Application of Statistical Methods for Enhancing Automatic Test Pattern Generation[1]

András Petri, dr. András Pataricza
Technical University of Budapest
Dept. of Measurement and Information Systems
E-mail: petri@mmt.bme.hu

Introduction

The goal of automatic test pattern generation (ATPG) is to find a method for searching of test patterns systematically for a given error, from the set of all possible input patterns. This generic task is NP-complete, so simple brute-force exhaustive search algorithms are completely hopeless for larger-scale practical systems. It is necessary to apply some kind of heuristic search algorithms. An attractive attempt was made earlier to build an automatic test generator based on high-level VHDL descriptions, in the frame of the FUTEG project [4]. This test generator contains some heuristic search enhancement ideas.

Fine-tuning of these heuristic algorithms requires experimental evaluation, with respect to the *fault coverage* properties of the test generation process, as it provides the comparison base with classical low-level ATPG methods.

One of the most important features implemented in the FUTEG ATPG is the support of *hierarchical test generation*, i.e. tests generated for the (relatively simple) individual components of a great complex circuit can be re-used when generating test for the whole system, further improving the overall fault coverage of the test generation (TG) process. Traditional fault coverage calculation methods, however, do not consider this hierarchical structure very well; usually the complex system must be re-evaluated completely in order to obtain coverage measures, even if the measures of the components are already available from previous experiments.

A recently developed method – called *stratified sampling* [1,2] – seems to work around this problem. It considers the fault space to be divided into disjoint classes or *strata*, and it provides *global fault coverage* information by combining the measures calculated in the different strata. *Confidence limits* of the results are also calculated.

These fault coverage measures can be used in different phases of ATPG for enhancing heuristic algorithms:

- During systematic TG, single faults are placed into the system in a specific order. The *scheduling* of the faults can determine whether information obtained from previous test pattern calculations can be re-used for generating tests for subsequent faults, thus decreasing the computation requirements. A better scheduling can be achieved if the interaction of different faults is estimated from the fault coverage measuses.
- From statistical calculations, *correlation* between different parts of the system is also revealed. This type of information can be exploited both for generating *correlated tests* (tests that detect more than one fault, due to correlation between the faults) and identifying *hard-to-test faults* (faults in parts which are only loosely correlated with other parts).
- Test patterns are checked by *fault simulation* in order to find further faults tested by the same pattern. During fault simulation, coverage measures can help finding *"almost tests"*, i. e. input patterns that do not result in an observable discrepancy on an output, but the effect of the fault is propagated relatively far. These input patterns are perfect candidates for subsequent test search sessions, as they can very probably be modified to get a good test with minimal effort.

[1] This research was partially sponsored by Hungarian Scientific Research Fund (OTKA Contract N°. T015728)

Definitions

For easier understanding of stratified sampling, the problem of determining the fault coverage of a test generation algorithm on a given system is presented in an alternate form, as the problem of determining the fault coverage of a system from fault injection experiments. It is obvious that the two problems can be derived from each other.

The *input space* of a system is considered as the Cartesian product of the set of possible faults and the set of input sequences that activate them. This set will be denoted as G. The fault handling capability of the system can be described as a variable H with a value of 1 if the system can handle the fault correctly and 0 otherwise. Therefore the probability of handling a fault correctly, called *coverage factor (c)*, is actually a *conditional probability* of H=1 if a given fault $g \in G$ occurs. If H is considered as a discrete random variable, *h(g)* denotes the value of H for a specific $g \in G$ and *p(g)* denotes the probability of g, the coverage factor can be formally defined as $c = \sum_{g \in G} h(g) p(g)$; it is actually the expected value of H over the set G.

It is trivial that the knowledge of the probability of the single fault/activation sequence pairs, *p (g)*, is essential for making any meaningful estimations about the coverage factor. If no such knowledge exists, only a rough approximation – that all fault/activation pairs are equally probable – can be made; in this case, a *coverage proportion* $\tilde{c} = \frac{1}{|G|} \sum_{g \in G} h(g)$ can only be calculated that gives a very rough estimate about the system's fault tolerance.

The relation between the coverage proportion and the coverage factor of a given system is determined by the *correlation* between the random variable H and the random variable P, that has the value *p(g)* for a given $g \in G$. From the definition of covariance, it can be shown that $\tilde{c} = c - |G| S_{HP}$. It means that the coverage proportion is optimistic with respect to of the coverage factor if the correlation between H and P is positive (e.g. if the fault-tolerant mechanism of the system is designed in a way that it handles the most probable fault/activation pairs better), and pessimistic if the correlation is negative.

Sampling

As it was stated earlier, testing all $g \in G$ and observing the value of *h(g)* is hopeless. Practically, a sample subset G^* of G can be only used. The subset is created by assigning a *selection probability*, denoted as *t(g)* to each element of G, such that $\forall g \in G$, $t(g) > 0$ and $\sum_{g \in G} t(g) = 1$. If $t(g) \equiv 1/|G|$, the sampling is *uniform*; if $t(g) \equiv p(g)$, the sampling is *representative*.

If Γ_i denotes the ith from the n elements of G^*, for c the following unbiased point estimator [2] can be given: $\hat{c}(\Gamma_1, \cdots, \Gamma_n) = \frac{1}{n} \sum_{i=1}^{n} h(\Gamma_i) \frac{p(\Gamma_i)}{t(\Gamma_i)}$. In the case of representative sampling, $\hat{c} = \frac{1}{n} \sum_{i=1}^{n} h(\Gamma_i)$, which is equal to the ratio of faults in G^* that were handled properly. In this case, the fault coverage estimated from G^* equals to the global fault coverage, so the sample G^* actually *represents* the behaviour of the whole system.

If the number of unhandled faults, called *deficiency number*, is denoted as X, the estimator for the *non-coverage factor* $\bar{c}$=1-c can be easily formulated as $\hat{\bar{c}}(X) = \frac{X}{n}$. As the non-coverage factor is easier to deal with, it will be used instead in the following.

Stratification

Stratification means partitioning the input space G into M disjoint classes (or strata). The definition of coverage factor can be rewritten into the following form:

$$c = \sum_{i=1}^{M} \sum_{g \in G_i} h(g)p(g) = \sum_{i=1}^{M} p(g \in G_i) \sum_{g \in G_i} h(g)p(g \mid g \in G_i) = \sum_{i=1}^{M} p_i c_i$$

where p_i is the relative probability of fault/activation pairs in class G_i and c_i is the coverage factor in class G_i. If the p_i values are grouped into the vector **p** and the non-coverage factors are grouped into the vector $\bar{\mathbf{c}}$, the overall non-coverage can be formulated as $\bar{c} = \mathbf{p} \cdot \bar{\mathbf{c}}^{\mathrm{T}}$.

In a stratified fault injection experiment, faults and activation sequences are injected in the different strata independently and the deficiency numbers X_i are composed to a deficiency vector **X**. One of the important issues is the distribution of the injections among the strata if the overall number of injection experiments is fixed.

The result of the experiment can be described with two vectors: $\mathbf{n} = [n_1, ..., n_M]$ is the number of injections in each stratum and $\mathbf{X} = [X_1, ..., X_M]$ is the observed deficiency numbers. The most common injection number distributions (or *sample allocations*) are the $\mathbf{n}_{\mathrm{H}}$ *homogenous* allocation where $n_i = n/M$, and the $\mathbf{n}_{\mathrm{R}}$ *representative* allocation, where $n_i = p_i n$.

The overall deficiency number can be estimated from the values of X_i; the most common estimators are the *arithmetic* ($Y_A = \frac{1}{n}\sum_{i=1}^{M} X_i$) and the *weighted* ($Y_W = \sum_{i=1}^{M} \frac{p_i}{n_i} X_i$) average estimator. Combining these possibilities, the following techniques can be considered for estimating the overall non-coverage:

- Y_A with $\mathbf{n}_{\mathrm{H}}$: $\hat{\bar{c}}_{AH}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{M} X_i = \frac{X}{n}$ (*naive estimator*[2]);
- Y_W with $\mathbf{n}_{\mathrm{H}}$: $\hat{\bar{c}}_{WH}(\mathbf{X}) = \frac{M}{n}\sum_{i=1}^{M} p_i X_i$;
- Y_A or Y_W with $\mathbf{n}_{\mathrm{R}}$: $\hat{\bar{c}}_{R}(\mathbf{X}) = \frac{1}{n}\sum_{i=1}^{M} X_i = \frac{X}{n}$ (*stratified estimator*).

It can be proven [2] that the variance of the stratified estimator, $\hat{\bar{c}}_{R}(\mathbf{X})$, is never greater than the variance of the estimator $\hat{\bar{c}}(X)$ for simple sampling. Thus, stratification provides an estimation that is not weaker than the simple one (in terms of variance) but requires significantly less computation due to the smaller input space sizes.

Confidence Limits

In the previous sections, (non-)coverage factor is treated as a random variable with some distribution and its value is estimated. Therefore the validity of these estimations must be evaluated and confidence measures must be determined.

An 100γ % *upper confidence limit* $\bar{c}_{\gamma}^{\uparrow}(X)$ for the non-coverage factor $\bar{c}$ can be defined as a value that is in 100γ % of the time greater than $\bar{c}$, supposed a large number of concrete values for X are considered.

The problem with confidence calculation is that the distribution of the coverage factor is not known. Based on the central limit theorem, the classic method in this case is the *approximation* with a normal distribution around of the expected value. Thus the upper confidence limit can be formulated as

[2]It is called 'naive' since it results in a biased estimation.

$$\bar{c}_\gamma^{\uparrow}(X) = \hat{\bar{c}}(X) + z_\gamma \sqrt{Var\{\hat{\bar{c}}(X)\}}$$

where z_γ is the 100γth standard normal percentile. In the case of stratified sampling, the individual non-coverage factors can similarly be approximated with normal distributions.

These confidence limit estimations, however, are valid only for very large total sample sizes. Moreover, the higher the coverage factor is, the higher is the probability that the deficiency number in one or more classes becomes zero. The above formula implies that these classes do not contribute any uncertainty to the overall estimate, which is practically untrue.

A more accurate confidence estimation can be created if the original sampling distribution is used instead of the approximated one. In this case, however, the mentioned definition of the upper confidence limit cannot be used; the new term of *confidence region* must be defined.

A 100γ % confidence region for the non-coverage vector $\bar{\mathbf{c}} = [\bar{c}_1, \ldots, \bar{c}_M]$ is defined as a function $I_\gamma(\mathbf{X})$ of the deficiency vector **X** such that for any given value of $\bar{\mathbf{c}}$, $I_\gamma(\mathbf{X})$ will contain the value of $\bar{\mathbf{c}}$ with a probability of γ. The upper confidence limit then can be defined as the upper frontier value of the confidence region. From the formula for overall non-coverage in the case of stratified sampling, a 100γ % upper confidence limit is $\bar{c}_\gamma^{\uparrow}(\mathbf{X}) = \max_{\bar{\mathbf{c}} \in I_\gamma(\mathbf{X})} (\mathbf{p} \cdot \bar{\mathbf{c}}^{\mathrm{T}})$. Thus the calculation of upper confidence limit is converted to a maximization problem.

Simulation experiments on theoretical systems [3] with pre-determined fault coverage measures show that this maximization problem can easily be solved with appropriate numerical methods, and the estimated and the pre-determined fault coverage factor is in a good agreement in most of the cases. The incorporation of stratified fault coverage calculations into the FUTEG ATPG system is currently under development.

Conclusions

In this paper, a concept of *stratified sampling* was described. This method provides fault coverage measures for a complex system from measures given for the individual components, using statistical estimation techniques. It fits very well to the hierarchical test generation scheme used in the FUTEG high-level ATPG program, and, according to the currently running experiments, it can be used for evaluation and optimization of test generation heuristics.

References

[1] D. Powell, M. Cukier and J. Arlat: *On Stratified Sampling for High Coverage Estimations.* Proc. of the 27^{th} International Conference of Fault-Tolerant Computing (FTCS-27), pp. 62-81, IEEE Computer Society Press, 1997.

[2] D. Powell, E. Martins, J. Arlat and Y. Crouzet: *Estimators for Fault Tolerance Coverage Evaluation.* Proceedings of the 23^{th} International Conference of Fault-Tolerant Computing (FTCS-23), pp. 228-237, IEEE Computer Society Press, 1993.

[3] M. Cukier: *Estimation of the Coverage of Fault-Tolerant Systems.* National Polytechnic Institute, Toulouse, France, 1996.

[4] A. Petri, B. Sallay, A. Pataricza, K. Tilly: *BudaTest: An architectural level heuristics-driven test generation tool.* Technical Report FUTEG-2/1996 (FUTEG PECO Project 9624), 1996, Budapest.