A CUSUM Control Chart to Monitor Wafer Quality

Sheng-Shu Cheng and Fong-Jung Yu

***Abstract*—**C-control chart assumes that process nonconformities follow a Poisson distribution. In actuality, however, this Poisson distribution does not always occur. A process control for semiconductor based on a Poisson distribution always underestimates the true average amount of nonconformities and the process variance. Quality is described more accurately if a compound Poisson process is used for process control at this time. A cumulative sum (CUSUM) control chart is much better than a C control chart when a small shift will be detected. This study calculates one-sided CUSUM ARLs using a Markov chain approach to construct a CUSUM control chart with an underlying Poisson-Gamma compound distribution for the failure mechanism. Moreover, an actual data set from a wafer plant is used to demonstrate the operation of the proposed model. The results show that a CUSUM control chart realizes significantly better performance than EWMA.

***Keywords*—**Nonconformities; Compound Poisson distribution; CUSUM control chart.

I. INTRODUCTION

THE yield has a direct impact on the manufacturing cost, so it is frequently used as an index for the evaluation of the integrated circuit (IC) manufacturing performance. Generally, in IC Manufacturing, the yield is affected by the number of defects on the wafer. The defect count for most IC wafer processes is generally monitored using a C-control chart as a tool for statistical process control. The C-control chart assumes that the occurrence of nonconformities is a Poisson distribution. References [1], [2] mentioned that the distribution of nonconformities on the wafers becomes clustered when the wafer area increases. Reference [3] believed that the clustering of nonconformities in wafers often violates the "independent" assumption of the Poisson model. Reference [4] used a revised c-chart, based on a Neyman Type-A distribution, to correct the error caused by using a conventional c-chart. References [5], [6] found that a negative binomial model is more appropriate to clustering in wafer nonconformities. The control model is also more flexible, because there are two parameters in a negative binomial model and only one parameter in the Poisson process. Reference [7] used fuzzy theory to modify a traditional C-control chart and then used this modified chart to simultaneously monitor nonconformities and their clustering on wafers. During wafer manufacture, a defective item normally has more than one defect. When the numbers of defective items fall in a Poisson distribution and

nonconformities in these defective items fall into other types of distribution, the defects in this process fall in a compound Poisson distribution [8].

Traditional Shewhart control charts are very popular in statistical process control. A major advantage of these charts is that they are relatively sensitive to large shifts in the production process. The exponentially weighted moving average (EWMA) and cumulative sum (CUSUM) control charts are two very effective alternatives to the Shewhart control chart, which can be used when small to moderate-sized shifts are of interest [9]. Reference [10] proposed the use of EWMA Control Charts for monitoring wafer quality, when the failure mechanism is a Negative Binominal Process. There is no certain method for the calculation of the smoothing parameter for an EWMA control chart. Reference [9] suggested that it could be set between 0.05 and 0.25, but the reference value for a CUSUM control chart can be decided using a sequential probability ratio test, whereupon the optimized average run length (ARL) for the out of control state can be determined [11].

In this study, a CUSUM control chart is constructed under the assumption that the failure mechanism exhibits a negative binomial distribution (Poisson- gamma compound distribution) in the production process. An actual data set from a wafer plant is used to demonstrate the operation of the proposed model.

II. LITERATURE REVIEW

In a manufacturing process of wafers, the assumption of randomness for a defect's location on a wafer and the independent relationships for different defects is made (i.e., the probability that the defects fall at any point in the wafer is the same) if a traditional c-chart is used to monitor the defect count, It results in a constant defect density. With this assumption, the probability of the occurrence of k defects in one die of the wafer is as equations (1-2):

$$P(k) = \frac{e^{-\lambda}\lambda^{k}}{k!} \qquad\qquad k = 1.2.3..., \qquad (1)$$

$$\lambda = AD . \qquad (2)$$

where λ is the average number of defects in the die, A is the area of the die, and D is the defect density.

Reference [6] pointed out the Poisson distribution model can predict the yield reasonably accurately when the area of the die is smaller than 0.25 cm². As the area of the die increases, the Poisson distribution model underestimates the average number of defects. Reference [12] proposed the concept of defect density and believed that it is different for dies and wafers. He proposed that the defect density must be expressed by a

S. S. Cheng is with the department of Business Administration, Yu-Da University, 36143, Taiwan (phone: +886-972-017566; e-mail: alpha@ ydu.edu.tw).

F. J. Yu is with the department of Industrial Engineering and Technology management, Da-Yeh University, 51591, Taiwan (phone: +886-938-253430; e-mail: fischer@mail.dyu.edu.tw).

World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:7, No:6, 2013

probability density function $f(\lambda)$. This function is quite different from the Poisson distribution model, which assumes that the defect density is a constant. The probability of the occurrence of k defects in [12]'s compound model is expressed as equation (3):

$$P(k) = \int_0^\infty \frac{e^{-\lambda}\lambda^k}{k!} f(\lambda) d\lambda .$$ (3)

where λ is the average number of defects and $f(\lambda)$ is the probability density function of λ. This model is also called the compound Poisson yield model.

Reference [13] proposed a negative binominal distribution model, which is one of the compound Poisson yield models. This compound Poisson model examines the distribution of defects on wafers from two perspectives:

(1) The number of clusters on the wafer follows a distribution with a defect density function, $f(\lambda)$,

(2) The number of defects in a cluster has another distribution (normally a Poisson distribution),

where λ itself is a random variable. The negative binominal model for the defect density function ($f(\lambda)$) is described by the Gamma distribution, denoted as $Y \sim Gamma(\alpha, \lambda/\alpha)$. The function for the gamma density is expressed as equation (4):

$$f(y) = \frac{(\lambda/\alpha)^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-y(\lambda/\alpha)}$$ (4)

So the compound Poisson distribution model can be expressed as equation (5):

$$P(x) = \int_0^\infty \frac{\lambda^x}{x!} e^{-\lambda} \frac{(\alpha/\lambda)^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-y(\lambda/\alpha)} dy$$ (5)

Then the following negative binominal distribution probability model can be obtained through deduction [14]:

$$P(x) = \frac{\Gamma(\alpha+x)}{\Gamma(\alpha)\Gamma(x+1)} \left(\frac{\alpha}{\lambda+\alpha}\right)^\alpha \left(\frac{\lambda}{\lambda+\alpha}\right)^x \qquad x=0,1,2\dots$$ (6)

where the parameter α is the clustering parameter. When α becomes smaller, clustering becomes more serious.

Since $\dfrac{\Gamma(\alpha+x)}{\Gamma(\alpha)\Gamma(x+1)} = \dfrac{(x+\alpha-1)(x+\alpha-2)\cdots\alpha}{x!} = \begin{pmatrix} x+\alpha-1 \\ x \end{pmatrix}$

, let $p = \dfrac{\lambda}{\lambda+\alpha}$, so the probability density function in equation (6) can be rewritten as equation (7):

$$P(x) = \begin{pmatrix} x+\alpha-1 \\ x \end{pmatrix} (1-p)^\alpha p^x \qquad x=0,1,2\dots$$ (7)

Meaning that the negative binomial distribution itself can be viewed as a compound Poisson distribution that has a random variable, λ, and is distributed according to $Gamma(\alpha, \alpha/\lambda)$. The mean and the variance of a negative binominal distribution can be derived by algebraic operation, as follows equations (8-9) [15]:

$$E(X) = \frac{p\alpha}{1-p} = \lambda$$ (8)

$$Var(X) = \frac{p\alpha}{(1-p)^2} = \lambda + \alpha^{-1}\lambda^2$$ (9)

The negative binomial distribution has two parameters, λ and α. Then the defect count can be modeled and the probability of the yield (p(x)=0) of a die can be calculated using equation (6).

III. CUSUM CONTROL CHART AND ARL

Average run length (ARL) is defined as the expected number of plotted points on a control chart before the first point exceeds the control limits. It is a common index, used to measure the efficiency of a particular control chart. An effective and efficient CUSUM control chart can provide the desired ARL and the value of the control limit in the continuous inspection chart should be selected to produce the desired ARL. The ARL is a criterion used to chart performance in this research. This section describes the features and applications of the negative binomial CUSUM control chart and the calculating method of ARL using the Markov chain approach proposed by [16].

A. CUSUM Control Chart

The CUSUM chart directly incorporates all the information in the sequence of sample values by plotting the cumulative sums of the deviations of the sample values from a target values. It is more sensitive to small and moderate shifts than the Shewhart chart, when used to monitor a manufacturing process. The CUSUM and EWMA control charts are two very effective alternatives and are comparables in terms of the run length performance for many processes. The traditional approach for monitoring process defects involves taking samples at each die with a negative binomial count of defects. If the average number of nonconformities in the process exhibits a negative binomial distribution, the expected value and the variance of the negative binomial distribution are derived using equations (8) and (9), with the known parameters for clustering.

Let x_t (t=1,2,...) be the count of defects obtained from the tth observation in the process, k is a reference value and is a constant when a process is in control, with parameters of average defect count, λ, and clustering, α. The control limit of CUSUM chart is used to detect an increase in the number of defects, λ. accumulates derivations from λ and α that are above target with statistic S_t. It can be expressed as equation (10):

World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:7, No:6, 2013

$$S_t = max\{0, S_{t-1} + x_t - k\} \quad t = 1, 2, ..., \qquad (10)$$

If statistic S_t exceed the decision interval, h, the process is considered to be out of control. Meaning that the process drifts or shifts off the target value, then the CUSUM signals and an adjustment is made to some controllable variable to bring the process back in line. In order to improve the sensitivity of a CUSUM at the beginning of the process, the fast initial response (FIR) or head-start value of the statistic, $S_0 = u$, $(0 \leq u < h)$ is used [17].

The proper selection of the reference values k and decision interval h is quite important, because they have a substantial impact on the performance of the CUSUM. It is usually recommended that these parameters be selected to provide good average run length performance. The reference values k can be obtained using the sequential probability ratio test (SPRT) to test a simple hypothesis, $H_0 : \lambda = \lambda_0$, against the simple hypothesis, $H_1 : \lambda = \lambda_1$, where λ_0 is the λ of in-control state, and λ_1 is the λ of out-of-control state, λ. Reference [11] proposed optimal stopping times for the detection of changes in distributions and showed that the reference value chosen using SPRT can be optimal, in terms of run length performance. Reference [18] proposed a method to calculate the reference value for CUSUM charts, based on Bernoulli and binomial counts. Under the assumption of a negative binomial distribution, the reference values k of a CUSUM control chart can be expressed as equation (11):

$$k = \alpha \left(ln \frac{\lambda_1 + \alpha}{\lambda_0 + \alpha} \bigg/ ln \frac{\lambda_1(\lambda_0 + \alpha)}{\lambda_0(\lambda_1 + \alpha)} \right) \qquad (11)$$

B. ARL Calculation for a NB CUSUM Chart

The performance of control charts used for statistical process control can be described by an ARL. And ARL can be calculated using the Markov chain approach proposed by [16]. For the purpose of monitoring an iid process's outputs, suppose that a one-sided control chart is used to monitor the quality characteristic, X. For a continuous charting statistic, there are m distinct values within the control limits that result in a finite number of possible values for the charting statistic. These values are treated as states of a finite Markov chain and all of the values that exceed the limits are incorporated into the absorbing states. Assuming an iid process ($X_1, X_2,...$) wherein the initial probabilities of the Markov are given, the transition probability matrix is used to determine both the mean and the distribution of the run length of the chart. It is used to evaluate the run length properties of a traditional Shewhart chart, a CUSUM chart and an EWMA chart.

A unified method to find the run length distribution and the ARL of a control chart, based on the finite Markov chain approach, is further described by [19]. In addition, this method yields the variance or standard deviation of the run length as a byproduct. Under the assumption of independence, the method for calculating the ARLs of a one-side CUSUM chart has two steps, as follows:

Step 1. Divide the area of the one-sided control chart into m states and then a sequence of discrete random variables of the charting statistic can be obtained.

Given m (m is a positive integer number) and a decision interval, h, let $S_t(m)$ be a finite-state homogeneous Markov chain on the space, Ω, so $\Omega = \{a_0, a_1,..., a_{m-1}, a_m\}$, where $a_0 < a_1... < a_{m-1} < a_m$ and $\{a_0, a_1,..., a_{m-1}\}$ are the transition states and a_m is an absorbing state. The represented values of these transition states are $a_i = i\,w$, for $i = 0,1,2,...,m-1$. The charting statistic value at time t and its mapping to the Markovian state correspond to:

$$Markovian\ state = \begin{cases} a_0 & if\ S_t \leq 0.5w \\ a_i & if\ (i\text{-}0.5)w < S_t \leq (i+0.5)w \\ a_m & if\ S_t > h, \end{cases}$$

where $w = 2h/(2m-1)$.

Step 2. Based on the charting statistic a Markov chain, $M_t(m)$, that has a finite state space, S, and a transition matrix, $P(m)$, of the form can be expressed as equation (12):

$$P(m) = \begin{bmatrix} R(m) & (I - R(m))1 \\ 0 & 1 \end{bmatrix} \qquad (12)$$

where the sub-matrix, $R(m)$, contains the probabilities of going from one in-control state to another in-control state and is an m x m matrix, the $(I - R(m))1$ is a column vector of probabilities corresponding to the in-control states to the absorbing state, I is the identity matrix and 1 is a column vector of ones.

If $S_t = max\{0, S_{t-1} + x_t - k\}$, and if the statistic, S_t, exceeds the decision interval, h, the process is considered to be out of control and the process is assumed to be initially in state 0. The one-step transition probability, P_{ij}, and the matrix of transition probability, R, from the state, i ($i=0,1,...,m$-1), into a new state, j ($j=0,1,...,m$-1), can be shown as equations (13-15):

$$p_{i,0} = p(x_t - k + iw \leq w/2)$$
$$= 1 - I_{\frac{\lambda}{\lambda+\alpha}}([k + (1/2 - i)w + 1], \alpha) \quad for\ i = ,1,2,...,m\text{-}1 \qquad (13)$$

$$p_{i,j} = p((j-1/2)w \leq x_t - k + iw \leq (j+1/2)w)$$
$$= I_{\frac{\lambda}{\lambda+\alpha}}([k + (j-i-1/2)w + 1], \alpha) - \qquad (14)$$
$$I_{\frac{\lambda}{\lambda+\alpha}}([k + (j-i+1/2)w + 1], \alpha)$$

World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:7, No:6, 2013

for $i = 0,1,2,\ldots,m-1$ and $j = 1,2,\ldots,m-1$,
where I is the incomplete gamma function and $[\bullet]$ is the Gaussian function. Since x_t is an integer random variable the Gaussian function replaces the first parameter in the incomplete gamma function, ie, $[\bullet]$ is used to obtain the integer value.

$$R = \begin{bmatrix} P_{0,0} & P_{0,1} & P_{0,2} & \cdots & P_{0,j} & \cdots & P_{0,m-1} & P_{0,m} \\ P_{1,0} & P_{1,1} & P_{1,2} & \cdots & P_{1,j} & \cdots & P_{1,m-1} & P_{1,m} \\ \vdots & & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots \\ P_{i,0} & P_{i,1} & P_{i,2} & \cdots & P_{i,j} & \cdots & P_{i,m-1} & P_{i,m} \\ \vdots & & \vdots & \cdots & \vdots & \cdots & \vdots & \vdots \\ P_{m-1,0} & P_{m-1,1} & P_{m-1,2} & \cdots & P_{m-1,j} & \cdots & P_{m-1,m-1} & P_{m-1,m} \\ P_{m,0} & P_{m,1} & P_{m,2} & \cdots & P_{m,j} & \cdots & P_{m,m-1} & P_{m,m} \end{bmatrix}$$

$$(15)$$

Let $N(m)$ denote the run-length random variable induced by the finite Markov chain. Then, the ARL can be obtained from equation (16), if a positive integer, m, is given.

$$ARL = E(N(m)) = \pi'_0(m)(I - R(m))^{-1}1 \qquad (16)$$

The term, $\pi'_0 = (1,0,\ldots,0)$, is the probability vector of an initial state with 1, corresponding to a specified state and zeros elsewhere.

IV. CUSUM CONTROL CHART FOR A POISSON GAMMA COMPOUND DISTRIBUTION

In the production process for ICs, a wafer defect detector is used to check nonconformities in the wafer and die, after each process. In order to validate the use of a CUSUM control chart for a Poisson Gamma compound distribution, the data collected by [20] was used. The size of the wafers was six inches and 101 wafers were collected. Each wafer was cut into 198 dies. The area of the wafer was 176.72 cm^2 and the area of each die was 0.73 cm^2. All 101 wafers exhibited clustering, as shown in Table I.

TABLE I
NO. OF NONCONFORMITIES IN SAMPLES OF 101 WAFERS

A	B	A	B	A	B	A	B	A	B	A	B	A	B
1	18	16	50	31	23	46	43	61	26	76	25	91	17
2	22	17	42	32	16	47	58	62	26	77	35	92	27
3	14	18	29	33	20	48	52	63	22	78	25	93	15
4	25	19	22	34	15	49	40	64	19	79	22	94	9
5	49	20	29	35	29	50	20	65	20	80	8	95	7
6	17	21	16	36	19	51	18	66	20	81	135	96	101
7	5	22	11	37	29	52	63	67	27	82	105	97	83
8	38	23	22	38	13	53	37	68	25	83	27	98	27
9	18	24	14	39	68	54	19	69	12	84	18	99	16
10	10	25	21	40	44	55	24	70	14	85	16	100	18
11	11	26	21	41	22	56	33	70	17	86	9	101	14
12	12	27	21	42	34	57	9	72	15	87	36		
13	8	28	19	43	23	58	18	73	8	88	24		
14	21	29	16	44	27	59	33	74	10	89	42		
15	13	30	25	45	14	60	22	75	17	90	17		

A: sample number
B: Number of Nonconformities

A. Calculation of ARL_0 for a NB CUSUM Control Chart

From Table I, one can get the average and the variance for these wafer nonconformities are 26.535 and 421.63, respectively. And the average defect density in the wafer is $D = 0.1501$ per cm^2, the average number of nonconformities for each die is calculated, using equation (2), to be $\hat{\lambda} = 0.1096$ and the clustering parameter $\hat{\alpha}$ is 1.0 from equation (9). Assume that the estimated error in these parameters is negligible and that the in-control process is a NB distribution with parameters, $\lambda_0 = 0.1096$ and $\alpha_0 = 1.0$, and that the shift of the parameter values, $\lambda_1 = 2 * \hat{\lambda}_0 = 0.2192$, the mean shifts to be detected quickly.

The reference value is k=0.1575 from equation (11). Supposing that the decision interval is divided into 2001 zones (m=2000), the ARL_0, when the negative binominal CUSUM control chart is in-control state is obtained using the above ARL calculation method for the upper and lower sides. For the purpose of comparison, the ARL_0 are also obtained by simulation, using 20,000 runs for each different situation. The calculation and simulation results for the different situations are summarized in Table II.

TABLE II
ARL_0 FROM CALCULATION AND SIMULATION

h	ARL0	simulation
3.00	191.73	191.36
3.50	287.46	286.55
3.84	369.35	369.26
4.00	414.95	414.74
4.50	606.16	605.30
5.00	844.75	847.06
5.50	1179.2	1181.44
6.00	1588.1	1625.08

From Table II, it is seen that ARL_0 for the calculation and simulation are almost the same for different widths of control limits and that they increase when the desired control limits increase. The ARL_0 can also be selected to be 370 if the same probability of type I error as the Shewhart control chart is desired. At this time, the control limits, h=3.84 can be chosen for this CUSUM chart for process control.

B. A Comparison of the Performance of the Cusum and Ewma Charts

In general, the use of a CUSUM charting procedure without a head start ($S_0 = 0$) to show the ARL of the CUSUM chart is common, but using the control statistics with a head start ($S_0 > 0$) allows faster detection of a process shift if it occurs at the beginning of the process. Usually, the value of the head start is a half of the decision interval ($S_0 = 0.5* h$) when no sample is taken at the beginning of the process. In order to make a fair comparison between the use or not of a head start, this study

World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:7, No:6, 2013

investigates both the zero state (s_0=0) ARL and the head start with a half decision interval (s_0=h/2) for this real example.

Firstly, it is supposed that the average number of nonconformities in the production process is a negative binominal distribution. The acceptable probability of type I error is 0.002 during the process control. At this time, ARL_0 is 500. The initial state of the process is at the 0-state when S_0=0, or at the (m/2)-th state when S_0=h/2. Secondly, it is assumed that the average number of nonconformities in the wafer increases by 2.5 times during the production process (i.e. from 26.535 to 66.338). At this time, the average number of nonconformities in the dies increases from $\lambda_0 = 0.1096$ to $\lambda_1 = 0.2740$.

Because the number of nonconformities also depends on the clustering parameter, the effect of the clustering parameter, α, on the ARL performance of CUSUM charts is also examined. Supposed there are three different clustering parameter values in a wafer. They are $\alpha = 0.7$, $\alpha = 1.4$ and $\alpha = 4.2$. It is seen that the average number of nonconformities in a wafer increases by 0.462σ, 0.478σ and 0.491σ, for these three different clustering parameters. In the first situation, the initial state of the process is at the 0-state, with S_0=0. The *ARL* algorithm uses the control limit of the previous CUSUM control chart. The average run lengths, denoted as ARL_1, are 35.26, 33.41 and 31.83, for the different reference values, k = 0.177, 0.178, 0.179, and the different upper decision intervals, h = 3.935, 3.686, 3.500. When the process shifts to $\overline{\lambda}_1$ for the respective different clustering parameter values ($\alpha = 0.7$, $\alpha = 1.4$, or $\alpha = 4.2$). If the initial state of the process is at the (m/2)-th state with S_0=h/2, the different parameters and values of ARL_1 can also be determined for different respective clustering parameter values ($\alpha = 0.7$, $\alpha = 1.4$, or $\alpha = 4.2$). These are summarized in table III, for the EWMA model proposed by Yu *et al.* (2011).

TABLE III
ARL_1 OF NB CUSUM & EWMA CONTROL CHART

	ARL_1	k	h	ARL_1	k	h	ARL_1	k	h
A	25.7	0.177	4.04	23.86	0.178	3.78	22.53	0.179	3.60
B	30.41			28.51			27.26		

A: CUSUM (S0=*h*/2)
B: EWMA*: Yu's (2011) EWMA charts

From Table III, it is seen that ARL_1 is 30.41 and 25.70 for the respective EWMA and CUSUM control charts with a probability of type I error of 0.002, an initial state of S_0=h/2 and a clustering parameter of 0.7. If a clustering parameter is 1.4 or 4.2, then ARL_1 will be 28.51 and 23.86 or 27.26 and 22.53 for the respective EWMA and CUSUM control charts with a probability of type I error of 0.002. This demonstrates that the performance of the CUSUM is better than that of the EWMA. All of the situations demonstrate the same characteristic for different combinations of clustering parameters and initial states.

V. CONCLUSIONS

The yield rate has a high impact on the production cost for the semiconductor industry. Therefore, the rapid detection of process variation is vital to cost reduction in process monitoring. This paper provides the calculating formula for a CUSUM's reference values, calculates one-sided CUSUM ARLs using the Markov chain approach and demonstrates the calculation of two-sided CUSUM ARLs, to construct a CUSUM control chart. The data from a real case are also used to illustrate the application of the proposed control chart. The results show that when there is clustering of nonconformities, a negative binomial CUSUM control chart effectively highlights die nonconformities. The performance of the CUSUM for this real case is also better than that of the EWMA control chart, so the negative binomial CUSUM control chart can more effectively monitor wafer nonconformities and reduces the cost of quality control, to allow enterprises in the semiconductor industry to remain competitive.

REFERENCES

[1] S. F. Liu and F. L. Chen, "A data clustering model for wafer yield loss in semiconductor manufacturing." *Journal of the Chinese Institute of Industrial Engineers,* 21(4), pp. 328–338, 2004.
[2] S. H. Tong and B. J. Yum, "A dual burn-in policy for detect-tolerant memory products using the number of repairs as a quality indicator." *Microelectronic Reliability*, 48, pp. 471-480, 2008.
[3] C. H. Stapper, "The effects of wafer to wafer defect density variations on integrated circuit defect and fault distributions." *IBM Journal of Research Development*, 29(1), pp. 87- 97, 1985.
[4] S.L. Albin and D. J. Friedman, "Clustered defect in IC fabrication: Impact on process control charts." *IEEE Transactions on Semiconductor Manufacturing*, 14(1), pp. 36–42, 1991.
[5] L. I. Tong and L. C. Chao, "Novel yield model for integrated circuits with clustered defects." *Expert System with Applications*, 34, pp. 2334–2341, 2008.
[6] J. A. Cunningham, "The use and evaluation of yield models integrated circuit manufacturing." *IEEE Transactions on Semiconductor Manufacturing*, 3(2), pp. 60–71, 1990.
[7] K. L. Hsieh, L. I. Tong and M. C. Wang, "The application of control chart for defects and defect clustering in IC manufacturing based on fuzzy theory." *Expert System with Applications*, 32, pp. 765–776, 2007.
[8] C. W. Chen, Y. S. Wu, and K. W. Su, "A study on detecting small shifts in quality levels with the geometric Poisson process using EWMA control schemes." *Journal of Quality*. 13(1), pp. 85–97, 2006.
[9] D. C. Montgomery, "*Introduction to statistical quality control.*" 3rd ed. John Wiley and Sons, New York, NY. 2005.
[10] F. J. Yu, Y. Y. Yang, M. J. Wang and Z. Wu, "Using EWMA control schemes for monitoring wafer quality in negative binomial process." *Microelectronic Reliability*, 51, pp. 400–405, 2011.
[11] G. V. Moustakides, "Optimal stopping times for detecting changes in distributions." *Annals of Statistics*, 14, pp. 1379-1387, 1986.
[12] B. T. Murphy, "Cost-size optimum of monolithic integrated circuits." *Proceeding of the IEEE*, 52, pp. 1537–1545, 1964.
[13] C. H. Stapper, "Defect density distribution for LSI yield calculations." *IEEE Transactions on Electron Devices*, 20, pp. 655–657, 1973.
[14] R. Winkelmann, "*Econometric analysis of count data.*" 5th ed. Springer-Verlag, Heidelberg, 2008.
[15] C. H. Stapper, "LSI yield modeling and process monitoring." *IBM Journal of Research Development*, 40(1/2), pp. 112–118, 2000.

[16] D. Brook, and D. A. Evans, " An approach to the probability distribution of CUSUM run length. *Biometrika*," 59(3), pp. 539–549, 1972.
[17] J. M. Lucas, and R. B. Crosier, "Fast Initial Response for CUSUM Quality Control Schemes: Give You CUSUM a Head Start." *Technometrics*, 24, 199-205, 1982.
[18] J. C. Chang, and F. F. Gan, "Cumulative sum charts for high yield processes." *Statistica Sinica*, 11, pp. 791-805, 2001.
[19] J. C. Fu, F. A. Spring, and H. Xie, "On the average run lengths of quality control schemes using a Markov chain approach." *Statistics and Probability Letters*, 56, pp. 369-380, 2002.
[20] L. C. Chao, "*The yield prediction model with neural network for integrated circuit-based on Poisson model*," Master's thesis, Department of Industrial Engineering and Management, National Chiao Tung University, Taiwan. 1997.