

Redundancy Yield Model for SRAMS

Nermine H. Ramadan, STTD Integration/Yield, Hillsboro, OR, Intel Corp.

Index words: Poisson's formula, yield, defect density, repair rate

Abstract

This paper describes a model developed to calculate the number of redundant good die per wafer. A block redundancy scheme is used here, where the entire defective memory subarray is replaced by a redundant element. A formula is derived to calculate the amount of improvement expected after redundancy. This improvement is given in terms of the ratio of the overall good die per wafer to the original good die per wafer after considering some key factors. These factors are memory area, available redundant elements, defect density and defect types with respect to the total reject die and defect distribution on the memory area. The model uses Poisson's equation to define the yield, then the appropriate boundary conditions that account for those factors are applied. In the case of a new product, knowing the die size, memory design, and total die per wafer, the model can be used to predict the redundancy yield for this product at different initial yield values. Optimizing the memory design by varying the number of memory blocks and/or redundant elements to enhance redundancy is also discussed. The model was applied to three products from two different process generations and showed good agreement with the measured data.

Introduction

Due to the continuing increase in the size of memory arrays, reaching a high yield from the same wafer is more challenging than ever. Redundancy is a way to improve the wafer yield and to reduce the test cost per good die by fixing potentially repairable defects. In order to forecast the volume of a certain product when redundancy is applied, it is important to estimate, as accurately as possible, the number of die gained after redundancy.

Redundancy is the process of replacing defective circuitry with spare elements. In SRAMs, rows and/or columns can be replaced, as well as an entire subarray. In a

previous study[1], a redundant yield estimation methodology was developed. It is applicable to row, column or block redundancy schemes. It distinguishes between repairable and non-repairable faults within a memory block. In order to apply this method, new CAD tools are required. This method is useful if row or column redundancy is used.

This paper will focus only on the yield estimation for block redundancy, as block redundancy was preferred over row and column redundancy for the SRAM architecture. It is usually easier to replace the entire subarray. This might seem like overkill; however, replacing the entire subarray allows for the replacement of defective peripheral circuits in addition to just the memory array elements. It also allows for the replacement of multiple bad bits, or other combinations of failing bits, rows and columns.

A yield multiplier M is defined as the ratio of the total good die after redundancy to the original good die per wafer, or

$$M = \text{total redundant good/original good} \quad (1)$$

so that the redundant yield, Y_{red} , is given as

$$Y_{\text{red}} = M \times Y \quad (2)$$

where Y is the initial yield. Forecasting of the redundancy yields is based on how accurately the factor M is calculated. A formula for M was obtained by using the correlated defect model. According to this model, an expression for the yield of die containing a number of defects, I , is given by

$$y_i = \frac{(n+I+1)! \times (DA)^I}{(I! \times n^{I-1}) \times (1+D \times A/n)^{n+1}} \times f^I \quad (3)$$

where

$$y_i = \text{yield of a die with } I \text{ defects}$$

D = average defect density (#/cm²)
 A = die area (cm²)
 n = correlation factor between defects
 f = fraction of the die area that contains the defects

The yield of die with zero defects can be obtained by setting $I = 0$ and $f = 1$ as

$$Y = 1 / \{ 1 + (A D / n) \}^n \quad (4)$$

With $n = 4$ and using equation (4) to substitute for the defect density, equation (3) becomes

$$y_i = Y \times ((I+3)(I+2)(I+1) / 6) \times f^I \times (1-Y)^{I/4} \quad (5)$$

Introducing g as the fraction of repairable defects, g varies depending on the number of repaired defects. An expression for M was obtained by summing y_i over the ratio of correctable defects and substituting in (2)

$$M = 1 + \sum_{I=1}^k ((I+3)(I+2)(I+1) / 6) \times (g f (1-Y)^{I/4})^I \quad (6)$$

M was calculated by entering arbitrary values of g and f in equation (6). However, there was no evidence to support the values of the repairable defect density represented by g used to calculate M .

Another formula was used to estimate the yield multiplier M . The yield is derived from Poisson's equation [2]

$$Y = \exp(-AD) \quad (7)$$

Instead of using a constant defect density, D , Murphy assumed several defect density distributions[3]. The most preferred distribution was a Gaussian. Stapper used a gamma distribution, which led to the following yield formula [4]

$$Y = 1 / \{ 1 + (A D / \alpha) \}^\alpha \quad (8)$$

where α is the average value of the coefficient of variation for the gamma function. The yield multiplier derived from the previous yield formula is given by

$$M = S \times (1 + 0.01 (L + I) A_{sb} D / k)^k \quad (9)$$

where

S = fuse programming success rate
 I = number of redundant elements
 L = number of subarrays
 A_{sb} = area of subarray (mm²)
 k = constant for MOS process

0.01=conversion from mm² to cm²

This simple formula is actually overestimating the redundancy improvement, since it assumes that all the defects are repairable.

In order to get a better estimate of the yield improvement, the nature and distribution of the defect need to be understood. These are taken into account in this model. When considering defects, it is important to realize that not all reject die are repairable: a die failing for a short, for example, cannot be repaired. Also the number of defective subarrays that could be repaired depends on the available redundant elements per memory block. This means that having more than one defect per die requires a certain distribution of those defects in order for redundancy to be successful.

Taking into account the above factors and using Poisson's equation to describe the yield, the present model was able to predict the redundant yield within the same range as shown by the real data. The following section illustrates how the key parameters affecting redundancy are used to develop the model.

SRAM Array Layout

Figure 1 shows the layout of a SRAM memory array. Before going into details, the following terms are defined as they will be used throughout the paper:

- *Subarray* This is a unit array of the memory area, and is shown as subarrays 0 to 72 in Figure 1.
- *Memory block or bloc k* This is a segment of the memory area, and is one of four rows shown in Figure 1.
- *Redundant element or elemen*. This is a spare subarray used to replace a memory subarray, and is given as subarrays R in Figure 1.

The die consists of two areas:

- *Repairable area* This includes all the circuitry in the subarray. In this model the repairable area is the sum of the areas of the memory subarrays.
- *Non-repairable area* This includes the periphery area. The redundancy elements are also considered part of the non-repairable area.

Block redundancy is illustrated in Figure 1. The defective subarray "4" is replaced by the redundant element R in the same memory block. This is done by programming the right fuses and shifting the array assignments.

defect density is used in the expression of λ , where D is found from the yield equation, equation (13), as

$$D = - \ln Y / A \tag{21}$$

Here Y is the random yield and is given by $Y = Ngd/N$ and is calculated from the data. Using A_{rep} instead of A, the expression for λ that will be used for the rest of the analysis is then

$$\lambda = F_{area} \times A D \tag{22}$$

Cumulative Model

Next, a better definition of Nrep is obtained: the number of die with one, two, or n defects. Poisson's equation is used to derive a formula for the number of die with a certain number of defects. Since

$$P_n = \{ \lambda^n \exp(-\lambda) \} / n! \tag{10}$$

where n is the number of defects, the following improvement factors can be defined:

M_1 = improvement factor from die with one defect
 M_2 = improvement factor from die with two defects
 M_n = improvement factor from die with n defects and is equal to

$$M_n = 1 + \sum Nrep_n / Ngd \tag{23}$$

where

$$\sum Nrep_n = N \sum \{ \lambda^I \exp(-\lambda) / I! \} \tag{24}$$

from $I = 1$ to $I = n$

The improvement factors are then given by

$$M_1 = 1 + \lambda \dots \dots \dots 1 \text{ def/die}$$

$$M_2 = 1 + \lambda + (\lambda)^2 / 2! \dots \dots \dots 2 \text{ def/die}$$

and for n defects per die

$$M_n = 1 + \lambda + (\lambda)^2 / 2! + (\lambda)^3 / 3! + \dots + (\lambda)^n / n! \tag{25}$$

Since there is a possibility of having more than one defect per block, λ must be multiplied by a so-called repair probability R_n , where R_n is the ratio of the combination of blocks and defects that can be repaired to the total number of combinations. This depends on the available number of redundant elements. M_n is then written as

$$M_n = 1 + R_1 \lambda + R_2 (\lambda)^2 / 2! + R_3 (\lambda)^3 / 3! + \dots + R_n (\lambda)^n / n! \tag{26}$$

An expression for R_n is found by using a binomial series expansion. If $G = X + Y$, where X and Y represent the number of blocks, the resulting binomial series is

$$G^n = (X + Y)^n = X^n + n X^{n-1} Y + {}^n C_2 X^{n-2} Y^2 + \dots + {}^n C_k X^{n-k} Y^k + Y^n \tag{27}$$

with

$${}^n C_k = n! / k! (n-k)! \tag{28}$$

as the coefficient of X. Note that this coefficient represents the number of terms with X raised to a certain power, where this power represents the number of defects on this block.

If G contains more than two terms, or more than two blocks, G is written as

$$G = X + Y + Z + \dots \text{ up to } b \text{ blocks}$$

and the series becomes

$$G^n = (X + Y + Z + \dots)^n = X^n + n X^{n-1} Y + n X^{n-1} Z + n X^{n-1} \dots + {}^n C_2 X^{n-2} Y^2 + {}^n C_2 X^{n-2} Z^2 + \dots + {}^n C_k X^{n-k} Y^k + {}^n C_k X^{n-k} Z^k + \dots + Y^n + Z^n + \dots \tag{29}$$

Knowing that each redundant element, e, can fix one defect, a term raised to the power of e+1 or higher indicates that it has more defects than elements and it cannot be fixed. This means that the number of possibly repaired blocks is equal to the total number of blocks and defect combinations minus the sum of coefficients of the terms raised to the power of e+1 or higher. All terms can be treated similarly, since all blocks are equal, and terms raised to the same power are collected together. Their coefficients can then be added together as well. Each coefficient in the previous series is repeated b-1 times for b terms. Except for the highest power in the series, it exists only b times. This means that the sum of coefficients can be written as

$$\text{sum} = \sum b (b-1)^{(n-k)} n! / k! (n-k)! \tag{30}$$

from $k=1$ to $k=n$, the number of defects. The number of repairable blocks is then

$$G_{rep} = (b)^n - \sum b (b-1)^{(n-k)} n! / k! (n-k)! \text{ from } k = e+1 \text{ to } k = n \text{ and } n \geq k \text{ always}$$

From the definition of R_n , the total combination of blocks and defects can be given by b^n . The repair probability is the ratio of the possibly repaired count to the total count, or

$$R_n = G_{rep} / b^n = \{ (b)^n - \sum b (b-1)^{(n-k)} n! / k! (n-k)! \} / (b)^n \quad (31)$$

and the formula for the cumulative improvement factor is

$$M_n = 1 + R_1 \lambda + R_2 (\lambda)^2 / 2! + \dots + R_n (\lambda)^n / n! \quad (32)$$

Note that this formula is applicable to up to $e \times b$ defects, which is the total number of elements and blocks; beyond that it is not useable. Higher order terms in the series are also negligible and can be ignored without affecting the improvement factor.

General Model

A general model is developed by including the effect of defect type in the previous improvement factor formula. The cumulative model is in fact overestimating the real data, because it assumes that all die are repairable. Studying the reject die data, it was found that only certain die could be fixed, namely raster type bins which occupy a certain fraction of the total reject die population. Adding to this the other restriction of obeying the previously described repair probability, only a certain fraction of those die is repairable. An efficiency factor η is introduced into the cumulative model. It is defined as the effective fraction of bad die repaired extrapolated at the maximum yield for a certain repairable area. It is calculated from

$$\eta = \gamma / F_{area} \quad (33)$$

where

$$\gamma = (Nbd_{cr} / Nbd) \times (Nrep / Nbd_{cr})$$

$Nrep$ = number of repaired die

Nbd = number of reject die

Nbd_{cr} = correctable reject die

which cancels out in the expression of γ . λ is then modified to

$$\lambda = \eta F_{area} \times A D \quad (34)$$

which is then used in the general model

$$M_n = 1 + R_1 \lambda + R_2 (\lambda)^2 / 2! + \dots + R_n (\lambda)^n / n! \quad (35)$$

This is the same as the cumulative model formula, equation (32), except for the expression of λ . γ is

obtained from the empirical data, so that one value of γ can be used for products from the same process generation. For a new process, γ from a previous process can be used, since its value is close from one process to the other.

Redundant Elements and Memory Blocks Optimization

In this section, how the number of redundant elements and memory blocks affects the yield improvement is studied. Increasing the number of spare elements increases the chance of repair. However, this impacts the repairable area, since the total area increases, while the repairable area is fixed. The dependency of the improvement factor on both the number of redundant elements and the repairable area is studied in order to check the possibility of improving redundancy by varying these two factors.

Since the improvement factor is a function of the repairable area and the defect density, and since the defect density is also a function of the area as calculated from the yield equation, equation (21), this equation is used to substitute for D in the expression for λ , equation (34), as

$$\lambda = \eta F_{area} \times (A D) = - \eta F_{area} \times \ln Y \quad (36)$$

The total die area is then written as

$$A = A_{rep} + A_{nrep} + 4 \times A_{el} \quad (37)$$

where A_{rep} is the repairable area and is equal to the area of the subarrays, A_{nrep} is the area of the die circuitry, and A_{el} is the redundant element area and is equal to the subarray area. Increasing the number of redundant elements by sets of 4, the total area is

$$A = A_{rep} + A_{nrep} + 4 \times e \times A_{el} \quad (38)$$

where "e" is the number of elements/block. The fraction of the repairable area is then

$$F_{area} = A_{rep} / (A_{rep} + A_{nrep} + 4 \times e \times A_{el}) \quad (39)$$

and

$$\lambda_i = - \eta A_{rep} \ln Y / (A_{rep} + A_{nrep} + 4 \times I \times A_{el}) \quad (40)$$

To study the behavior of the improvement factor with e and A_{rep} , start with the general yield improvement factor formula, equation (35)

$$M_n = 1 + R_1 \lambda + R_2 (\lambda)^2 / 2! + R_3 (\lambda)^3 / 3! + \dots + R_n (\lambda)^n / n!$$

In the case of adding extra redundant elements, the die area, and hence λ , is also changing, so that for each case with a certain number of elements the value of λ is different. The improvement factor formula is written as

1 element, n defects:

$$M_{n,1} = 1 + R_1 \lambda_1 + R_2 (\lambda_1)^2 / 2! + R_3 (\lambda_1)^3 / 3! + \dots + R_n (\lambda_1)^n / n!$$

2 elements, n defects:

$$M_{n,2} = 1 + R_1 \lambda_2 + R_2 (\lambda_2)^2 / 2! + R_3 (\lambda_2)^3 / 3! + \dots + R_n (\lambda_2)^n / n!$$

e elements, n defects:

$$M_{n,k} = 1 + R_1 \lambda_e + R_2 (\lambda_e)^2 / 2! + \dots + R_n (\lambda_e)^n / n! \quad (41)$$

For the number of defects less than or equal to the number of elements per block, the die is always repairable, i.e., $R_n = 1$ for all terms with $n \leq e$.

On the other hand, if a die has n defects, where $n > e \times b$, the die is never repairable. The improvement factor formula is then written as

1 element, $n \leq 1$ defects:

$$M_{n,1} = 1 + \lambda_1$$

2 elements, $n \leq 2$ defects:

$$M_{n,2} = 1 + \lambda_2 + (\lambda_2)^2 / 2!$$

e elements, $e < n \leq e \times b$ defects:

$$M_{n,e} = 1 + \lambda_e + (\lambda_e)^2 / 2! + \dots + (\lambda_e)^e / e! + R_{e+1} (\lambda_e)^{e+1} / (e+1)! + \dots + R_n (\lambda_e)^n / n! \quad (42)$$

where R_n follows the expression given by equation (31).

Next the effect of increasing the number of blocks per memory area on the redundant yield is studied. Dividing the memory area into a larger number of blocks also increases the chance for repair, since each block is accompanied by one redundant element. However, there is a certain maximum number of blocks, after which the increase in improvement is negligible, since the larger order terms in the series start to diminish. In this analysis, the total number of subarrays and F_{area} , are kept constant, but the size of the subarray is changed depending on how the memory area is divided. The number of redundant elements per block e is still one.

The general formula, equation (35), is used here, where the number of blocks b is changed in the repair probability term R_n given by equation (31).

Results and Discussion

This report describes a model that calculates the redundancy yield. The amount of improvement depends on some key factors: the repairable area, available redundant elements, defect density and types of defects and their distribution on the die. The memory area is the area that contributes to redundancy, since the rest of the die area cannot be fixed and has to be functional. Only the random defect density is considered here as the defect category that is potentially repairable. The number of available redundant elements also determines how much improvement can be gained. If there is one redundant element per memory block, only one defect per block can be fixed. The type of defects is another important factor in estimating the redundancy yield. Raster defects such as bits, columns or rows (where bits can represent individual or clustered defects as long as they fall in the same memory block) are considered repairable. Although the number of defects that could be fixed equals the number of redundant elements available, those defects have to follow a certain distribution on the memory array according to the repair probability described in the text.

Figures 2 through 4 show the improvement factor versus the initial yield calculated by the three models: simple, cumulative, and general. Data is compared to three products from two different process generations.

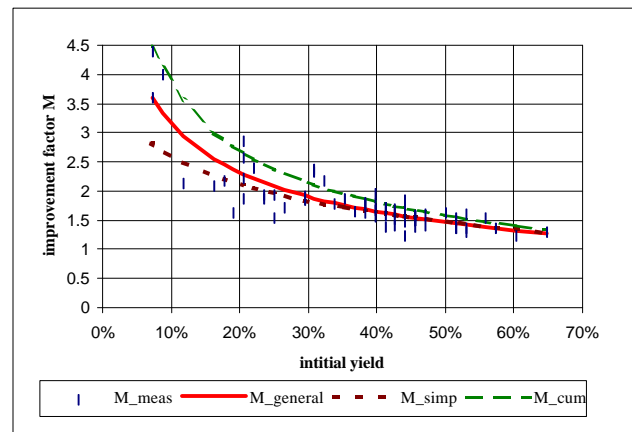


Figure 2: Three formulas compared to data measured on product 1

Comparing the formulas of the improvement factor, the closest fit to the actual data was obtained when all the factors affecting redundancy were accounted for (general model). The simple model underestimates the data, since it assumes the repair of die with one defect only, which is not the real case. The cumulative model is overestimating the data. It considers all types of defects and assumes all of them are repairable, if their count is equal to or less than the number of redundant elements. Thus, it ignores the restriction of allowing one defect per block.

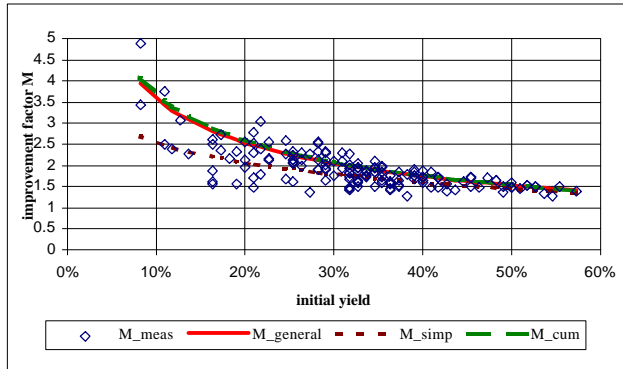


Figure 3: Three formulas compared to data measured on product 2

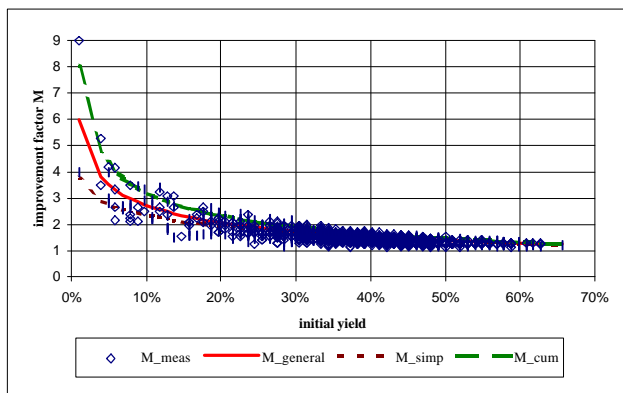


Figure 4: Three formulas compared to data measured on product 3

The effect of varying the number of redundant elements is shown in Figure 5. The effect of adding more redundant elements is mostly seen at a lower initial yield. It was observed that the improvement in yield is significant up to two extra sets of elements for a die of originally one redundant element per block. Beyond that, the effect of decreasing the repairable area is dominating, so that the two factors cancel out, and the overall improvement is unchanged.

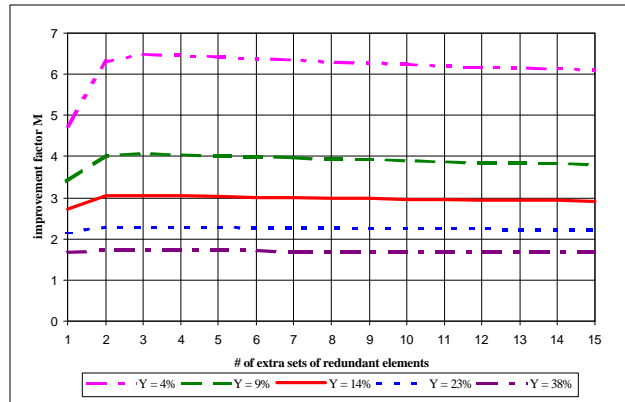


Figure 5: Improvement factor when extra redundant elements at different initial yield Y are added

Figure 6 shows the effect of dividing the memory area into a large number of blocks. Again the enhancement in the yield multiplier is observed at a lower yield. With an increase in the initial yield, an improvement in the redundant yield was observed up to six blocks. Beyond that, the effect of more blocks per repairable area is not noticeable, since the higher order terms in the multiplier formula are negligible and do not add extra value to the yield multiplier.

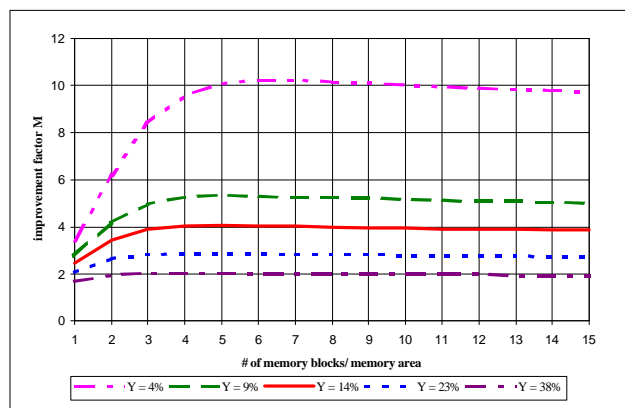


Figure 6: Improvement factor when the number of memory blocks at different initial yield Y is increased

Conclusion

A model for calculating the redundancy yield is developed and described in this paper. Poisson's equation plus the effect of some redundancy-influencing factors are used to derive a general yield multiplier formula. The memory area is considered the only portion of the die area where redundancy is applied. The random defect density is used here as the only defect category that contributes to redundancy. From the defect population,

only a fraction of it can be repaired depending on the nature of the defect. According to the die design, the number of repairable defects depends on the available redundant elements per memory block. This means that the number of defects must be below a certain value, and the defects have to follow a certain distribution throughout the memory area to enable redundancy. An efficiency factor is introduced and empirically evaluated to account for the repairable defects. Combining those factors, a general formula is derived and shows good agreement with the actual data. Knowing the properties of a new product and using the efficiency factor for the process generation, the redundancy yield of a new product can be predicted. The formula can also be used to study the impact of varying the number of redundant elements and memory blocks on the final result. Thus, a better design that optimizes the number of redundant elements, memory size with respect to the total die area, and the number of blocks in the memory area might result in a more efficient redundancy scheme.

and Engineering Physics from the University of Wisconsin, Madison in 1986 and 1992, respectively. In 1994 she joined Intel in Oregon and is currently working as a Senior Integration Engineer in Sort/Test Technology Development. Her e-mail address is nermine_ramadan@ccm.ra.intel.com

Acknowledgments

I would like to thank Dan Grumbling for initiating this project, Tim Deeter for the useful discussions and comments during the development of the model, and Mike Mayberry for his continuous support and guidance throughout this work.

References

- [1] Jitendra Khare, et al. Accurate Estimation of Defect-Related Yield Loss in Reconfigurable VLSI Circuits. *IEEE Journal of Solid State Circuits*. Vol. 28, No. 2, February 1993.
- [2] R. M. Warner, Jr. Applying a Composite Model to the IC Yield Problem. *IEEE Journal of Solid State Circuits*. Vol. SC-9, No. 3, June 1974.
- [3] B. T. Murphy. Cost Size Optima of Monolithic Integrated Circuits. *Proc. IEEE*. Vol. 52, December 1975.
- [4] C. H. Stapper. On Murphy's Yield Integral. *IEEE Trans. Semiconductor Manufacturing*, Vol. 4, November 1991.

Author's Biography

Nermine Ramadan received a B.Sc. in Nuclear Engineering from the University of Alexandria, Egypt in 1982, and a M. Sc. and Ph.D. in Nuclear Engineering