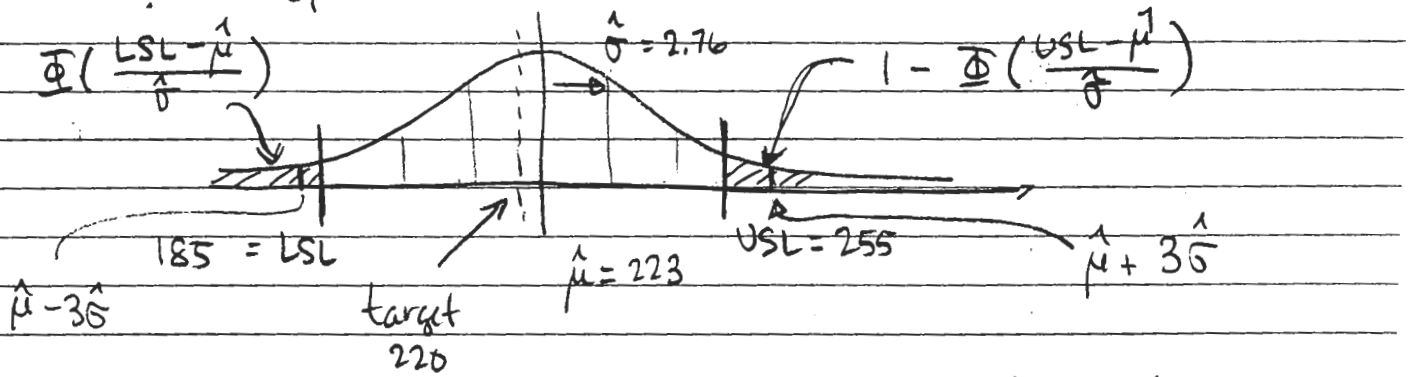
Part b Estimate PROCESS mean and std. dev.

- As discussed above, $\hat{\mu} = \bar{\bar{x}} = 223$

$\hat{\sigma} = 12.7$ //

Part c

Given spec limits 220 ± 35, does process meet specifications?



Whether or not this process "meets specifications" is a matter of degree or definition: it depends on how many nonconforming parts we are willing to live with.

- Because our process is not centered, we should use the C_{pk} measure of "process capability":

$$C_{pk} \triangleq \min\left\{ \frac{USL-\mu}{3\sigma}, \frac{\mu-LSL}{3\sigma} \right\} = \frac{USL-\mu}{3\sigma}$$

(since our distribution is shifted toward the upper CL),

or

$$C_{pk} = \frac{255-223}{3(12.7)} = 0.84$$

Generally a $C_{pk} < 1$ is not considered a capable process, as it will generate a substantial fraction of non-conforming parts:

- Fraction non-conforming:

$$\text{fract} = \Phi\left(\frac{LSL-\hat{\mu}}{\hat{\sigma}}\right) + \left(1 - \Phi\left(\frac{USL-\hat{\mu}}{\hat{\sigma}}\right)\right)$$

$$= \Phi\left(\frac{185-223}{12.7}\right) + 1 - \Phi\left(\frac{255-223}{12.7}\right)$$

$$= \underbrace{1-\Phi(2.99)}_{1-0.99861} + \underbrace{1-\Phi(2.52)}_{1-0.99413}$$

$$= 0.00139 + 0.00587 = 0.00726$$

or 0.726% of parts are nonconforming

or about one out of every 138 parts,

or about 7,260 ppm (parts per million) fail. //

Part d Mean-centering the process minimizes the nonconforming parts. In this case, we move μ to 220, so

$$C_p = \frac{UCL-LCL}{6\sigma} = \frac{255-185}{6(12.7)} = 0.918 \text{ //} \quad \text{(better than 0.84)}$$

Now

$$\text{nonconforming} = 2\left(1-\Phi\left(\frac{UCL-\mu}{\sigma}\right)\right) = 2\left(1-\Phi\left(\frac{35}{12.7}\right)\right)$$

$$= 2(1-\Phi(2.756)) = 2(1-0.99705) = 0.59\% \text{ //}$$

Problem 3 $\bar{x}$ & S chart

- Sample size $n = 4$

$\bar{x}$ chart:		S chart:	
UCL = 710		UCL = 18.08	
CL = 700		CL = 7.979	
LCL = 690		LCL = 0.0	

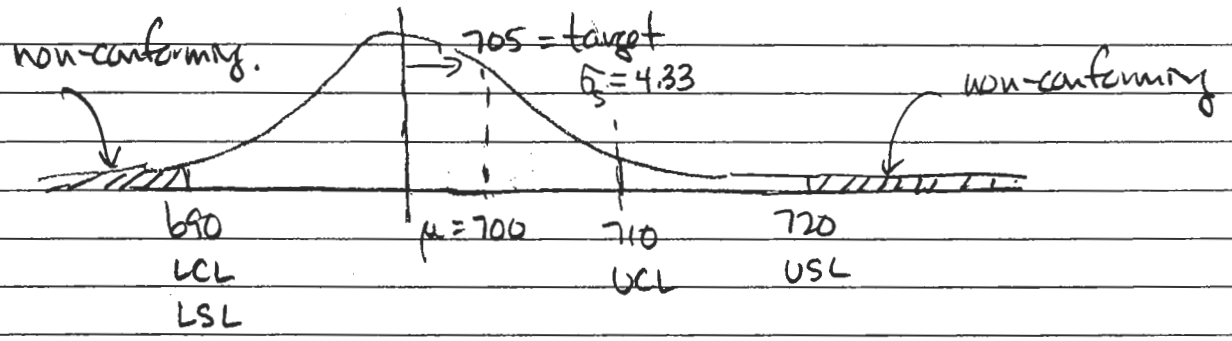
Part a: Estimate μ and σ

- $\hat{\mu} = \bar{\bar{x}} = 700$ //
- Assuming that CL of S chart was based on an estimated value for the process variation $\bar{s}$, we have

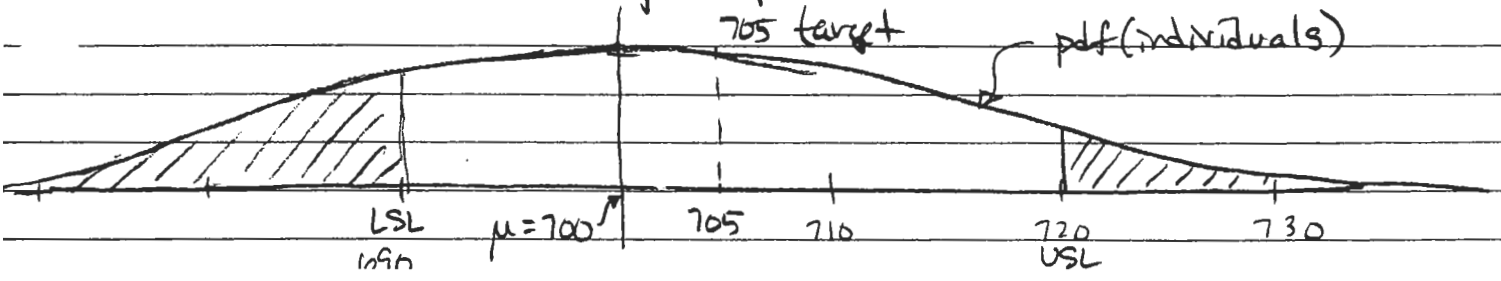
$$\hat{\sigma} = \frac{\bar{s}}{c_4} = \frac{7.979}{0.9213} = 8.66 \text{ //}$$ using $c_4 (n=4)$ from Appendix

Part b For spec limits at 705 ± 15 and assuming normally distributed output, find fraction nonconforming:

- One might be tempted to draw a picture such as



The above picture is WRONG! The UCL & LCL are based on the $n = 4$ sampling distribution, not the individual part distribution. The right picture is (with $3\sigma = 26$)



- The out of spec region (shaded above) gives:

$$\text{fraction non-conforming} = \Phi\left(\frac{LSL-\mu}{\sigma}\right) + 1 - \Phi\left(\frac{USL-\mu}{\sigma}\right)$$

$$= \left[1 - \Phi\left(\frac{700-690}{8.66}\right)\right] + \left[1 - \Phi\left(\frac{720-700}{8.66}\right)\right]$$

$$= \left[1 - \Phi(1.1547)\right] + \left[1 - \Phi(2.3095)\right]$$

$$= (1 - 0.8749) + (1 - 0.98956)$$

$$= 0.1251 + 0.01044$$

$$= 0.1355$$

or 13.6% nonconforming.

Part c Assuming $\hat{\sigma} = 8.66$ as previously estimated, the α error of the $\bar{x}$ chart is

$$\bar{x} = 700 \pm k\hat{\sigma}/\sqrt{n} = 700 \pm 10$$

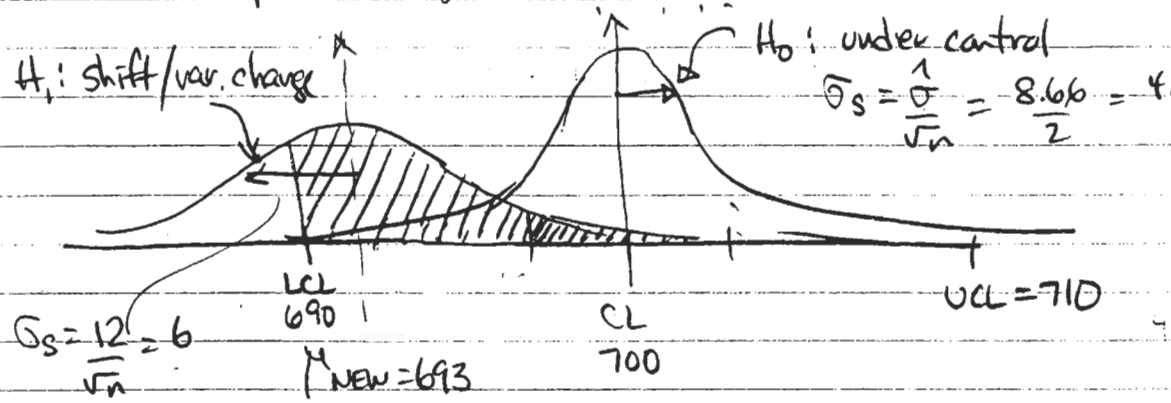
or $k\frac{\hat{\sigma}}{\sqrt{n}} = 10$ so $k = \frac{10\sqrt{4}}{8.66} = 2.31 = z_{\alpha/2}$

So for a $\pm 2.31\sigma$ control chart, we have

$$1 - \alpha/2 = \Phi(k) = \Phi(2.31) = 0.98956$$

False alarm prob: $\alpha = 2(1 - 0.98956) = 0.0209$ or 2.09%

Part d Mean shifts to 693 and σ moves to 12 (from 8.66). Find probability of detection on $\bar{x}$ chart on 1st sample.



Power $\triangleq 1 - \beta$

We need $\Pr(\text{trigger} \mid H_1) = 1 - \Pr(\text{don't trigger} \mid H_1)$

shown in shade on previous page.

That is, probability of a miss is the probability that given H_1 is true, the point is still within the H_0 control limits, or

$$\beta = \Pr(\text{miss}) = 1 - \Phi\left(\frac{696-693}{6}\right) = 1 - \Phi(-0.5)$$

$$= \Phi(0.5) = 0.691$$

$\therefore$ Power $= \Pr(\text{detect}) = 1 - \beta = 0.309$ //

ARL to detect $= \frac{1}{\text{Power}} = \frac{1}{1-\beta} = 3.24$ runs //

Part e

$\Pr(\text{type II error}) = \beta$ for a shift of μ to 693 & σ to 12,

$= 0.691$ as determined above. //

Problem 4

- Show individuals & moving range chart based on above data.

For MR chart, we use $n=2$ and form virtual samples to find $\overline{MR}$ & $\bar{x}$ from the data:

$\bar{x} = 1.611$

$\overline{MR} = 0.0024$

For $n=2$, we have factors
$d_2(2) = 1.128$
$D_4(2) = 3.267$
$D_3(2) = 0$

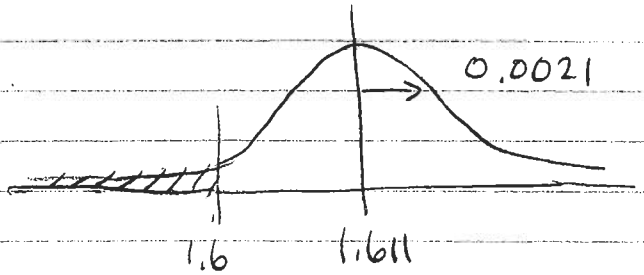
- Moving-Range chart:
$UCL = D_4 \overline{MR} = 0.0078$
$CL = \overline{MR} = 0.0024$
$LCL = D_3 \overline{MR} = 0$ //

- Individual chart:
$UCL = \bar{x} + 3\frac{\overline{MR}}{d_2} = 1.617$
$CL = \bar{x} = 1.611$
$LCL = \bar{x} - 3\frac{\overline{MR}}{d_2} = 1.604$ //

- Our estimates for mean & std. dev. of process:

$\hat{\mu} = \bar{x} = 1.611$ and $\hat{\sigma} = \frac{\overline{MR}}{d_2} = 0.0021$ μm //

- Percentage rework needed?



$\text{Rework Prob} = \Phi\left(\frac{1.6 - 1.611}{0.0021}\right)$

$= 1 - \Phi(5.15)$

$= 1.33 \times 10^{-7}$ OR

$= 133 \times 10^{-9} = 133$ ppb (parts per billion)

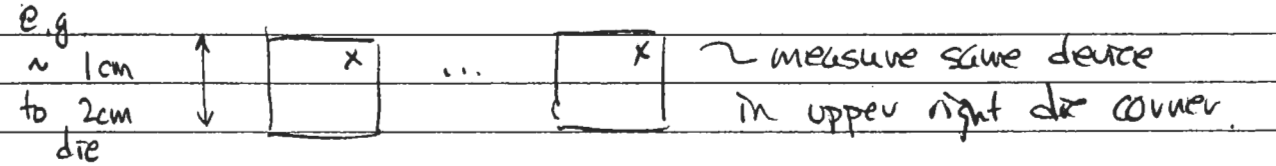
Problem 5

This is a hard problem - I hope you spent some time thinking the possibilities and issues through. Let me start with some definitions and assumptions.

- As written, the problem does not clearly distinguish between lot to lot and "batch" variation. Usually a "lot" is 24 or 25 wafers that travel together in a single wafer carrier. (Side note: the lot size is likely to be different for 300mm fab lines!) Some fab steps aggregate multiple lots into a larger group or "batch" for batch operations, e.g. a high temperature furnace step may run 50 or 100 wafers at a time.

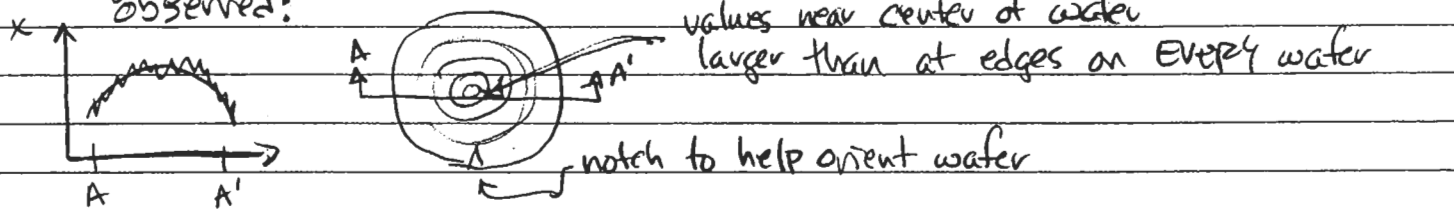
  In this discussion we will assume that no additional batching is done - that is, we consider the batch size to be identical to the lot size.

- The second assumption we will make is that the same "critical dimension" (CD) is measured on the same device within the die whenever a measurement is called for, e.g.

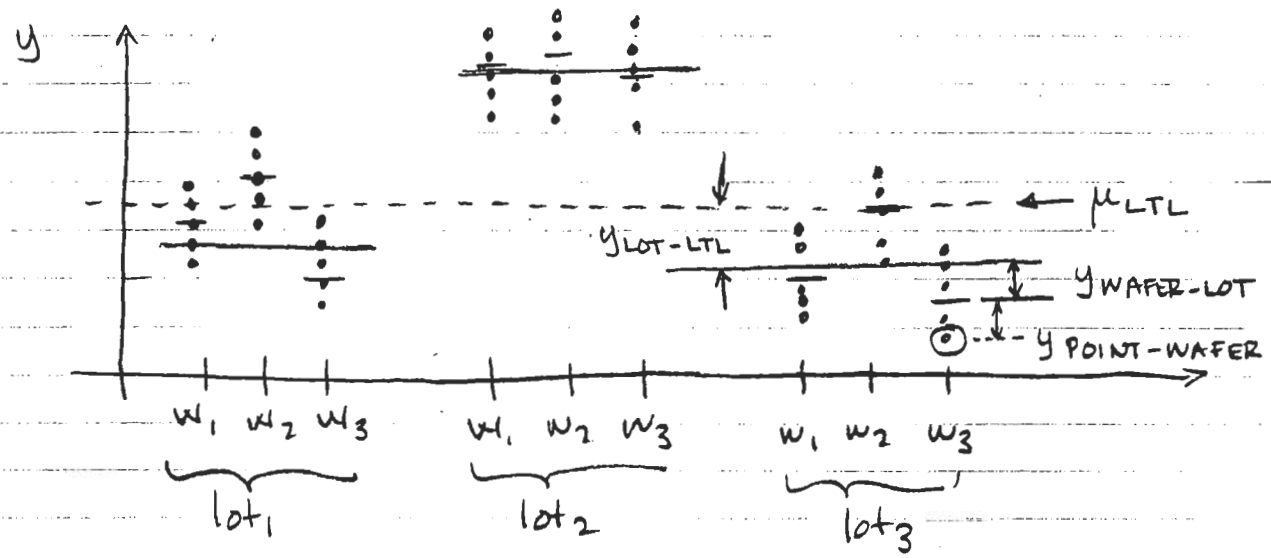


  This assumption allows us to ignore any additional or systematic within-die variation.

- For purposes of education, I will consider the case where within-wafer variation may be fairly complicated. In this case, suppose there is both <u>systematic</u> wafer-scale dependencies and <u>random</u> within-wafer fluctuations. This is quite common, e.g. a "bowl" or "bull's eye" pattern is often observed:



- The following picture illustrates some key variation components arising from the problem scenario. Imagine we have 3 wafers in each lot, and measure 5 pts each wafer.



Some features to note:

- μ_{LTL} is the overall lot to lot mean (or grand lot mean)
- $y_{LOT-LTL}$ is the difference between any one lot average and the grand lot mean
- $y_{WAFER-LOT}$ is the difference between any one wafer average and the within lot mean
- $y_{POINT-WAFER}$ is the difference between a specific measurement point and the current wafer mean

So we can think of any particular measurement y as

$$y = \mu_{LTL} + y_{LOT-LTL} + y_{WAFER-LOT} + y_{POINT-WAFER} \quad \cdots$$

where $y_{LOT-LTL} \sim N(0, \sigma^2_{LTL})$ ... lot to lot variance

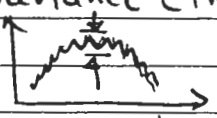
$y_{WAFER-LOT} \sim N(0, \sigma^2_{WTW})$ ... wafer to wafer var.

- That is, we will assume that there is no systematic dependence of wafer on location within the lot.

We can further decompose the $y_{POINT-WAFER}$ based on the fact that within the wafer we have both systematic (fixed) and random variation:

$$y_{POINT-WAFER} = f(position) + y_{POINT-POINT} + \varepsilon_{MEASUREMENT}$$

where $f(position)$ is a deterministic function of where the measurement is taken on the wafer

$y_{POINT-POINT} \sim N(0, \sigma^2_{WIW})$ ... random within wafer variance (the "noise" on )

$\varepsilon_{MEASURE} \sim N(0, \sigma^2_M)$

This last element is the random error from the measurement tool. We will assume $\varepsilon_{MEASURE}$ is negligible.

- Now we can ask what variance in our data point y is composed of. Assuming each component is independent of the other:

$$Var\{y\} = Var\{\mu_{LTL} + y_{LOT-LTL} + y_{WAFER-LOT} + f(pos) + y_{POINT-POINT}\}$$

$$\sigma^2_y = \cancel{Var\{\mu_{LTL}\}} + \sigma^2_{LTL} + \sigma^2_{WTW} + \cancel{Var\{f(pos)\}} + \sigma^2_{WIW}$$

$Var\{\mu_{LTL}\} = 0$ – constant mean

$Var\{f(pos)\} = 0$ – fixed for any particular measurement point

Part a Single measurement on a wafer selected at random from each lot.

- In this case, we take our measurement from the same location on the wafer every time (e.g. at wafer center). This ensures $Var\{f(pos)\} = 0$ for a fixed position.

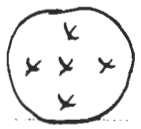
Thus our measurement y (e.g. circled point on figure of previous page) contains three variation components: σ^2_{LTL}, σ^2_{WTW}, and σ^2_{WIW}.

- What control charts to use?

Since we only have individual measurements we might choose an x and moving range (MR) chart. The x control limits should be based on σ^2_y, that is, the sum of lot to lot, wafer to wafer, and within wafer variation. An MR would estimate this lumped variance.

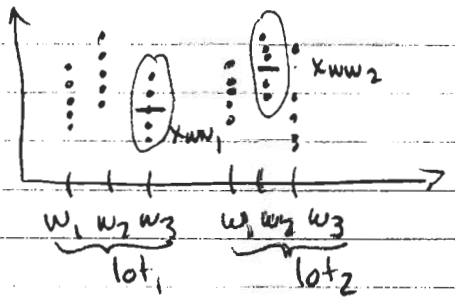
Note that an alarm would only indicate that the lumped variance or a mean shift somewhere had occurred; we do not have enough resolution to figure out which component (mean or variance of LTL, WTW, or WIW) has changed.

(Part b) Now we measure 5 points on the wafer instead of a single point, i.e.



and calculate x_{WW} as 5 pt. average and R_{WW} as 5 pt. range.

- The question is not clear, but I will assume we still sample one wafer at random from each lot, e.g.



Now $x_{WW} = \frac{1}{5}\sum_{i=1}^{5} y_i$ captures variance as

$$\sigma^2_{x_{WW}} = \sigma^2_{LTL} + \sigma^2_{WTW} + \frac{\sigma^2_{WIW}}{5}$$

In this case, the systematic within wafer variation may

result in a constant offset (the same for every wafer), but we get the benefit of a five point sample to reduce the effect of within-wafer random variance.

The exact nature of R_{WW}, the 5 point range, is a little tricky. Because it contains systematic elements it cannot directly estimate σ^2_{WIW}. However, an R chart would monitor just the within wafer random variation. Imagine we measure our 5 points on wafer i: $y_1 \ldots y_5$. Here W_i indicates the wafer to wafer, lot to lot, and lot mean parts which are the same for each y_j.

$$y_1 = W_i + 27 + N(0, \sigma^2_{WIW})$$
$$y_2 = W_i + 7 + N(0, \sigma^2_{WIW})$$
$$y_3 = W_i + 3 + N(0, \sigma^2_{WIW})$$
$$y_4 = W_i + 17 + N(0, \sigma^2_{WIW})$$
$$y_5 = W_i + 8 + N(0, \sigma^2_{WIW})$$

($f(\cdot)$)

Then $R(y_1 \ldots y_5) \approx (27-3) + R(\text{due to } \sigma^2_{WIW}, n_p = 5)$.

Thus an R chart would likely show $\bar{R}_{WW}$ having both 24 & random variance elements in it, and σ_{RWW} would just be due to σ^2_{WIW}.

$$UCL = \bar{R}_{WW} + 3\sigma_{RWW}$$
$$CL = \bar{R}_{WW}$$
$$LCL = \bar{R}_{WW} - 3\sigma_{RWW}$$

So a shift in within-wafer variance should trigger the R_{WW} chart, but shifts in wafer mean or lot mean would not trigger this chart.

(Part c) Measure 1 point on every wafer, (we'll use $n_W = 3$ consecutive wafers rather than $n_W = 5$ as mentioned in the problem set so we match with my pictures.)

- $X_{BW} = \frac{1}{n_W} \sum_{j=1}^{n_W} y_j$ – an average within the lot

$R_{BW} = \text{range}\{y_j\}$ – range within the lot

So now

$$\text{Var}\{X_{BW}\} = \sigma^2_{XBW} = \sigma^2_{LTL} + \frac{(\sigma^2_{WTW} + \sigma^2_{WIW})}{n_W}$$

Reduction by # wafers in lot sampled.

- Necessary to run separate charts for each point?

Answer: No. We could further reduce the impact of within wafer variation by averaging n_p points (e.g. $n_p = 5$) from each wafer first; in which case we would have

$$\sigma^2_{\bar{x}_{BW}} = \sigma^2_{LTL} + \frac{(\sigma^2_{WTW} + \sigma^2_{WIW}/n_p)}{n_w}$$

Consider the limit as $n_p \to$ large, and $n_w \to$ large. Then we would effectively have a control chart just to capture the lot to lot variation.

(Part d) Randomly selected sites on the wafer: would this change our answer to part c?

Answer: No. BUT we could be susceptible to including systematic spatial variation in our variance component, e.g.

$$\sigma^2_{\bar{x}_{BW}} = \sigma^2_{LTL} + \frac{(\sigma^2_{WTW} + [\sigma^2_{f(pos)} + \sigma^2_{WIW}]/n_p)}{n_w}$$

Indeed, often $\sigma^2_{f(pos)} \gg \sigma^2_{WIW}$, so our control chart now may have hugely expanded control limits due to this component, making it even harder to detect other small changes.

In both part d and part c, R_{BW} will also be expanded by the systematic variation, making it hard to interpret in detail. However, R_{BW} would block out σ^2_{LTL} and be only sensitive to the wafer to wafer and lower level components. As in part (b), we would primarily pick up wafer to wafer (between wafer) and "lower" components of variance. (By "lower" I mean also the within wafer variance).

(Part e) What would we recommend for batch-to-batch variability monitoring?

- Here, I'll think of "batch to batch" as lot to lot variability.

The recommendation would be to take as many measurements as economically feasible to separate components. Begin by trial data collection to estimate σ^2_{WIW} (using fixed spatial sampling patterns), σ^2_{WTW}, and σ^2_{LTL}. One would estimate in that order; once we know σ^2_{WIW} then we can estimate σ^2_{WTW} (using multiple wafer data); then we can estimate σ^2_{LTL} (using multiple lot data).

Based on those estimates, we could then decide how many points n_p or n_w wafers are needed to reduce those elements of variance and let us monitor σ^2_{LTL} with as tight limits as possible.