Lecture 29: Productivity and process yield

Contents

1 Introduction

Process and device evaluation in the fab is used to make sure that the fabrication goals are met, at the individual process level. These goals are usually defined as a *process window* for each step in the fabrication process. This could be a maximum level for defect density and for a growth process a range of acceptable thickness values. Semiconductor IC manufacture is a complex process, with hundreds of individual steps. Errors in even one step have the potential to dramatically affect the functioning of the final product. Even one *killer defect* can cause device failure. The goal of the manufacturing process is to improve the **yield** of the process. There are three major yield measuring parameters

1. **Wafer fabrication yield** or **fab yield** - this is defined as the ratio of the total number of wafers that come out of the fab (after the end of all the individual processes, including measurement) to the total number of wafers that were started in the fab.

Table 1: The major yield measurement points in IC fabrication. There are three yield parameters, wafer yield is defined in the fab while sort and packaging yield are defined for processes after the fab.

Major yield measuring points	
Manufacturing stage	**Measured**
Wafer fabrication yield	$\frac{\text{Wafers out of fab}}{\text{Wafers started in the fab}}$
Wafer sort yield	$\frac{\text{Functioning dies on wafer}}{\text{Total dies on wafer}}$
Packaging yield	$\frac{\text{Packages passing final testing}}{\text{Good dies started for packaging}}$

2. **Wafer sort yield** - this is the ratio of the number of good (functioning) dies (after electrical testing) to the total number of dies in the wafer.

3. **Packaging yield** - this is the ratio of the total number of packaged dies that pass final electrical testing (e-test) to the number of good dies that are started for packaging. This measures the yield of the packaging process, which is a series of steps similar to fabrication process.

The major yield points are summarized in table 1. The different ways the yield points are related is shown in figure 1.

2 Fab yield

Fab yield or wafer fabrication yield measures the output of the wafer fabrication process. It is a *cumulative yield*, since wafers go through a number of individual steps within the fab. For each step or process, it is possible to define a yield parameter. This is called a **station yield** and is defined as

$$\text{Station yield} = \frac{\text{number of wafers leaving the station}}{\text{number of wafers entering the station}} \tag{1}$$

Thus, the fab yield is defined as the product of the individual station yields, since the output of each step is the input for the next

$$\text{fab yield} = \text{station yield 1} \times \text{station yield 2} \times \text{station yield 3} \times ..... \tag{2}$$

For an overall high fab yield, the individual station yields must also be high. Also, as the number of fab processes increase, the fab yield must decrease

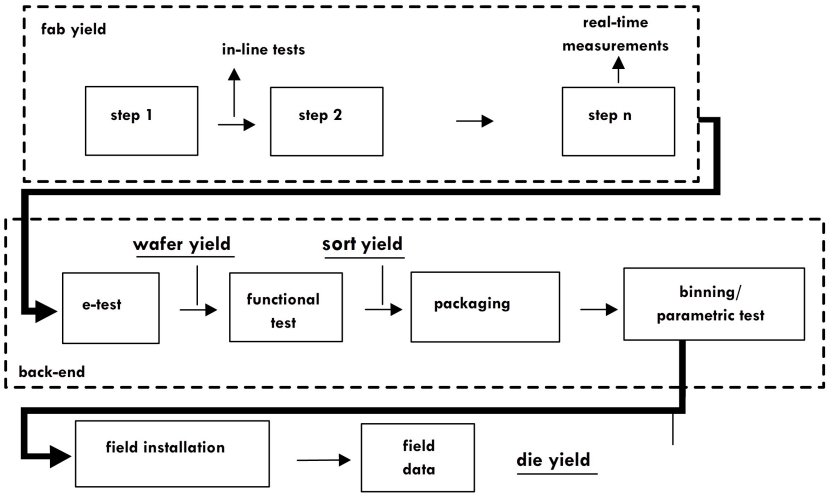

Figure 1: The three yield points and their relation to IC process flow. Fab yield is related to the individual station yields in the fab. Sort yield is related to e-test and die yield is after packaging. Adapted from *Semiconductor manufacturing and process control - May and Spanos.*

since the station yield can never be larger than 1. Consider a sample ten step process as shown in table 2. These are various steps in the formation of a MOSFET. The individual station yields are all shown in percentages. Even though these yields are all high (each station yield is above 90%), the cumulative yield of the process is only 83.9. This decrease in overall yield will become more pronounced as the number of process steps increases. There are various factors that affect the overall fab yield

1. Number of process steps - for a 50 step process to get 75% yield, the individual steps must each have a yield greater than 99.4%.

2. Wafer breakage and warping - this is related to the handling of the wafers. The handling process is now automated in most commercial fabs. Table 3 lists the number of handling steps in an oxidation process. Each step has the potential to lower yield, by causing wafer breakage. Warping happens when wafers are heated to and cooled from high temperatures e.g. rapid thermal annealing process.

3. Process variation - this is related to the process window for each step in the fabrication. This is also called **spec limit**. If the spec limits are too loose, then the process variation can cause lower yield. On the

Table 2: Ten step process with individual station and cumulative yields, listed in percentages. As we go down the process flow, the cumulative yield decreases since the output of each step is the input for the next

Step	**Wafers in**	**Station yield**	**Wafers out**	**Cumulative yield**
Field ox	1000	99.5	995	99.5
S/D masking	995	97	965	96.5
S/D doping	965	99.3	958	95.8
Gate mask	958	99.0	948	94.8
Gate ox	948	99.0	938	93.
Contact hole mask	938	92.0	862	86.2
Metal layer dep	862	99.0	853	85.3
Metal layer mask	853	100	853	85.3
Alloy metal layer	853	99.5	848	84.8
Passivate	848	99	839	83.9

Table 3: Wafer handling steps in a typical oxidation furnace. Each handling step can cause defects in the wafer affecting yield. The goal is to minimize wafer handling. Adapted from *Microchip fabrication - Peter van Zant.*

Substep	No. of wafer handlings
Wafers are removed from carrier and placed in cleaning boat	2
Wafers are cleaned, rinsed, and dried	1
Wafers are removed from cleaning boat, inspected, and placed on oxidation boat	2
Boat is removed from furnace	0
Wafers are removed from boat and placed back in the carrier	1
Test wafers are removed from carrier and measured	2
Total no. of handlings	8

other hand, too tight a spec limit can lower individual station yield and hence fab yield.

4. Process defects - these are isolated defects that happen during processing. They could be due to equipment issues or incoming wafer issues.

5. Mask defects - these are caused due to defects in the hard mask like dust, cracks in the quartz, damaged Cr layer. Some mask defects are shown in figure 2.

Batch processes should have higher process control (i.e. higher station yield) than serial processes, since more than one wafer is handled at a time. Typically, Si wafers have a higher yield than GaAs, since GaAs is more brittle. For GaAs, even partial wafers are processed in the fab due to their brittle nature, while for Si, partial or broken wafers are discarded.

3 Wafer sort yield

Wafer sort yield refers to the number of functioning (good) dies in the wafer that comes out of the fab to the total number of dies in the wafer. There are a number of factors on which sort yield depends on

1. Wafer diameter - as the wafer diameter increases, number of edge dies decreases. This increases sort yield.

2. Die area - as the individual die area increases, yield decreases. So, larger wafers are more cost effective for larger dies.

3. Number of processing steps - this is related to fab yield in that as the number of process steps increases, the background defect density also increases. The effect of wafer size and die size on the number of wafers is shown in figure 3.

4. Circuit density - this defines the smallest defect size that can cause a wafer to short. Not all defects in a wafer are fatal, but as the circuit density increases, the number of fatal defects increases.

5. Defect density - a high background defect density in the fab can cause dies to fail. This effect is more pronounced for larger dies, as shown in figure 4.

6. Crystal defect density - this relates to the quality of the Si wafer that is used for IC manufacturing. Existing defects like dislocations can be increased during processing, which can affect yield.

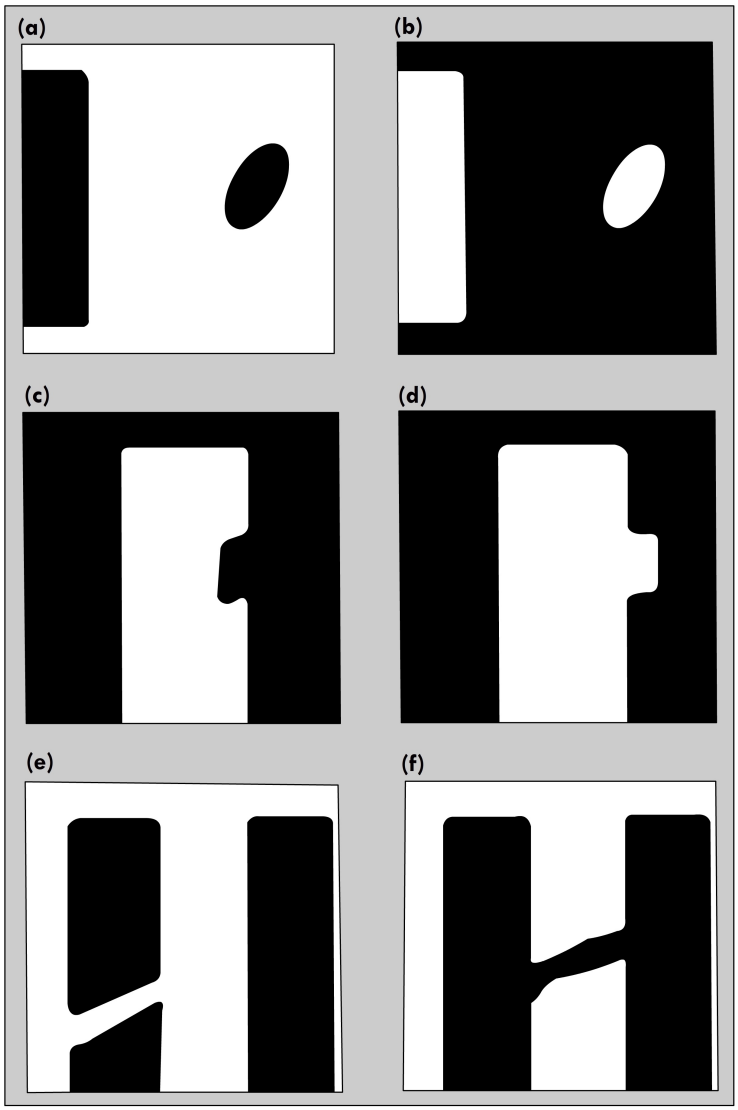

Figure 2: (a) - (f) Typical mask defects that can get transferred to the wafer during lithography. Not all of these defects are killer defects. Adapted from *Microchip fabrication - Peter van Zant.*

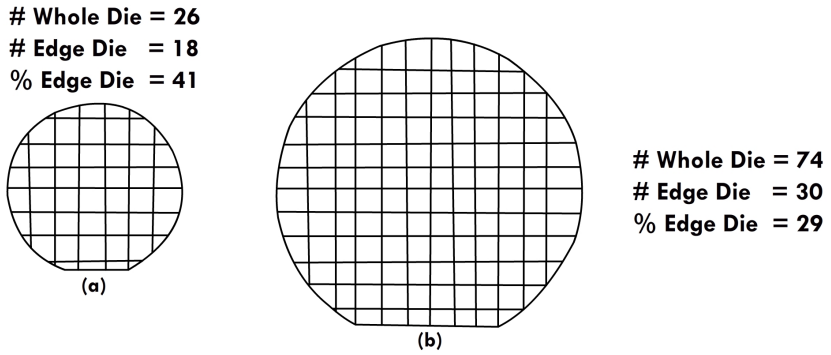

Figure 3: Relation between wafer size and die size effects. (a) Smaller dies have larger number of edge dies (b) Larger wafers have smaller edge dies, for the same die size. This leads to greater yield. Adapted from *Microchip fabrication - Peter van Zant.*

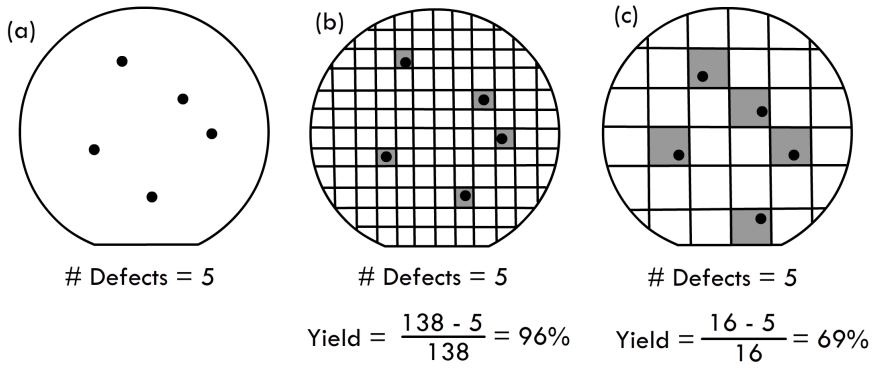

Figure 4: Effect of five random defects, i.e., background defect density on (a) wafer (b) wafer with smaller dies (c) wafers with larger dies. Larger die sizes will have lower yields. Adapted from *Microchip fabrication - Peter van Zant.*

7. Process cycle time - larger residence time for wafers in the fab can increase the defect density. Wafers that are stored in the fab between processes, can get contaminated and lower yield.

Die sizes and defect density are also related. For smaller die sizes, there are more dies per wafer. For the same number of defects, sort yield will be larger if the die size is smaller.
Packaging yield is related to the various steps in the packaging process. After sort, the individual wafers are separated and the packaged. After packaging the dies are again measured. The overall yield is then

$$\boxed{\text{Overall yield } = \text{ wafer-fab yield} \times \text{sort yield} \times \text{Packaging yield}} \quad (3)$$

Thus, all three yield points determine the final output of the IC process.

4 Yield models

Yield models relate the process steps, defect density and chip (die) size parameters to the wafer-sort yield. The defects that are used in the yield models are the *killer defects* (those that cause a die to fail). Hence, the defect density used is smaller than the actual defect density, since not all defects are killer defects.

4.1 Poisson model

The exponential or the Poisson model is the simplest yield model to be developed. Consider a wafer with N dies and n defects that are randomly distributed in the wafer. Then the probability of a given die having k defects is given by a simple Poisson distribution

$$p(k) \; = \; \frac{\exp(-m)m^k}{k!} \quad (4)$$

where m is n/N. Thus, the probability of a die having zero defects (good die) is then given by substituting $k = 0$ in equation 4.

$$p(0) \; = \; \exp(-m) \; = \; \text{Yield } (Y_r) \quad (5)$$

If we define chip defect density (D_0) to be the number of defects per die per unit area (i.e. $n/(N\ A)$) then

$$Y_r \; = \; \exp(-AD_0) \quad (6)$$

where A is the die area.

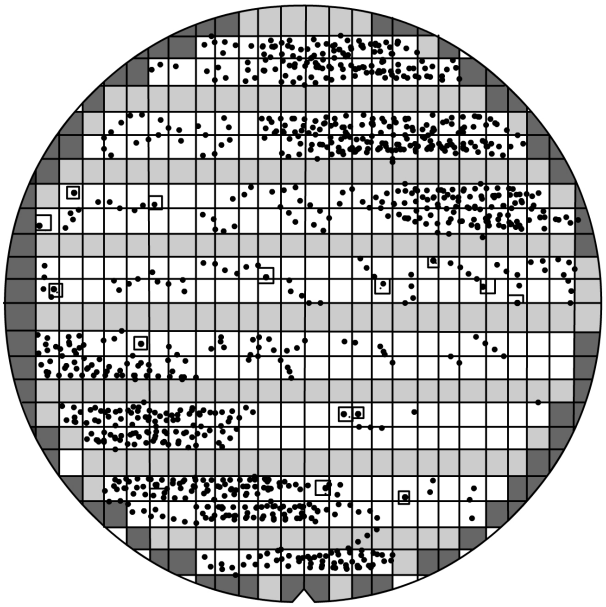

Figure 5: Clustering of defects on a wafer. While there is a random population of defects, there is clustering at the lower left corner and upper right corner of the wafer. This type of clustering happens when a certain process or a subset of processes are responsible for the defects. Source *http://www.geek.com/chips/ibm-toshiba-and-sony-form-32-nm-alliance-561548/*

4.2 Murphy Yield model

Usually, defects are not randomly distributed in the wafer but are clustered at certain locations, depending on the process that causes them. Another term that that is used to describe this is *decoration* of defects. A non-random defect distribution is shown in figure 5. Here, the defects are clustered in the lower left corner and in the top and top right. This type of clustering happens when a certain process or a subset of processes are responsible for the defects. For such situations, it is possible to define an yield integral given by

$$Y_r \;=\; \int_0^\infty f(D)\;\exp(-AD)dD \tag{7}$$

where $f(D)$ is the distribution of defect density. There are various models that define this function, $f(D)$. For the Poisson model, $f(D)$ is a delta function given by

$$f(D) \;=\; \delta\;(D \;-\; D_0) \tag{8}$$

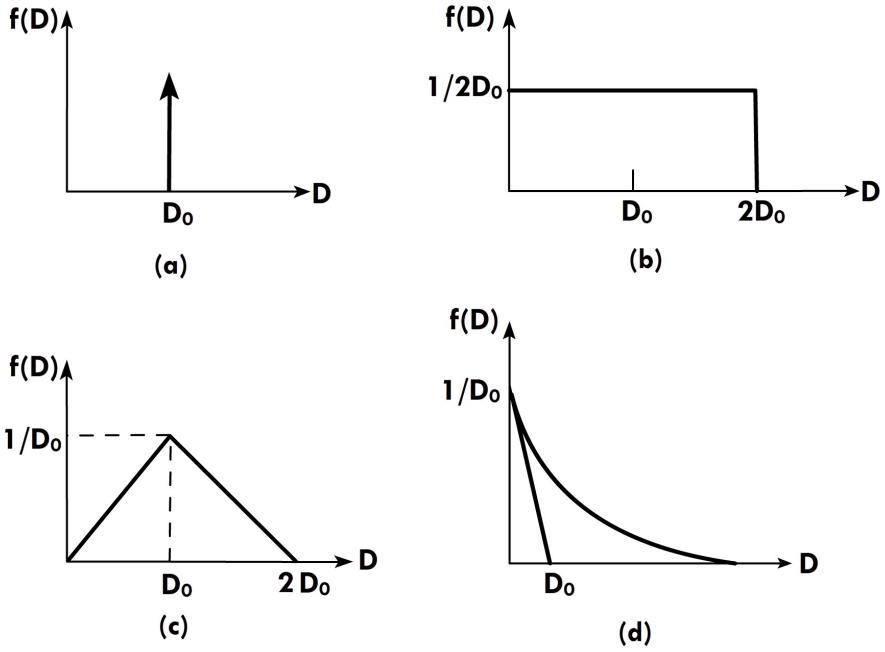

Figure 6: Various defect distribution functions, $f(D)$ vs. defect density. (a) Poisson (b) Uniform distribution (c) Triangular and (d) Seeds model. These models can be used *a priori* to calculate the yield of a process. Adapted from *Semiconductor manufacturing and process control - May and Spanos.*

where D_0 is the average defect density, used in equation 6.
There are different defect distributions, as shown in figure 6. One model, from Murphy, assumes an uniform density function, as shown in figure 6(b). The evaluation of the yield integral for this gives

$$Y_{\text{uniform}} = \frac{1 - \exp(-2D_0A)}{2D_0A} \tag{9}$$

Another yield distribution model is a Gaussian model, which can be approximated as a triangular function to get a closed form solution. This density distribution is shown in 6(c) and gives the yield function

$$Y_r = (\frac{1 - \exp(-AD_0)}{AD_0})^2 \tag{10}$$

The triangular yield model is widely used in the industry for estimating the effect of process defect density.
The Seeds model is based on the fact that high yields are obtained by a

large population of low defect densities and a small population of high defect densities. This is given by a distribution function, as shown in figure 6(d).

$$f(D) \ = \ \frac{1}{D_0} \ \exp(-\frac{D}{D_0}) \tag{11}$$

This gives a yield function given by

$$Y_r \ = \ \frac{1}{1 \ + \ D_0 A} \tag{12}$$

A more general model for the exponential distribution is based on the Bose-Einstein distribution function which gives a yield function

$$Y_r \ = \ \frac{1}{(1 \ + \ D_0 A)^n} \tag{13}$$

where n is the number of mask levels, for lithography defects. This model is used when lithography steps are the critical steps, as far as defect generation is concerned. As die size increases, (i.e. A increases) the yield decreases. This is shown in figure 7. These defects mentioned here are only the killer defects. The overall defect density is higher, since it also involves non-critical defects.

5 Wafer fabrication costs

Yield is an important factor in the semiconductor industry, since it affects the overall cost of the wafer. A high yield is essential to reduce the cost per die (or chip). But, yield is not the only parameter affecting cost. There are other parameters that affect the cost of a die. Fabrication costs can be divided into two main categories

1. **Fixed costs** - fixed costs are those that are present regardless of whether the fabrication facility (fab) is used or not. Typically, these include overhead costs like administration, facilities, research (including personnel salaries), and equipment (most commercial fabrication equipment are very expensive and bought on a lease basis). Other overhead costs include circuit design cost and fab maintenance costs.

2. **Variable costs** - variable costs are those that depend on the amount of chips produced i.e. the volume of manufacture. These include materials, labor (since staff are required to run a production fab), and yield. Material costs can be either material that are directly used for the chip (like Si or GaAs or metals) or indirect materials like masks, reticles, and chemicals, which are used in the various processes.

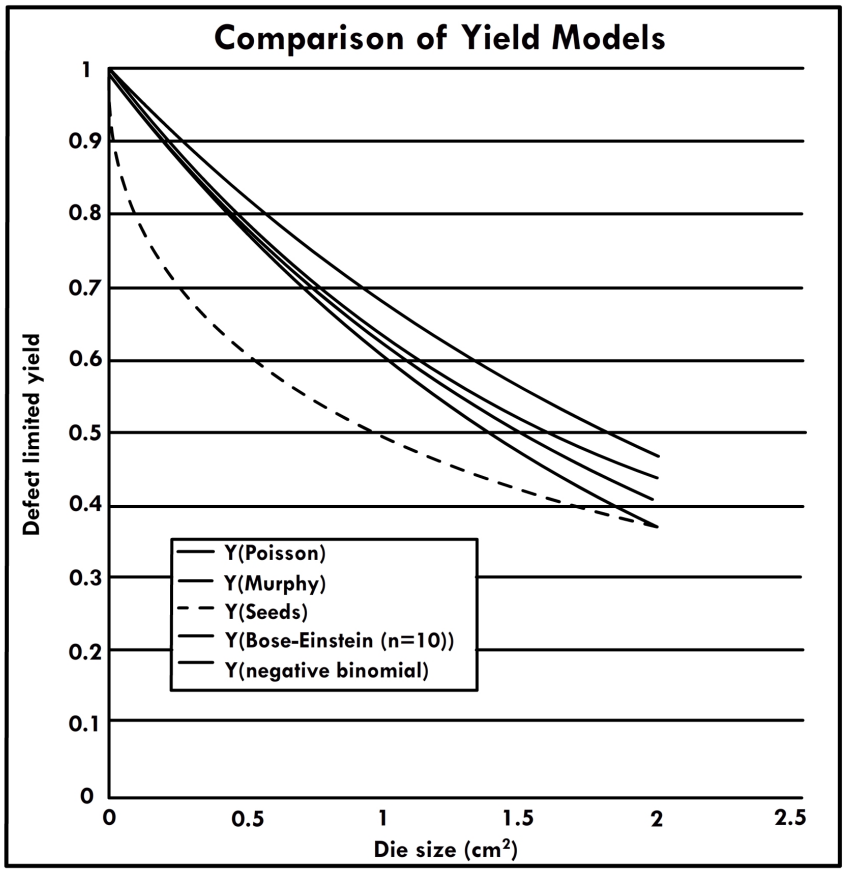

Figure 7: Defect yield density as a function of die size for different models. In all of these the yield reduces with die size. The steepest drop is for the Seeds model, which assumes an exponential dependence for the defect density distribution. Adapted from *Microchip fabrication - Peter van Zant.*

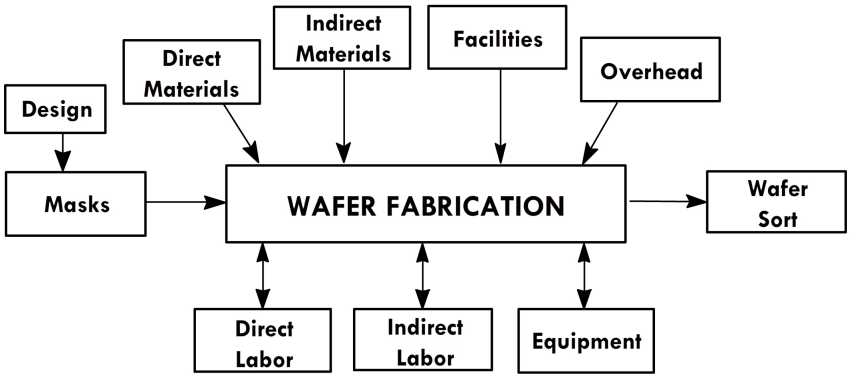

Figure 8: Components of the wafer fabrication cost. Both direct and indirect costs feed into the final die cost. Adapted from *Microchip fabrication - Peter van Zant.*

The fabrication costs are summarized in figure 8.

Because of the huge capital expense in maintaining and running a fab there are very few **integrated device manufacturers** (IDMs). IDMs own and operate their own wafer fabrication facilities and control all operations from design to packaging. The overall cost is high due to the overhead. Intel, IBM, and Samsung are some well known IDMs that are currently operating. Most other semiconductor companies are **fabless semiconductors**. They make the circuit design and contract with a **merchant foundry** for fabrication. AMD and Qualcomm are two examples of fabless semiconductors that contract with foundries like TSMC (Taiwan semiconductor manufacturing company) and Global Foundaries to manufacture their chips. Companies like Samsung, that are IDMs, also do contract work for other companies, most notably Apple.

During the early stages of IC manufacture, 100 mm wafers were used for production. Over time, wafer sizes have steadily increased with 300 mm wafers being currently used and 450 mm wafers under development. The lifetime of the different wafer sizes are shown in figure 9. Larger the wafers, more dies can be manufactured (for the same die area) and hence cost per die is lower. But the transition from one wafer size to the other leads to an increase in equipment cost, since new equipment has to be installed that can handle the larger wafers, a process called *retooling* the fab. In the long run, there is a lower equipment cost per die, for larger wafers. Handling larger wagers is slower, so the number of tools in the fab have to be increased to maintain the same production time. Hence, there is a trade-off between increased fixed

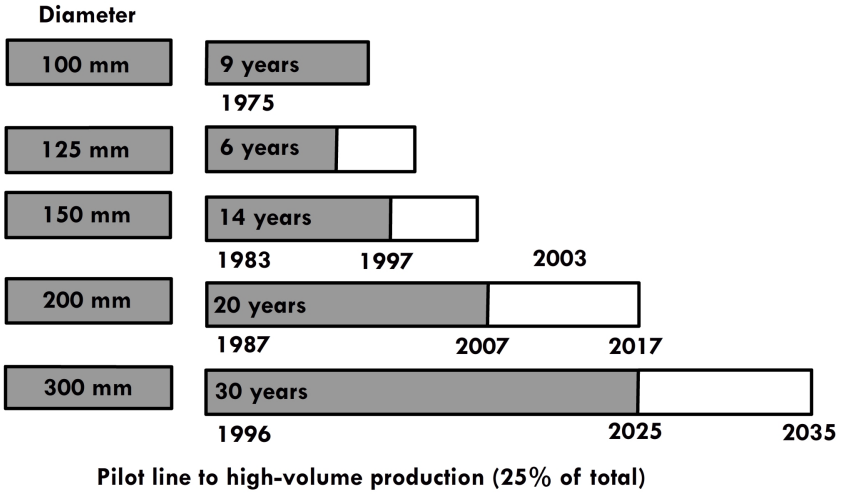

Figure 9: An approximate lifetime of different wafer sizes from the start of IC fabrication. Larger wafers lower cost per die, but there is increased cost during the transition from one wafer size to the next. The next generation is 450 mm wafers, which would be introduced in 2018. Adapted from *Microchip fabrication - Peter van Zant.*

costs and decreased variable costs when wafer size is increased.
Fab yield affects overall cost, since it is a part of the variable cost of manufacturing. If the die yield is low, then the overall cost goes up. The **yielded die cost** takes into account fab yield in calculating cost per die. To give a simple example, if the wafer cost is US $ 3000 and there are 300 dies per wafer, then the cost per die is $ 10. But, if yield is only 50% then the cost per die is $ 20, since only half these dies are usable. This is the yielded die cost. Thus, a higher yield is always better. The fab yield for a particular product depends on its maturity, as shown in figure 10. Initial yield during start and development is low and then yield rapidly increases once the development cycle is mastered. During production, typical fab yields are greater than 90% (this is the cumulative yield of all the individual stations). During the initial development process, yielded die cost is very high and then slowly decreases as yield increases. The development lifetime is usually 2 years, shown in figure 11 for dynamic random access memory chips (DRAMs). While high yield is desirable in a fab, another factor that must also be high is **wafer throughput**. Throughput is defined as the number of wafers processed per hour. This defines the productivity of the fab.
Overall, it is possible to define a *cost of ownership* (C_w) which defines the

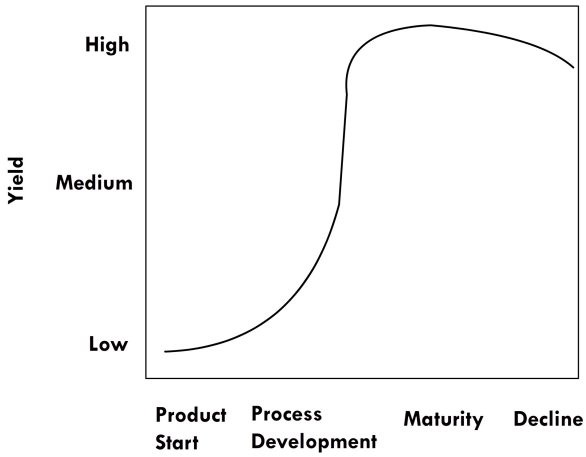

Figure 10: Dependence of fab yield on the maturity of a product. During development, yield starts very low, until the first *yielding die* is reached. After that, yield increase rapidly during development and saturates during maturity. Yield decreases slightly during transition to the next technology, since the number of wafers produced decreases. Adapted from *Microchip fabrication - Peter van Zant.*

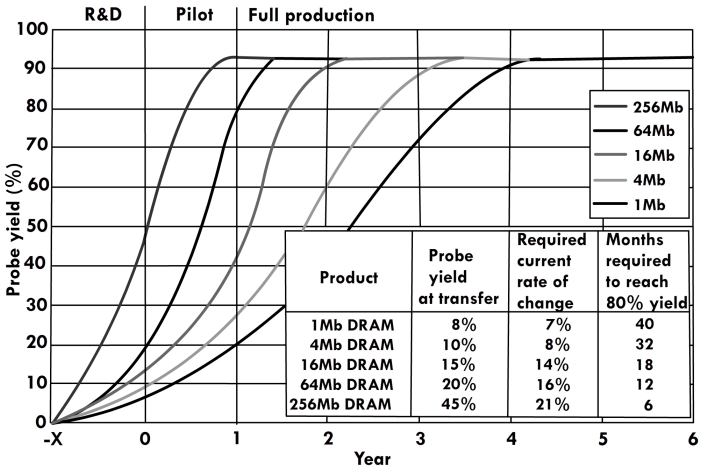

Figure 11: Yield vs. years in development and production for a DRAM product cycle. Typical yields during production are above 90%. Adapted from *Microchip fabrication - Peter van Zant.*

final cost of the finished wafer. This is given by

$$\boxed{C_w = \frac{\$F + \$V + \$Y}{L \times TPT \times Y_{TPT} \times U}} \tag{14}$$

The various factors in equation 14 are as follows

$ F = fixed cost

$ V = material, labor, and process cost when the tool is running

$ Y = cost of wafers scrapped due to defects

L = lifetime of the tool

TPT = wafer throughput

$\mathbf{Y_{TPT}}$ = yield factor

U = tool utilization factor that reduces tool available time from the maximum value

The trade-off between these various factors affect the cost of ownership. Thus, a high yield is desired but not at the cost of wafer throughput. Similarly, larger wafers can decrease wafer cost and material cost, but can lead to higher fixed cost and lower throughput. Thus, yield plays an important role in deciding overall die cost but it is only one of the parameters that decides cost.