ANALYZING THE COST OF REWORK

by David McGoveran

(This paper was originally published in: BTU Users Group Newsletter, No. 2, April 1983. pp.13-21, BTU Engineering, Billerica, MA. It has been reformatted and a few typographical errors corrected.)

I. INTRODUCTION

This article presents an elementary analysis of the financial impact of rework in the semiconductor FAB. Both empirical data and theoretical calculations indicate that the reworked yield (good die per wafer) drops off exponentially with the number of rework cycles. While it is true that the specific parameters involved are unique to a given FAB and product, it is possible to provide a general analysis.

The cost of wafer production is dependent upon a great many variables: equipment depreciation, labor (both direct and indirect), materials costs (e.g. raw materials and DI water), step yields, and die yields. In order to examine the cost of reworked wafers, it is not only necessary to take all these variables into account, but also to consider the impact that rework has on the use of critical resources. These costs are often overlooked when the decision to rework is made.

Several key relationships are established in this article: (1) the relationship between rework and yield; (2) the relationship between yield and wafer value; (3) the relationship between rework and wafer value; (4) the relationship between rework and capacity; (5) the relationship between rework wafer cost, wafer value, rework rates, and line position. These functions represent a family of curves or surfaces. We provide some analytical insight into these complicated relationships below.

II. RESOURCE USAGE CALCULATIONS

We begin the analysis by examining the amount of time required to process wafers at any given step. From this calculation we can establish the step throughput or units per hour. The step line yields must be known for each step before we can proceed and must be provided empirically. Once the step line yields are known, we can calculate the step related material cost, and the labor added, each adjusted to the step yield. A number of cumulative factors are then calculated: the cumulative material cost, the cumulative labor added, the direct overhead, the indirect overhead, and finally the direct wafer cost. We examine the die yield, taking rework into account and establish relationships between the FAB throughput in wafers per hour, the yield, the profit margin, and the cost of rework.

The hours required to process each wafer through a given operation (e .g. machine) is the inverse product of the cycle time, the bath size (wafers per cycle), the percentage of total shift time that the machine is available for use (percentage uptime) and the percentage of machine uptime that the machine is actually used {percentage utilization). Because we normally measure machine downtime, the machine percentage uptime is calculated as the arithmetic inverse of the machine percentage downtime.

$$HPU_m = \sum_{o} [l / (U/C_o * CT_o * UT_o * (1-DT_o)] \qquad (II.1)$$

where

m = step number
o = operation number
HPU = hours per unit
DT_o = % downtime for the equipment involved
UT_o = % utilization of the equipment (% time used)
U/C_o = units per cycle
CT_o = minutes cycle time for one batch on the equipment used including wafer handling associated with loading, unloading, and transferring wafers between pieces of equipment.

In calculating the HPU for the rework portion of a given process step, one must sum over the equipment used in rework and add the HPU for any additional operations. This quantity is added to the normal HPU for each rework cycle that the wafer goes through. Thus, the photoresist rework cycle for HPU calculations includes spin, softbake, align and expose, develop, inspect, and, additionally, photoresist strip.

$$RHPU_{pr} = HPU_s + HPU_b + HPU_a + HPU_d + HPU_i + HPU_e \qquad (II.2)$$

where

RHPU = HPU for rework
pr = photoresist work
s = spin
b = softbake
a = align and expose
d = develop
i = inspect
e = photoresist strip

Note that the HPU for reworked wafers applies not only to the wafers that are actually reworked, but also to those wafers which are in the same batch. This is due to the difficulty of adjusting such upstream operation rates as the etch rates for a partial batch so that it will match the processing of the unreworked batch remainder. Thus the unreworked wafers in the batch must waH for the reworked wafers to catch up. Some FABS do not follow this philosophy, however, and therefore in calculating the step HPU one must be careful to do a weighted average based upon the correct wafer quantities. Thus, the reworked wafer quantity for HPU calculations is the number of wafers which are reworked plus those which waH for reworked wafers. Obviously, Ha FAB does not use the "catch-up" method when doing rework, this second wafer quantity will be zero.

The total number of wafers per hour that can be processed through any operation (machine), process step, or process line is the inverse of the hours required per wafer processed through the operation, step, or line, respectively. The number of wafers per hour is a routinely calculated productivity statistic. Note that the total hours required for processing through the line must be calculated first; that is, this productivity statistic is dependent on the definition of the process line segment over wh1ch productivity is to be measured . In particular, this is due to the fact that the sum of the inverses is not the same as the inverse of the sums.

The total machine capacity is just the product of the wafers processed per cycle, the cycle time, and the uptime. The used machine capacity is the product of the total machine capacity and the utilization. This quantity represents processing of both reworked and non-reworked wafers. The unused capacity is just the product of the total machine capacity and the arithmetic inverse of the utilization.

$$C_o = U/C_o * CT_o (l - DT_o) * UT_o \tag{II .3}$$

$$CU_o = U/C_o * CT_o * (l - DT_o) * (l - UT_o) \tag{II.4}$$

$$WC_o = C_o + CU_o = U/C_o * CT_o (l - DT_o) \tag{II .5}$$

where

C_o = used capacity for the operation

CU_o = unused capacity for the operation

WC_o = total capacity available for the operation in wafers per batch.

III. STEP YIELD CALCULATIONS

The step yield is the number of wafers out divided by the number of wafers entering the step. This number is not representative of the step without taking rework into account more carefully. Ideally, one would like to measure line yield down to the operation level. However, this is not generally practice and so we do not complicate the calculation in this article.

$$FY_m = WO_m / WI_m \tag{III.1}$$

where

FY_m = step yield for step m

WO_m = good wafers out of step m

WI_m = total wafers entering step m

The rework step yield for a given wafer is approximately the $(n+l)^{th}$ power of the step yield where n is the number of rework cycles the wafer passes through. Since the degree of handling and the opportunity for catastrophic processing errors is increased proportionally for each pass through the step, the FAB yield drops geometrically.

$$RFY_{m,i} = FY_m * (FY_m)\ N_i \tag{III.2}$$

where

RFY_m = rework step yield for step m

N_i = number of rework cycles required for wafer i to pass inspection. Note that N_i is 0 for unreworked wafers.

We calculate a normalized step yield as a weighted average of the individual wafer step yields. For a given period of time (step cycle time), the normalized step yield is just the sum of the individual normalized wafer step yields divided by the total number of "wafers" processed through the step.

$$NFY_m = \sum_i (RFY_{m,i}) / (WC_m - DW) \tag{III.3}$$

where

NFY_m = normalized step yield

WC_m = step wafer capacity

$$DW_m = WC_m - WI_m + \sum (N_i) \text{ available capacity}$$

The step rework rate is defined as the percentage of the total wafers processed at that step that must be reworked. Reworking a wafer more than once requires adding to the total wafer count as well as the rework wafer count. (We refer only to major rework - not redeveloping as rework here.)

$$RR_m = (\sum_i (N_i)) / [WI + \sum_i ((N_i - l))] \tag{III.4}$$

where

RR_m = number of reworks/number of wafers that could have been processed normally.

When a station is capacity limited, the available capacity must be taken into account.

$$RR_m = (\sum_i (N_i)) / [WC_m - DW_m] \tag{III. 5}$$

Note that the adjusted wafer capacity for the step must not be negative, as indicated by the definition of capacity given above.

The cumulative FAB yield is the product of the step yields. Rework may be taken into account by using a weighted average in calculating the step yields.

$$CFY = \prod_m NFY_m$$

$$= \prod_m [(l - RR_m) * FY_m * (RR_m * (FY_m) N + l)] \tag{III.6}$$

IV. CALCULATING COSTS

We can now begin calculating the costs associated with wafer processing. At any given step, the yielded material cost is the material cost divided by the step yield.

$$MC_m = MC_m / NFY_m \tag{IV.1}$$

where

MC_m = material cost at step m

MCY_m = material cost at step m yielded

The cumulative material cost yielded is the cumulative material cost yielded from the previous step divided by the current step yield.

$$C(MC)Y_m = \sum_m (MCY_m) \qquad \text{(IV.2)}$$

where

$C(MC)Y_m$ = the cumulative yielded material cost up through step m.

The labor added is given by the hourly pay rate divided by the yielded units per hour and the step yield . Note that the yielded units per hour is just the step units per hour divided by 1 plus the rework rate. However, the labor added must be adjusted by the ratio of production hours to hours paid, denoted by SC here .

$$UPHY_m = UPH_m \; / \; (l + RR_m) \qquad \text{(IV.3)}$$

where

$UPHY_m$ = the number of wafers per hour yielded at step m.

$$LAY_m = \$HPR_m \; / \; NFY_m * UPHY_m * SC_m \qquad \text{(IV.4)}$$

where

LAY_m = the labor added at step m

$\$HPR_m$ = the hourly pay rate at step m

SC_m = percentage of time the equipment is covered by operators

= ST_m/PT_m

PT_m = production hours available per shift

ST_m = shift time in hours

The cumulative labor added is just the current step labor added plus the cumulative labor added through the previous step.

$$C(LA)Y_m = \sum_m LAY_m \qquad \text{(IV.5)}$$

where

$C(LA)Y_m$ = cumulative labor added yielded through step m.

The direct overhead up to any step is given as a percentage (the direct overhead rate) of the cumulative labor added at that step.

$$\$DOH_m = C(LA)Y_m * \$DOHR \tag{IV.6}$$

where

$\$DOH_m$ = indirect overhead through step m
$\$DOHR$ = indirect overhead rate.

We also account for the indirect overhead which is given by

$$\$IOH_m = C(LA)Y_m * \$IOHR \tag{IV.7}$$

where

$\$IOH_m$ = indirect overhead through step m
$\$IOHR$ = indirect overhead rate.

The total overhead is then

$$\$OH_m = \$IOH_m + \$DOH_m \tag{IV .8}$$

where

$\$OH_m$ = total overhead through step m.

The direct wafer cost up to any step is the sum of the cumulative material cost yielded, the cumulative labor added yielded, the cumulative direct overhead, and the cumulative indirect overhead.

$$\$DWC_m = \$C(MC)Y_m + \$0H_m \tag{IV .8}$$

where

$\$DWC_m$ = direct wafer cost through step m.

V. PRODUCTIVITY ANO COST/BENEFIT CALCULATIONS

The GDPH is a key productivity measure for a process line, given by the wafers per hour out times the die yield. However, the rework die yield is different from the normal die yield. for example, a reworked wafer sees an extra masking step in the process. Although not all the defects accumulated in the first pass are passed on to the second, the additional time spent in stripping the wafer all but "compensates" by exposing the wafer to more contaminants roughly doubling for every minute exposure in a class 100 environment. Thus, we must calculate a normalized die yield (which is determined by the effective number of masking steps each wafer goes through) before we can calculate GDPH. This may be determined by taking the sum of the total number of aligns where the wafer survives without catastrophic error divided by the number of wafers which survive through any given masking step.

$$n' = \sum_{m} [(NFY_m * WI_m + RFY_m * WI_m) / NFY_m * WI_m] \qquad (V.1)$$

The expected die yield can now be calculated using the effective number of masking steps.

$$Y = B^{n'} * C \qquad (V.2)$$

where

Y = good die per wafer
B = percent yield

The percent yield B is given by:

$$B = 1 / (1 + FDtd^2) \qquad (V.3)$$

where

d = die size on a side in mm (calculation assumes 0.2mm between die)
F = fraction of active or fatal area
t = time in Class 100 environment in minutes
D = defect density in defects/square mm/minute

The number of die per wafer unyielded is just:

$$C = [(A' / (d + 0.2)^2] * K \qquad (V.4)$$

where

A' = useful area of the wafer in square nm
K = fraction of whole die sites exposed (K = l for 10:1; K = 0.87 for 1:1 scan)

The total number of hours per unit is estimated by calculating the cumulative hours per unit and taking the additional time due to rework into account at each step, using the appropriate step rework rates.

$$THPU_m = \sum ((1-RR_m)*HPU_m = (RR_m*RHPU_m)] \qquad (V.5)$$

where

$THPU_m$ = total hours processing per wafer through step m.

The total number of wafers per hour is then just the inverse of this number.

$$WPH_m = 1/THPU_m \qquad (V.6)$$

Note that the cycle time of the FAB is increased by rework.

We now calculate the GDPH as the die yield calculated with n* masking steps multiplied by the number of wafers out of the last step per hour.

$$GDPH = K(B)^{n'} * C * WPH_e \qquad (V.7)$$

where

WPH_e = the wafers per hour throughput of the last step.

Having obtained a die productivity statistic, the average cost per good die is then calculated as:

$$\$/GD \;+\; WPH_e * \$DWC_e/GDPH \qquad (V.8)$$

where

$\$DWC_e$ = the final direct wafer cost
$/GD = the cost per good die.

We can estimate the cost of doing rework then by calculating the $/Good 0ie using different rework rates and effective number of mask levels. The decision as to whether it is beneficial to rework a given wafer is then a question of the total added costs in terms of expected $/Good 0ie versus the present $/Good Die. If this quantity is greater than the $DWV of the wafer to be reworked, there is not net benefit, but a net loss and the wafer should be scrapped rather than reworked.

$$SRC = \$/GD_r - \$/GD_u - SDWC_r \qquad (V.10)$$

where

SRC = cost of rework

$\$/GD_r = \$/\text{Good Die with reworking allowed}$

$\$/GD_u = \$/\text{Good 0ie without reworking}$

$\$DWC_r = \text{direct wafer cost of the wafer to be reworked.}$

However, this does not take into account the relative value of wafers to be reworked at each masking step in the process. The value of a wafer at any given point in the line is inversely proportional to the remaining expected wafer cost at that point.

$$\$DWV_m = k*[l / (\$DWC_e - \$DWC_m)] \qquad (V.11)$$

where

k = some constant factor dependent on profit margins, overhead rates, etc.
$/DS = sale price per packaged die.

For a given per die market value (including package and final test costs), the value per wafer is just GDPW * Price/die . The profit margin is then the unit sale price minus the cost per die for processing, packaging and final test.

$$\$PMPD = \$/DS - \$/GD - \$PKG - \$FT \qquad (V.12)$$

where

$PMPO = profit margin per die
$FT = final test costs per die
$PKG = packaging cost per die.

The differential profit margin per die between reworked and non- reworked die is just the absolute value of the difference between the sums for good die and reworked die.

$$\$DPM = \sum_{r} (\$/DS - \$/GD_r - \$PKG - \$FT) - \sum_{\mu} (\$/DS - \$/GD_{\mu} - \$SPK - \$FT) \qquad (V.13)$$

where

$DPM = the differential profit margin
r = die from reworked wafers only
u = die from unreworked wafers only

VI. SUMMARY

As expected, the decision to rework is dependent upon a number of factors then. Clearly, reworking decreases the per unit profit margin. The relative benefit of reworking when the line is not capacity limited at a potential rework step is driven by the expected differential per unit profit margin; this in turn is dependent upon the normalized rework rate (as defined above), the number of times the average wafer has been reworked (and thus the average number of masking steps a given wafer sees}, and the increase in cumulative FAB yield due to a decrease in step yield for masking steps where rework is done.

Having laid an analytical foundation in this issue, the next issue of the BTU User's Group Newsletter will present an example, clarifying the relationships and demonstrating the method by which the tools developed here may be applied. In addition, we will point out some ways in which Bruce System software tools can be of aid in analyzing rework in your FAB.