Partial error tolerance

Vladimir Simić
Computer Science Department
Faculty of Electronic Engineering
University of Nis
Serbia

DAAD Workshop “Embedded System Design”
Nis, Serbia
5 – 9 July 2010
Partial Defect Tolerance

- Concept
- Formal definition
- Example of bit plane array
- Concluding remarks
Compromising

Area – Time – Defect Tolerance

Goal:

- to achieve required defect tolerance, with optimal chip area or optimal using of chip resources, with acceptable time requirements
Design considerations

As scaling approaches the physical limits of devices and fabrication technology, designers will increasingly have to consider qualitative changes.

The key concerns include increasing process variations, defect rates, and infant mortality rates.

Dealing with defects, errors and faults

As VLSI scaling continues along its traditional path, we will soon be in a situation where chips will have billions of devices and thousands of defects.

Defects, errors, faults

Defect → Error → Fault
Fault tolerance

- Defect tolerance (DT) is the property that enables a system to continue operating properly in the event of a fabrication defect on some of its components.

- Fault tolerance (FT) is the property that enables a system to continue operating properly in the event of failure of some of its components, at the cost of hardware, time or information quantity.

An interesting question is: if some signal processing device has a minor hardware defect, will it still produce results that are good enough for the end user? If so, they could also be sold rather than be discarded.

Error tolerance is an alternative concept. Error tolerant systems neither detect nor correct error. The circuit is said to be an Error-Tolerant (ET), with respect to an application, if (1) it contains defects that cause internal and may cause external errors, and (2) the system that incorporates this circuit produces acceptable results.

Relaxing the requirement of 100% correctness for devices and interconnections may dramatically reduce costs of manufacturing, verification, and testing (and they can be sold rather than be discarded).

Partial error tolerance (PET) is a compromise between the FT and ET.

PET allows the fabrication process improvement (yield) at the cost of additional chip area.

Formal definition of PET

\[ V \text{ – set of vertices} \quad E \text{ – set of edges} \]
Basic terms

Let the architecture be represented by the data flow graph $G$.

Given a directed graph $G = (V, E)$, where $V = \{v_1, v_2, ..., v_n\}$ is a finite set of vertices and $E$ is a finite set of edges;

An edge $e \in E$ is an ordered pair $(v_i, v_j)$, where $v_i, v_j \in V$ and an edge $(v_i, v_j)$ means that vertices $v_i$ and $v_j$ are connected.
In order to define PDT we will define:

- Error propagation
- Error significance
- Error significance map
- Partial error tolerance
Def. 1. Error propagation


Let $v_i$ and $v_j$ be vertices and let $e_{i,j}$ denote $(v_i, v_j) \in E$.

We define an error propagation as the relation

$$\xi \subseteq V^2, (v_i, v_j) \in \xi.$$ 

The fact $(v_i, v_j) \in \xi$ we denote by $\xi_{i,j}$, which holds if an error which accrues within node $v_i$ causes an error within an error-free node $v_j$.

Note: Error propagation is a transitive relation.

$(v_1, v_9) \in \xi$, i.e. $\xi_{1,9}$
Def. 2. Error significance

Let $Y = \{y^0, y^1, \ldots, y^{0^\perp}\}$, $Y \subseteq V$ be the subset of architecture's output nodes.

We call the set $M_\eta$ error significance for the output bit $y^\eta$, iff

$$v_i \in M_\eta \iff (v_i, y^\eta) \in \xi.$$  

Note: Error significance marks the part of the architecture that must be error free in order to have output bit $y^\eta$ error free.
Let two dimensional ordering of an architecture be defined as a function $f_o: \mathbb{V} \rightarrow \mathbb{N}^2$, where $\mathbb{N}$ is a set of natural numbers. We call the function $f_o$ the ordering function.

We define Error Significance Map, $M_\eta=(m_{p,q})$, for the output bit $y^\eta$, as a matrix with elements:

$$m_{p,q}^\eta = \begin{cases} 1, & \exists v_i \in M_\eta, f_o(v_i) = (p,q) \\ 0, & \forall v_i \in M_\eta, f_o(v_i) \neq (p,q) \end{cases}$$

$$M_4 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$
We define a degree of Partial Error-Tolerance (PET) of an architecture as a function from the set \( \{0,1,...,l_0-1\} \) into the subset of \( V \), such that

\[
P_{ET}(\alpha) = \bigcup_{\eta=l_0-\alpha}^{l_0-1} M_\eta
\]

\[
P_{ET}(2) = M_4 \cup M_3
\]
Having in mind that error propagation is a transitive relation, the error significance map of the BP array can be obtained from transitive closure, which gives information about all paths within the array.
Transitive closure of bit-plane array

\[
A^* = \begin{bmatrix}
0 & G_C & G_C^2 & \ldots & G_C^{m\cdot k_C} \\
0 & 0 & G_C & \ldots & G_C^{m\cdot k_C - 1} \\
& & \vdots & & \vdots \\
0 & 0 & 0 & G_C & \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

\[
\left(g_{i,j}^C\right)^d = \begin{cases} 
1, & j + d \geq i \geq j \\
0, & \text{other}
\end{cases}
\]

Example of bit-plane array
Implementation results

<table>
<thead>
<tr>
<th></th>
<th>$k_C$</th>
<th>$m$</th>
<th>$n$</th>
<th>$l_0$</th>
<th>$m \cdot k_C$</th>
<th>$P_{ET}(\alpha)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arr1</td>
<td>4</td>
<td>8</td>
<td>8</td>
<td>16</td>
<td>32</td>
<td>$0 \leq \alpha \leq 16$</td>
</tr>
<tr>
<td>Arr2</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>24</td>
<td>32</td>
<td>$0 \leq \alpha \leq 24$</td>
</tr>
<tr>
<td>Arr3</td>
<td>4</td>
<td>8</td>
<td>24</td>
<td>32</td>
<td>32</td>
<td>$0 \leq \alpha \leq 32$</td>
</tr>
</tbody>
</table>

Table 1. Parameter sets for implemented arrays

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>Arr1</th>
<th>Arr2</th>
<th>Arr3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of cells</td>
<td>Impl. [kG]</td>
<td>No. of cells</td>
</tr>
<tr>
<td>32</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>24</td>
<td>/</td>
<td>/</td>
<td>2304</td>
</tr>
<tr>
<td>16</td>
<td>1536</td>
<td>41.4</td>
<td>2218</td>
</tr>
<tr>
<td>8</td>
<td>1450</td>
<td>39.6</td>
<td>1962</td>
</tr>
<tr>
<td>4</td>
<td>1344</td>
<td>37.3</td>
<td>1770</td>
</tr>
<tr>
<td>2</td>
<td>1274</td>
<td>35.8</td>
<td>1658</td>
</tr>
<tr>
<td>1</td>
<td>1236</td>
<td>34.9</td>
<td>1598</td>
</tr>
<tr>
<td>0</td>
<td>512</td>
<td>19.4</td>
<td>768</td>
</tr>
</tbody>
</table>

Table 2. Number of basic cells required for array implementation, and actual gate count by FPGA implementation of different sizes of arrays with $\alpha$ as parameter
Implementation results

\[
\text{Saving} = \frac{\text{FFT} - P_{ET}}{\text{FFT}} \cdot 100\%
\]
Concluding remarks

- We proposed a concept of systems partially tolerant to errors, and presented the design of a partially error-tolerant bit-plane array.
- Partial tolerance to errors is employed allowing some of the cells to produce errors.
- The design of partially error tolerant bit-plane FIR filter is facilitated by deriving the error significance map for the bit-plane array. Starting the development with acceptable margins of error in the output result, we obtained the error significance map from the transitive closure of the BP array.
- The array cells out of the area marked by error significance map could produce errors, but without any significant influence on the high order bits of the resulting word.
- By introducing the concept of partially error-tolerant bit-plane array, designers achieve one more degree of tradeoff freedom.