IC HTOL Test Stress Condition Optimization

Brian Peng[1], Ing-Yi Chen[2] , Sy-Yen Kuo[1], Colin Bolger[3]
Department of Electrical Engineering, National Taiwan University, Taiwan, R.O.C.[1]
Department of Computer Science & Information Engineering, National Taipei University of Technology, R.O.C.[2]
MPE division, VIA Technology Incorporation, Taiwan, R.O.C.[3]
E-mail: brian_peng@via.com.tw, ichen@csie.ntut.edu.tw, sykuo@cc.ee.ntu.edu.tw, colin_bolger@via.com.tw

Abstract

HTOL (High Temperature Operation Life) test is used to determine the effects of bias and temperature stress conditions on solid-state devices over time. It simulates the devices' operating condition in an accelerated manner, and is primarily for device reliability evaluation. This paper addresses an SA (Simulated Annealing) method used for the HTOL test stress condition decision-making that is an optimization problem. The goal is to reduce the resources for the HTOL test, hardware or time, under reliability constraints. The theory of reliability statistic model and the SA algorithm are presented. In our optimization algorithm, we need to calculate the accurate HTOL stressed power for the next optimization loop since the Vs (Stressed Voltage) that is optimized will affect not only Afv (Voltage Acceleration Factor) but also Aft (Thermal Acceleration Factor). A curve-fitting algorithm is applied to get reasonable accelerated factors and reliability calculations. The model selection process and statistical analysis of fitted data by different models are also presented. Experimental results with different stress condition priorities and different user settings are given to demonstrate the effectiveness of our approach.

1. Introduction

As ULSI technology continues down scaling, and the critical width and spacing geometries become smaller, tiny residues and particles play more important roles as wafer yield killers and reliability failure stimuli. However, these defects cannot be easily screened out either by the chip level or the package level functional tests. This means some devices shipped to customers have a potential reliability issue over an expected period of time. The reliability tests are used to evaluate the product life and guarantee the product reliability within the warranty period by using accelerated conditions to simulate operating life over a shortened test period. The HTOL (High Temperature Operation Life) test is one such reliability test, and requires the application of a high temperature and voltage stress on the semiconductor devices over long periods, for a small sample size, to evaluate the lifetime and failure rate of the larger population. Statistical analysis is important in order to calculate the FIT (Failure In Time), MTTF (Mean Time to Failure) based on the HTOL test results. Due to the expense of test time, HTOL board, test socket, it is necessary to find an optimized stress condition to minimize the costs of HTOL testing. Similar problems as semiconductor burn-in stress condition decision-making have been addressed in [1], [2]. An SA-based optimization algorithm [3] is proposed to reduce the resources of HTOL testing. SA algorithm is widely used for solving some well known combinatorial optimization problems [4]-[10]. Using different priority sets, the program will find the best stress condition to meet the reliability constraints. Therefore, we can save costs

Proceedings of the 19th IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems (DFT'04)
1550-5774/04 $ 20.00 IEEE

IEEE COMPUTER SOCIETY

while still meeting the required reliability level. Due to the cross relationship of the Vs, Tah and the Ws (Stressed Power), the program must intelligently capture the estimated Ws values due to different Vs and Tah values, which are randomly changed during optimization. A best-fit curve algorithm helps to fit experimental data and output the Ws for the optimization routine. Many curve fit techniques and analysis methods are discussed in previous works [11], [12]. In our work, a least square method with *Gauss Jordan Elimination* is utilized to solve for our required parameters. The rest of this paper presents the reliability tests, HTOL test and algorithms for the optimization. In section 2, the statistical failure rate calculation for the HTOL test is introduced for our proposed stress condition optimization. In section 3, the algorithm for our optimization problem is introduced. In section 4, the experiment results are discussed. In section 5, conclusions are given.

2. Failure Rate Calculation of HTOL test

The failure rate is calculated using chi-square failure distribution based on a constant failure rate concept. The Activation energy (Ea) is determined by the failure mechanisms [13]. Here, we take an average Ea as 0.7eV for the calculation. We use the Arrhenius equation to determine the acceleration factor when changing from Tjl (Junction Temp) to Tjh (Junction Temp @ Tah). The thermal acceleration factor Aft is

$$Aft=\exp\{(\frac{Ea}{K}*\frac{1}{T_{jl}}-\frac{1}{T_{jh}})\}$$

Where, Tjl calculate the junction temperature of the DUT under normal conditions using the known case temperature (Tcl).

$$T_{jl}=T_{cl}+(W_u*R_{jc})$$

The below equation shows the calculation method for the junction temperature (Tjh) of the DUT under stress conditions, which are calculated from the known atmospheric temperature (Tah).

$$T_{jh}=T_{ah}+(W_s*R_{ja})$$

Tah is the atmospheric temperature under stress conditions (Temperature in °C + 273), Tjh is the junction temperature under stress conditions (˚K), Tcl is the case temperature under normal working conditions (˚K), Tjl is the junction temperature under normal working conditions (˚K), Wu is the normal operation power consumption (watts), Rja is the thermal resistance co-efficient from air to junction, Rjc is the thermal resistance co-efficient from case to junction, Ea is the thermal activation energy (eV), k is the Boltzman's constant (8.61423*10-5 (eV/°K)).

The voltage acceleration factor Afv is as defined in the below equation:

$$Afv=\exp\{(\frac{K}{t_{ox}}*(V_s-V_u)\}=\exp\{\beta(V_s-V_u)\}$$

Where, Vs is the stress voltage, Vu is the normal Voltage for device use and β is the voltage acceleration factor.

Reliability Failure Rates (λ)

This calculation uses a Chi-Square probability distribution to approximate the reliability failure rate curve.

$$\lambda=\{(\frac{\chi^2}{2}*\frac{1}{(Af*N*H)})\}$$

IEEE COMPUTER SOCIETY

Where, λ is the failure rate, Af is the overall acceleration factor, N is the sample size, H is the test period in hours and χ^2 is the inverse chi-square distribution factor.

Failures In Time (FIT)

FIT is defined as failures in 1 billion device hours. Obviously such extensive test periods are not normally feasible under standard operating conditions so that accelerated testing is required.

$$\mathrm{FIT}=\{(\frac{\chi^2}{2}*\frac{1}{(Af*N*H)})*10^9\}$$

3. Optimization Algorithms

Figure 2 shows the stress condition optimization flow. The default parameter settings are necessary to initialize the program in a reasonable state. After the settings, Aft, Afv and FIT are calculated based on the default settings and the equations we mentioned in section 2. If FIT achieves the desired reliability level, the program will be terminated and from this we get the optimal settings of Tah, Vs, SS (Sample Size) and Hr (Stress Time). The main sub-routine is the SA Algorithm, which includes the priority settings for Tah, Vs, SS, Hr.

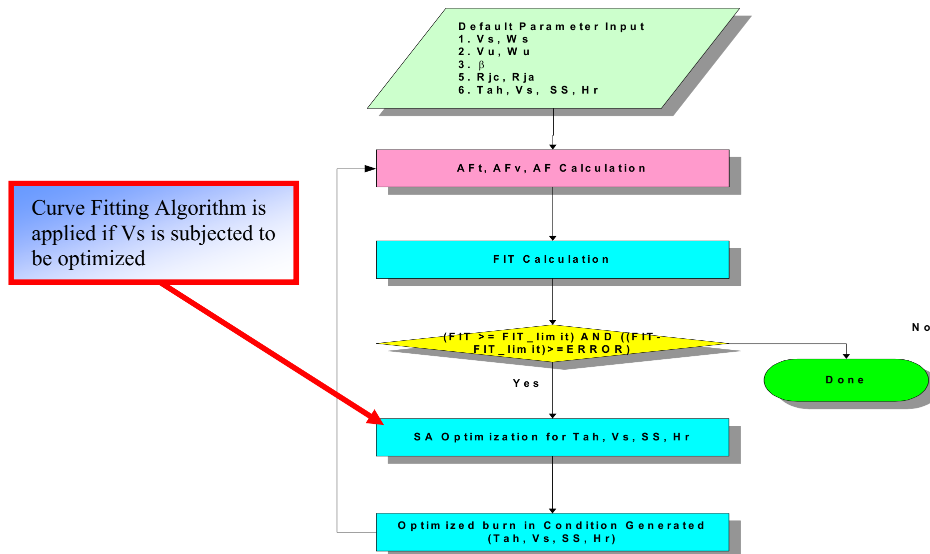

Figure 2. The HTOL stress condition optimization flow.

3.1. Proposed SA-based Algorithm

Annealing is a process that has been used to temper metals throughout much of history in many civilizations. This process involves heating and then slow cooling (rather than fast cooling or quenching) in order to prevent brittleness in metals, plastics, or glass. When they are allowed to cool slowly, strong bonds form between atoms or molecules rather than imperfections. Thus long periods are spent at near freezing temperatures. Figure 3 shows the pseudo code of proposed SA based algorithm for our optimization problem.

Anneal (I, S, Iteration)

PRE: I = Instance of problem

S = Initial feasible solution (Vs_int, Tah_int, SS_int, Hr_int) to I

h(S) = Objective function = (Tah_pri * Tah_Fact * Tah) + (Vs_pri * Vs_Fact * Vs) + (SS_pri * SS_Fact * SS) + (Hr_pri * Hr)

IEEE COMPUTER SOCIETY

```
POST: S = Feasible/ Optimized solution (Vs_opt, Tah_opt, SS_opt, Hr_opt)
Determine initial temperature T
do
   tries = moves = goodmoves = 0
  do
      NewS = random neighbor of S (The solution must meet constraints of Tah, Vs, SS, Hr)
         ΔS = h(NewS) - h(S)
      if (FIT_Est <= FIT_limit)AND((FIT_limit-FIT_Est)<=ERROR)
         tries = tries + 1
         rnd = random number between 0 and 1
      if ΔS <0 or rnd < e(- ΔS /T) then
         S = NewS
         moves = moves + 1
      if ΔS < 0 then goodmoves = goodmoves + 1
  while ((goodmoves < iteration) and (moves < 2*iteration) and (tries < 20*iteration))
    T=0.85T
while ((moves > 0.05 tries) and (T > 0.01))
```

Figure 3. Proposed SA-based algorithm.

The famous formula for the probability of any particular configuration of atoms is $e^{-E_i/k_B T}$, where E_i is the energy of the configuration, T is the temperature, and k_B is Boltzmann's constant. The configurations with minimum energy become optimal solutions. Some properties of this SA process are: (a) The probability of moving to a new configuration decreases as the temperature decreases. (At high temperatures, lots of changes may occur; At low temperatures, very few.) (b) The probability of moving to a new configuration increases as the change in energy decreases. (It is more likely to change to a configuration with an energy level that is not too much higher than the current configuration.) (c) It is possible to move away from local optima because probabilities determine whether or not to move. (d) Smaller changes are allowed at lower temperatures. Annealing temperature is a controversial issue in simulated annealing. Obviously, optimization problems have no temperature, but the above algorithm calls for one, which makes defining an initial temperature rather interesting philosophically. For adapting the SA Algorithm to our optimization flow, the cost function h(S) needs to be defined. Here, our cost function is h(s) = (Tah_pri * Tah_Fact * Tah) + (Vs_pri * Vs_Fact * Vs) + (SS_pri * SS_Fact * SS) + (Hr_pri * Hr). The constraint of the cost are Tah_min <= Tah <= Tah_max, Vs_min <= Vs <= Vs_max, SS_min <= SS <= SS_max, Hr_min <= Hr <= Hr_max, (FIT_Est <= FIT_limit) and (FIT_Est - FIT_limit <=ERROR). Tah_pri, Vs_pri, SS_pri, Hr_pri are the weighting parameters, which depended on the priority of stress condition to be optimized. Tah_Fact, Vs_Fact, SS_Fact, Hr_Fact are the factors that modulate data variation range before the start of the optimization process. For instance, if the data range of Hr is larger than the other three parameters, Tah_Fact = (Hr_max-Hr_min) / (Tah_max-Tah_min), Vs_Fact = (Hr_max-Hr_min) / (Vs_max-Vs_min), SS_Fact = (Hr_max-Hr_min) / (SS_max-SS_min). There is no Hr_factor shown above because it is a reference term for the other three factors and should be equal to 1. Tah_max/ min, Vs_max/ min, SS_max/ min, Hr_max/ min are determined by the limitation of the device characteristic or resources. After the SA optimization, the expected configuration is the minimized value of Tah, Vs, SS and Hr under desired reliability level constraints.

3.2. Curve Fit Algorithm for Ws Estimation

During optimization, Vs, Tah, SS and Hr are randomly generated to calculate the FIT rate and the cost to be compared with the last generated cost. However, Vs and Tah are not

IEEE COMPUTER SOCIETY

perfectly independent. As Vs increases, both the voltage and thermal acceleration factors also increase. Therefore, a 3-dimension curve fit algorithm is applied in getting accurate Ws values to be used in Aft calculations as Vs increases. Figure 4 illustrates the flow and analysis of the curve fit. Based on the response variables and independent variables determined by step 1 and 2, one can estimate a prediction equation to fit our input data as a multiple regression. The multi- regression model is $Yi=\beta_0+\beta_1X_{1i}+\beta_2X_{2I}+.....+\beta_\kappa X_{ki}+\varepsilon_i$, i=1,2,....,n . Where, Y is the expected response variable ("Is" in our application) that we're going to fit, X1, X2, ……Xk are the independent or expected variables which are the different stress input variables (Tah, Vs) in our experiment, β_0, β_1, …. β_k are the modeling parameters which are also called *Partial Regression Coefficients*, ε_i is the random error items and $\varepsilon_i : N(0, \sigma^2)$, i=1, 2, ..., n , The regression model is $\hat{Y}=\beta_0+\beta_1X_1+\beta_2X_2......+\beta_kX_k$. The error between actual data Y_i and expected data $\hat{Y}$ is ε_i, i.e. the vertical distance between Yi and $\hat{Y}$ ($\varepsilon_i=Y_i-\hat{Y}$). We use least square method to get the point estimate of parameters β0, β1, β2…….The sum of least square is $D=\sum_{i=1}^{n}(Yi-\hat{Y})^2$.

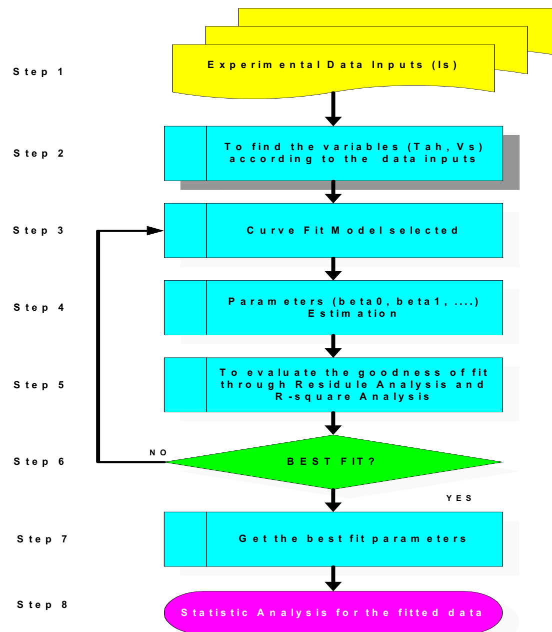

Figure 4. Curve fit flow and analysis for the Ws generation.

To find β0, β1, β2…. which minimize D, D is partially differentiated by β0, β1, β2…and let it equal to zero.

$$\frac{\partial D}{\partial \beta_i}=0,\ (i=1\sim n),\ \text{to find the proper value of the } \hat{\beta}_i.$$

By partial differential process and adjusting:

$$n*\beta_0+\sum_{i=1}^{n}\beta_1x_{i1}+\sum_{i=1}^{n}\beta_2x_{i2}+\sum_{i=1}^{n}\beta_3x_{i3}+.....=\sum_{i=1}^{n}y_i$$

$$\sum_{i=1}^{n}\beta_0x_{i1}+\sum_{i=1}^{n}\beta_1x_{i1}^2+\sum_{i=1}^{n}\beta_2x_{i1}x_{i2}+\sum_{i=1}^{n}\beta_3x_{i1}x_{i3}+.....=\sum_{i=1}^{n}x_{i1}y_i$$

$$\sum_{i=1}^{n}\beta_0x_{i2}+\sum_{i=1}^{n}\beta_1x_{i1}x_{i2}+\sum_{i=1}^{n}\beta_2x_{i2}^2+\sum_{i=1}^{n}\beta_3x_{i2}x_{i3}.....=\sum_{i=1}^{n}x_{i2}y_i$$

.

.

$$\sum_{i=1}^{n}\beta_0x_{in}+\sum_{i=1}^{n}\beta_1x_{i1}x_{in}+\sum_{i=1}^{n}\beta_2x_{i2}x_{in}+\sum_{i=1}^{n}\beta_3x_{i3}x_{in}+.....=\sum_{i=1}^{n}x_{in}y_i$$

IEEE COMPUTER SOCIETY

β0, β1, β2 ……βn, are the optimal values that we can get by means of *Gauss-Jordan Elimination*. Therefore, the expected equation is $\hat{Y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ...$, which represents the expected value of stressed current "Is". To evaluate our regression models, the residual analysis and R^2 (Coefficient of determination) are used to determine which model is the best one. If the model is not suitable, some expected variables must be changed, added, removed, or a search of other suitable regression models and then repeat the regression flow until the best-fit model is found.

4. Experiment Results and Discussion

4.1. Curve Fit Results and Analysis

The curve FIT experiment is implemented by hardware and software. In hardware, the 10 DUTs (Device Under Test) we selected for the measurement are the VIA Technology Inc. advanced 0.13um product. The HTOL board is used to load DUTs for the HTOL testing. 6 steps of temperature (25°C, 60°C, 75°C, 100°C, 125°C, 140°C) are applied on the DUTs. There are 6 voltage levels settings (1v, 1.1v, 1.2v, 1.3v, 1.7v) applied on the DUTs. The minimum and maximum of stressed voltage must be considered based on the device characteristics. In software, BIST (Built In Self Test) pattern is applied on the DUTs to cover the embedded hard macro HTOL stressing. The SCAN test pattern is performed during the HTOL stressing to ensure the high toggle rate of sequential logics of the whole chip. Figure 5 shows four different graphs of regression models and fitting results. Obviously, the model for Figure 5(a) is a poor fit model compared with the other three. The fitting results (Regression Curve) of Figure 5(b), (c), (d) are almost the same. To find the best candidate of curve Fit model, the R^2 coefficient is utilized. The current measurement of the device is based on 1.1v stressed voltage under different stressed temperature since it's hard to compare the 3 dimension (Is, Tah, Vs) fitting result by different regression models in one graph. The fitting results of "Is" versus Vs graph while fixing Tah are verified and similar to the result as above. Table 1 shows the R^2 and R^2(Adj) comparison of different regressive models. The model with the cube term as Figure 5(d) is the best model since the R^2 and R^2(Adj) value is larger than other three. Figure 6 shows the 3D curve fitting result of the regressive model with cube terms we selected as introduced before. The "Is" to Tah form a polynomial with cube term while "Is" versus Vs is the same. We can obviously find that the Is, Tah and Vs are mutually dependent.

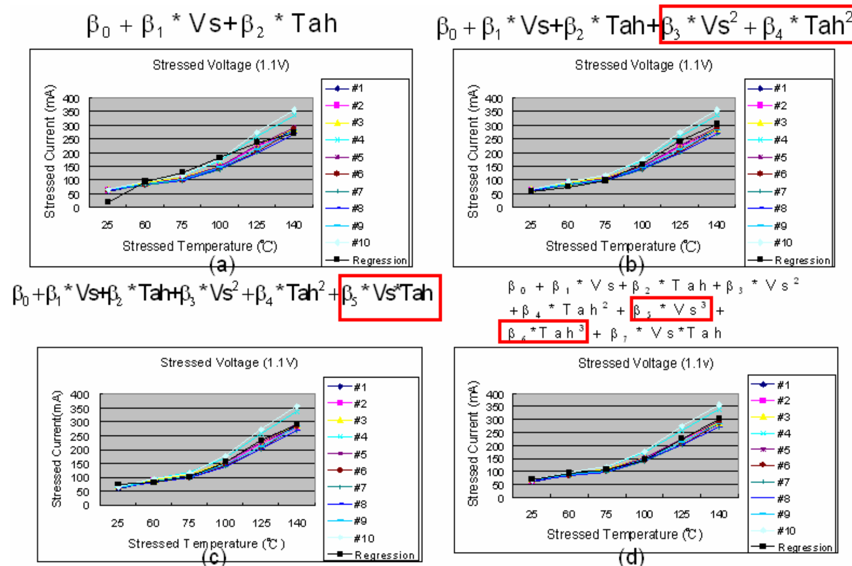

Figure 5. Curve fit model comparison.

IEEE COMPUTER SOCIETY

Table 1. R^2 and R^2(Adj) coefficient comparison of different models.

	Linear	With quadric term	With mutual term	With cube term
SST (Total Sum of Squares)	4799315	4799315	4799315	4799315
SSE (Sum of Squares for Errors)	728322	423402	378534	356024
R square (1-SSE/SST)	0.848244593	0.91177866	0.921127494	0.925817747
R square adjust (1-(SSE/(n-k-1)/Syy/(n-1)))	0.999850487	0.999913082	0.999922293	0.999926914

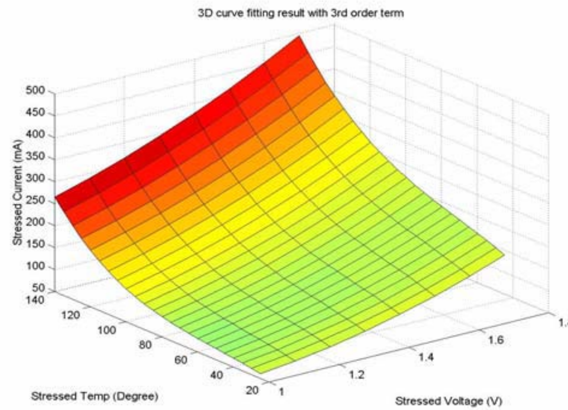

Figure 6. 3D graph of the curve fit.

4.2. Optimization Results and Analysis

Table 2 shows the SA results for different priority settings of the stress conditions. Programs are running under the following environment, CPU: Intel Pentium 4 Mobile CPU 1.4Ghz, Memory: 256MB DDR, OS: Windows XP Home Edition, Compiler: Microsoft Visual C++ 6.0. The variable range of Vs is 1V to 2V, 125°C to 140°C for Tah, 80pcs to 240pcs for SS and 1000hrs to 2000hrs for the Hr value. Column 2–6 compare the program execution times for different fixed value settings. For column3, the execution time is shorter than column2. This is because the average FIT rate during the optimization is lower than the FIT_limit we set, the larger Wu result in the generated FIT rate during optimizing is closer to the FIT_limit. The same results are shown for the comparison of column 3, 4, 5 to Column1. In Column 7-10 are the results with different priority settings compared with column 2. The final optimized configurations are shown in rows of best Tah, best Vs, best Hr and best SS. As shown, the Tah, Vs, Hr, SS values follow the priority we set.

IEEE COMPUTER SOCIETY

Table 2. Optimization results of different priority sets.

Opt. Priority	Tah>Vs>SS>Hr					Vs>Tah>SS>Hr	SS>Vs>Tah>Hr	Hr>Vs>SS>Tah	Tah>SS>Vs>Hr
Beta	3	3	3	3	3	3	3	3	3
Vu (volt)	1	1	1	1	1	1	1	1	1
Init. Vs (volt)	1	1	1	1	1	1	1	1	1
Max. Vs (volt)	2	2	2	2	2	2	2	2	2
Wu (watt)	1	2	1	2	1	1	1	1	1
Init. Ws (watt)	2	2	2	2	2	2	2	2	2
Tcl (℃)	70	70	85	70	70	70	70	70	70
Init. Tah (℃)	125	125	125	125	125	125	125	125	125
Init. SA Temperature	1.00E+07	1.00E+07	1.00E+07	1.00E+07	1.00E+07	1.00E+07	1.00E+07	1.00E+07	1.00E+07
Rjc (℃/w)	25	25	25	30	25	25	25	25	25
Rja (℃/w)	30	30	30	30	25	30	30	30	30
Init. Hr (hrs)	1000	1000	1000	1000	1000	1000	1000	1000	1000
Init. SS (pcs)	80	80	80	80	80	80	80	80	80
C.L.	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
DF	2	2	2	2	2	2	2	2	2
Init. FIT	85.1	346.7	202.1	114.4	126.5	85.1	85.1	85.1	85.1
Best Tah (℃)	125	125	125	125	125	127	128	136	125
Best Vs (volt)	1.1	1.3	1.17	1.01	1.02	1	1.25	1.02	1.24
Best SS (pcs)	181	236	238	239	194	191	80	191	90
Best Hr (hrs)	1998	1991	1918	1999	1993	1905	1837	1000	1879
Best Ws (watt)	0.21	0.37	0.29	0.22	0.21	0.21	0.34	0.26	0.33
Final FIT	198.3	199.9	198.8	194.6	195.5	197.1	191	191	194.2
MTTF (device-hrs)	5042864	5002501	5030181	5138746	5115090	5073566.717	5235602.094	5235602.094	5149330.587
Final SA Temperature	476.84	616.26	648.79	616.3	430.4	139.2	350.5	430.4	350.5
Program Execution Time(s)	16	10	9	11	14	16	16	15	16

5. Conclusions

Reliability tests are important for evaluating the normal IC component reliability in an accelerated manner. Customers would not accept the product if it is unreliable. HTOL testing is one of product reliability test methods, providing voltage and thermal acceleration to the DUTs. Various stress conditions during HTOL testing affect the failure rate and reliability calculation. The stress conditions must be optimized to satisfy both the desired reliability level and testing costs. A simulated annealing algorithm is successfully applied for our HTOL stress condition optimization. During optimization, a curve fit method is utilized for generating a feasible stress power while Vs increases during optimization. The developed tool can easily generate a set of reasonable stress conditions predetermined in accordance with reliability constraints.

6. References

[1] Don-Hae Chi, Way Kuo, "Burn-In Optimization under Reliability & Capacity Restrictions", IEEE Trans. Reliability, vol.38, NO. 2, pp. 193-198, 1989 Jun

[2] W. Kuo, "Reliability enhancement through optimal burn in", IEEE Trans. Reliability, REL-33, pp. 145-156, 1984

[3] S. Kirkpatrick, C. D. Gellatt, Jr., M. P. Vecchi, "Optimized by Simulated Annealing ", SCIENCE, 1983 May

[4] Feng-Tse, Cheng-Yan Kao, and Ching-Chi Hsu, "Applying the Genetic Approach to Simulated Annealing in Solvong Some NP-Hard Problems", IEEE Trans. on Systems, Man. and Cybernetics, vol. 23, NO. 6, November/ December, pp. 1752-1767, 1993

[5] N. Deeb, Sargent and Lundy, "Simulated Annealing in Power Systems", IEEE, 1992

[6] B. Rosen, "Function optimization based on advanced simulated annealing", IEEE Workshop on Physics and Computation-PhysComp, pp. 289-293, 1992

[7] A. Corana, M. Marchesi, C. Martini, and S. Ridella, "Minimizing multimodal functions of continuous variables with the simulated annealing algorithm, ACM Trans. Mathl. Software 13(3), 1987

[8] P.J.M van Larhoven and E.H.L Aarts “Simulated Annealing: Theory and Applications”

[9] Emile Aarts and Jan Korst “Simulated Annealing and Boltzmann Machines”

[10] D.F. Wong and H.W. Leong and C.L. Liu “Simulated Annealing for VLSI Design”

[11] M. Alhanaty, M. Bercovier. Curve and surface and design by optimal control methods. CAD V.33, pp167-182, 2001

[12] Harvey Motulsky and Arthur Christopoulos, "Fitting Models to Biological Data using Linear and Nonlinear Regression", 2003

[13] JEDEC PUBLICATION, JEP122-A “Failure Mechanisms and Models for Semiconductor Devices”, 2001 December.

IEEE COMPUTER SOCIETY