Fault Clustering in deep-submicron CMOS Processes

Jan Schat, NXP Semiconductors Germany

Abstract

The fraction of ICs that pass all production tests but fail in the application is called the defect level. Defect levels depend on the average number of defects per IC, and also on the clustering of these defects. High clustering leads to a higher yield and a lower defect level.

This paper compiles the coefficients for defect clustering using research findings from 1970 until 2001. Because recent data for deep submicron processes are missing in the literature, the clustering coefficient has been calculated using scan fail distributions of ICs in a 180 nm process. Clustering coefficients show a steady trend towards higher defect clustering. This is beneficial, but it is probably not sufficient to achieve today’s ambitious target of ‘zero defects’.

1 Introduction

As stated in the 2005 edition of the International Technology Roadmap for Semiconductors (ITRS), chipmakers face a steady increase in “Zero Defects per Million” requirements [1]. Each defective part that escapes production test and is shipped to the customer causes enormous costs to the manufacturer. Hence the defect level is an economic value that needs to be considered before the production of any IC type in order to evaluate its profitability. Furthermore, manufacturers must not only make various attempts to reduce defect levels, but also to predict them.

A defect denotes a physical imperfection leading to a device malfunction; for example, shorting a signal to GND or VDD caused by a particle, an open connection caused by an open via, a connection from gate to channel caused by a defective gate oxide, etc. In this paper, a defect includes what is sometimes called a fatal defect. A fault denotes the electrically observable effect of a defect, i.e. stuck-at 0, excessive signal delay, etc. A defect can cause more than one fault, e.g. if a particle is big enough to cause a contact between several wires. Defect level means here the relative number of test escapes (i.e. the number of devices that are not functional, but pass the production test). Hence the defect level does not cover early lifetime failures.

It has long been observed that defects in ICs are not randomly distributed, but are clustered. This defect clustering is believed to be due to the fact that clusters of particles from the manufacturing machinery reach the wafer surface before they have a random distribution in the gas or liquid used for processing [2].

There are many reasons why quantifying defect clustering has long been an important concern:

- Investors planning to put money into new equipment, or a new fab must know the extent of the defect clustering as well as the defect density in order to accurately calculate the expected yield and thus the profitability of their investment.
- Given the cost of each defective part that is shipped to the customer, the defect level is important in terms of calculating the profitability of a new device. At a given defect density, higher defect clustering means a lower defect level.
- Defect clustering at a given defect density influences not only the yield and the defect level, but also the reliability [3], because lifetime failures are often attributed to the same kind of defects as those reducing the yield.
- Production test patterns are usually designed to detect single faults. Defect clustering implies, however, that frequently there are multiple faults. This has implications on test coverage because not only is there a risk of two faults neutralizing each other, but also there is a chance that a fault which would be non-observable as a single fault, becomes observable in the presence of another fault. Hence defect or fault clustering may either increase or decrease the probability that a faulty IC is recognized as defective [4].
- Fault tolerance and self-repair, which are often used in memories, can drastically increase yield. Defect clustering, however, may lead to too many defects for an IC to be repairable. This must be taken into consideration when predicting the bottom line yield of fault-tolerant ICs [5].
- Clustering of defective dies on a wafer is also widely discussed. Sometimes so-called ‘good dies in bad neighborhood’ (dies passing all tests, but being surrounded by bad dies) are screened out because they are believed to pose a reliability or test escape risk [3].

In all of the cases above, given a constant mean defect density, defect clustering is beneficial. Ideally, we would have total defect clustering, so that all the defects of one wafer are concentrated in one die, thus securing optimal
yield, high detectability of this one defective die, and high reliability of all other, defect-free dies.

While all these issues have been valid since the early days of IC production, some of them have become increasingly important as customer requirements for low defect levels continue to increase. Because defect levels are referred to in many contracts between semiconductor suppliers and their customers, detailed investigation of defect clustering is an important issue.

Such investigations, however, are very expensive, and are hampered by the fact that the yield and defect level for a given chip area and process are proprietary and heavily guarded. Therefore in writing this paper, it has been necessary to gather whatever sparse data are available from relevant literature to get an indication of both the extent and the trend of defect clustering in current processes.

1 Defect statistics

Since the mid-seventies, the standard model has been the negative binomial distribution of the number of defects per IC [7], hence assuming a locally varying defect density. This model has proven to comply well with measured data [5-12].

Following this distribution, De Sousa and Agrawal combined defect clustering and multiple faults per defect to fault clustering [8], and have come up with a concise and practical two-parameter formula:

\[ Y_T (T) = (1 + \frac{\lambda}{T})^{-\alpha} \]  

with \[Y_T\]: apparent yield (devices passing the test) 
\[T\]: test coverage 
\[\alpha\]: fault clustering parameter 
\[\lambda\]: average number of faults per IC 

The clustering parameter \(\alpha\) can range between 0 (which means extreme clustering) and \(\infty\) (which means no clustering, so that the Poisson distribution is obtained).

2 Defect level

In the negative binomial distribution, the probability that an IC has exactly \(k\) faults, \(P(k)\), can be calculated [6] as

\[ P(k) = \frac{\Gamma(k + \alpha)(\lambda / \alpha)^k}{k!\Gamma(\alpha)(1 + \lambda / \alpha)^{k+\alpha}} \]  

\(\Gamma\) is the gamma function, i.e. the generalization of the factorial to real arguments. Obviously, \(P(0)\) is the fraction of ICs without fault, also known as the yield.

A hypothesis that is confirmed e.g. in [4], is: In an IC with multiple faults, the (average) fault coverage of each fault is independent of the fault coverage of other faults, so that with a coverage \(T_i\) for single faults, the coverage for \(T_i\) for \(k\) faults is

\[ T_k = 1 - (1 - T_i)\;^k \]  

\(k\): number of faults 
\(T_i\): fault coverage for exactly 1 fault 

The defect level \(DL\) can then be calculated as the ratio of all defect ICs shipped to all good ICs shipped. Hence:

\[ DL = \sum_{i=1}^{\infty} \frac{P(k=i)(1-T_i)}{P(k=0)} \]  

And since \((1-T_i)\) is rapidly decreasing with \(i\), and \(P(k=i)(1-T_i)\) is largest by far for \(i=1\), (in other words, mainly one-fault ICs contribute to the defect level), so that

\[ DL \approx \frac{P(k=1)(1-T_1)}{P(k=0)} \]  

Figure 1 shows the defect level at a constant \(\lambda\) (average number of defects per IC), but different defect clustering coefficients \(\alpha\). The test coverage is assumed to be \(T_i = 99\%\). As can be seen, variations in \(\alpha\) severely influence the defect level.

3 Prior work

Several papers have already evaluated defect clustering using field data [4-7,9-12,14-17], however before comparing the results, some differences must be borne in mind:

- Some authors evaluate all ICs but others throw away ‘gross defect ICs’ before evaluating the rest [9,11,12,15]. They define ‘gross defect’ ICs to be those with ‘too many’ defects (e.g. more than 40 defects in [9]). Not considering those ICs is founded on the assumption that these ICs do not suffer from defect clustering, but from other defect mechanisms such as scratches. ‘Gross defect ICs’ usually account for some percent of all ICs.

- Authors mostly derive the clustering coefficient \(\alpha\) from tester data, so they actually deal with faults and not with defects. On the other hand, in [7,9,11,14], particles – and thus potential defects – were detected using optical methods, so these defect statistics are not directly comparable with the fault statistics of other authors.
The defect clustering coefficient $\alpha$ can be derived either by fitting eq. (1) to real tester data, or by using the equation how it is defined for the negative binomial distribution, namely [11]:

$$\alpha = \lambda / (\sigma_{\lambda}^2 - \lambda)$$

(6)

This can lead to deviations like in [7], where $\alpha$ was computed to 0.62 using the first method, and 0.68 using the second method.

The clustering parameter $\alpha$ is generally believed to be independent of the chip area, but in practice for very large chip areas a noteworthy dependency has been found [11]: $\alpha$ was found to vary from 0.38 to 0.73 depending on the area.

The results of the literature survey are summarized in Table 2. The publication date given in the second row is an indication of the process used, which is likely to be a ‘typical mature’ process for that time. If the same data had already been published earlier (as e.g. in [7] and [15]), the earlier publication date is given. Of course the process itself is far more interesting than the publication date, but it is given only in the papers [5,9,16] as 0.5 µm, 1.0 µm and 1.25 µm, respectively. By communication from the author of papers [10,15] these processes are known [18], namely 3.5 µm and 0.85 µm. The third row indicates if faults (from electrical analysis) or defects (from optical inspection) were considered. For papers where $\alpha$ was calculated using different areas or algorithms, the whole span of the calculated values is given. To present a complete picture, the average number of defects or faults, $\lambda$, is also given, although this number is of course directly proportional to the (unknown) chip size.

To the authors’ best knowledge, no study compiling all these values for $\alpha$ has been previously published. Also, [5,15] seem to be the most recent studies, dealing with ICs in a 0.5 µm [5] and 0.85 µm, while today’s mainstream processes are 65 nm to 180 nm. This paper therefore is intended to close these two gaps, as well as raise general awareness of the importance of defect clustering.

4 Data processing

To obtain results from present deep sub-micron CMOS processes, wafer test data from an IC in a 180 nm process were used. Data of several wafers from several lots were taken. To prevent mix-ups with other defect sources like scratches, the wafer maps were inspected manually, and only wafers with a random-like wafer map were considered so that particles were likely to be the dominant source of defects. For the same reason, an IC type with a mature production status was used, so that timing problems or other parametric issues could be excluded from contributing to the faults.

Test-data feedback from the test center included the failing (scan) test, and the first vector failing in this test. This can easily be transformed to the $Y_a(T)$ needed to calculate $\alpha$ and $\lambda$ using (1). The stuck-at test coverage $T$ for each pattern, as needed in (1), was obtained from the ATPG tool that generates the scan test patterns.

5 Results

The values $Y_a(T)$ have been found to comply well with (1) and thus with the negative binomial distribution. The maximum error (eq. (1), calculated $Y_a$ to observed $Y_a$) was only 0.3%. Gross defects were not considered explicitly, however before the scan test an $I_{def}$ test was performed, in order to eliminate a large proportion of the ‘gross defect’ ICs. The calculated values for $\alpha$ and $\lambda$ were $\alpha = 0.08$, and $\lambda = 1.2$. This is an $\alpha$ lower than reported in all previous literature, but in line with the trend of decreasing $\alpha$ with smaller processes. These values of $\alpha$, $\lambda$ and the test coverage $T_1$ (99.1 %) lead, according to eq. (5) to a defect level of approximately 0.18 %. This is only approximately 1/3 of the defect level that would be expected at the same yield, but without defect clustering.

The actual defect level rate can be estimated, because some automotive customers return all failing parts for detailed analysis. A fraction of these are indeed test escapes, i.e. ICs that passed all tests on the tester, but failed in the application. The number of these ICs is compared to the number of ICs shipped to each customer, and the ratio is then the observed defect level. The exact defect level is proprietary data and cannot be given here. It can be said, however, that it is much lower than the 0.18% calculated here.

Figure 2 shows the historical trend of $\alpha$ as given in Table 2. The solid line shows ten values of fault clustering, (from the literature) obtained by tester measurements, as well the value (far right) from work carried out in the course of this paper. These values show a clear decline with smaller processes. The bold straight line is an approximation of the 11 values.

The slope indicates that $\alpha$ is halving every seven to eight years. The correlation coefficient $R^2$ is 0.83, which

<table>
<thead>
<tr>
<th>Ref.</th>
<th>[7]</th>
<th>[6]</th>
<th>[13]</th>
<th>[4]</th>
<th>[16]</th>
<th>[10]</th>
<th>[11]</th>
<th>[14]</th>
<th>[12]</th>
<th>[9], SRAM</th>
<th>[9], Test chip</th>
<th>[5]</th>
<th>[17]</th>
<th>[15], Sematech</th>
<th>[15], DELCO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faults or&lt;br&gt;defects</td>
<td>Defects</td>
<td>Faults</td>
<td>Faults</td>
<td>Faults</td>
<td>Faults</td>
<td>Defects</td>
<td>Faults</td>
<td>Faults</td>
<td>Defects</td>
<td>Faults</td>
<td>Faults</td>
<td>Faults</td>
<td>Faults</td>
<td>Faults</td>
<td>Faults</td>
</tr>
<tr>
<td>Process</td>
<td>1.25&lt;br&gt;µm</td>
<td>3.5 µm</td>
<td></td>
<td>1.0&lt;br&gt;µm</td>
<td></td>
<td>0.5 µm</td>
<td></td>
<td>0.85 µm</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.68&lt;br&gt;...</td>
<td>0.62</td>
<td>1.89</td>
<td>2.382</td>
<td>0.98&lt;br&gt;...</td>
<td>2.28</td>
<td>1.7</td>
<td>0.748&lt;br&gt;...</td>
<td>0.8</td>
<td>0.38&lt;br&gt;...</td>
<td>1.18</td>
<td>1.8</td>
<td>0.487</td>
<td>0.23</td>
<td>./</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>2</td>
<td>0.86</td>
<td>2.333</td>
<td>0.18&lt;br&gt;...</td>
<td>6.5</td>
<td>2.8</td>
<td>2.6</td>
<td>0.85</td>
<td>22.1</td>
<td>2.06</td>
<td>1.98</td>
<td>1.96</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Table 2: Clustering parameter $\alpha$ and faults per IC $\lambda$ as found in the literature |
Figure 2 Clustering parameter $\alpha$ and trend line

means - considering the heterogeneity of the clustering measurements - a relatively narrow distribution around the trend line. As mentioned earlier, a smaller clustering parameter at a given defect density means a smaller defect level and a higher yield.

The dashed line shows three values of defect clustering obtained by optical inspection. Even though the three values are far too few to show a trend, it is obvious that they do not exhibit the same trend that can be seen in fault clustering. A purely hypothetical explanation could be that this might be due to an increasing number of faults per defect.

6 Conclusion

It is almost unanimously agreed in the literature that faults and defects adhere to the negative binomial distribution, hence besides the mean number of faults, only one parameter is interesting: the clustering parameter $\alpha$. Two things about this parameter are important: its value for current processes, and its expected value for future processes.

Despite its importance, fault clustering has received relatively little attention in current literature. Moreover, important information like the process used is nearly always missing, which makes it difficult to precisely compare results. Also, missing literature values for deep-submicron processes have meant that we have had to carry out our own measurements with an IC in a 180 nm process. Using these, together with the values from the literature, we have been able to draw a trend line for $\alpha$ in past and present processes, which indicates that defect clustering within ICs is increasing with every new process generation. We like to point out that this observation needs to be backed up with more data, preferably pertaining to new processes.

Knowing the trend of $\alpha$ might be helpful when making decisions regarding shrinks or smaller processes, because $\alpha$ influences both the yield and the defect level.

The absolute value of $\alpha$ for current processes is important because as described above, clustering has been found to reduce the defect level to 1/3 of what would be expected without clustering, but with the same yield. This is a nice benefit, but it is unlikely to be enough to meet today’s ambitious target of “zero defect parts per million”.

References

[18] E-mail from Vishwani D. Agrawal to Jan Schat, 18. September 2007