Exploiting Error Detection Latency for Parity-based Soft Error Detection

Gökçe Aydos*
*University of Bremen
Bremen, Germany
goekce@cs.uni-bremen.de

Goerschwin Fey∗‡
‡German Aerospace Center
Bremen, Germany
goerschwin.fey@dlr.de

Abstract—Local triple modular redundancy (LTMR) is often the first choice to harden a flash-based FPGA application against soft errors in space. Unfortunately, LTMR leads to at least 300% area overhead. We propose a parity-based error detection approach, to use the limited resources of space-proven flash-based FPGAs more area-efficiently; this method can be the key for fitting the application onto the FPGA.

A drawback of parity-based hardening is the significant impact on the critical path. To alleviate this error detection latency, pipeline structures in the design can be utilized. According to our results, this eliminates from 22% to 65% of the critical path overhead of the unpipelined error detection. Compared with LTMR, the new approach increases the critical path overhead of LTMR by a factor varying from 2 to 7.

I. INTRODUCTION

Field-programmable gate arrays (FPGAs) are often utilized in space avionics. The avionics must be protected from ionizing radiation in space. In the absence of a shield (e.g., magnetic field of the earth), a high energy particle can traverse through a digital circuit and induce significant amount of charge, which can cause soft errors. These errors are not permanent and can be corrected e.g., by a reset. In flash-based FPGAs, soft errors mainly happen in the flip-flops (FFs) of an FPGA application in form of bitflips. The FPGA configuration bits do not have to be protected, because flash memory has a negligible soft error rate.

One of the popular flash-based FPGAs is the ProASIC3 [1]. This FPGA has 130 nm feature size and a lower logic density compared to SRAM-based FPGAs in the market, but it is the state-of-the-art FPGA for space applications ([2], [3]). This FPGA does not have inherent protection against soft errors in FFs, and the standard hardening solution is local triple modular redundancy (LTMR). In LTMR, the application FFs are triplicated and their outputs are voted, which protects against single bitflips. LTMR is shown in Figure 1. Unfortunately, triplication has a significant area overhead of at least 300%.

Hardening a circuit against soft errors on design level is mostly done by redundancy (e.g., in LTMR by space redundancy). LTMR detects and corrects errors locally, which comes at the cost of additional space. Alternatively, a part of the area redundancy in the FPGA may be eliminated by combining space and time redundancy, such as: only error detection on hardware and on-demand transaction-retry (i.e., recomputation) on software, if the error rate is less than a dozen bitflips per year. In our previous work [4], we evaluated a parity-based error detection (PBED) approach to use the limited resources of space-proven flash-based FPGAs more area-efficiently, which can be the key for fitting the application onto the FPGA.

Triplication [5] and parity-based code, which is a concurrent error detection technique (CED) [6], are well-known. Recomputation for achieving error correction has also been proposed [7], [8].

In our previous work [4], we applied LTMR and PBED-based hardening on a reference architecture and experimentally compared circuit area overhead, maximum frequency and the needed processing time using an example mission under fault injection. The results for a fixed cluster size showed that at least 30% of the area overhead caused by LTMR can be saved by implementing PBED and correcting the errors with time redundancy. In this previous implementation the impact on the critical path of the circuit was significant. Our contributions in this work consist of:

• a pipelined error detection and handling approach to alleviate the impact of PBED on the critical path of the circuit
• the concept of a parity-based hardening tool, which implements this approach
empirical comparison of the pipelined- with approach from our previous work [4]

empirical comparison of PBED and LTMR using the new approach

In the following sections, we will firstly present how PBED is applied on an example processing architecture and the system requirements for applying PBED. In Section III, we introduce the pipelined error detection approach and describe how to apply it on a circuit automatically. In Section IV, we compare pipelined error detection to our previous approach and LTMR using experimental results.

II. ERROR DETECTION BASED HARDENING ON AN EXAMPLE DESIGN

LTMR detects and corrects the bit error locally in the same clock cycle, but PBED detects an error on the module level and this error can live for further cycles until correction. In the latter approach, there is an error correction latency and this latency places additional demands on the system. These are discussed in the following example.

Figure 2 shows an on-board data handling architecture. The processor runs the mission software and uses the FPGA as an extensible interface module for communicating with the subsystems. In this architecture the processor and subsystems are assumed to be sufficiently hardened against soft errors, but the FPGA must be hardened by design.

The subsystems are manipulated via memory access, i.e., the mission software sends a data packet, which is written to a particular memory area in the FPGA. In turn, this memory access can trigger an action in the subsystem. The circuit in the FPGA for decoding the data packets are shown in Figure 3. Circuit (A) queues packets sent by the processor, circuit (B) decodes the packets and writes to respective memory addresses in circuit (C). Circuits (A) and (C) are assumed to be sufficiently hardened (e.g., by using LTMR) and circuit (B) is hardened using the gray components. Figure reused from [4].

Sending a response is not only important for flow control but also for detecting an error. Due to the fact that we allow errors in circuit (B) that cannot be corrected immediately, the software must use transaction-based processing. This gives the software the opportunity to repeat the last processing request (i.e., resend the last packet) after a timeout, if circuit (B) cannot send any response due to a recovery event. Transaction-based processing is visualized in Figure 4.

The goal of isolation is that an error in circuit (B) does not propagate to the rest of the system. This can be achieved by masking the output signals. If the circuit interface includes control signals, further resources can be saved by only masking the control signals like write- and read-enable in Figure 3.

III. PIPELINED ERROR DETECTION AND HANDLING

In PBED, the error signal can be generated directly or by using pipelining, where the latter gives better timing results with insignificant area overhead. In this section, we explain the latter technique more in detail and show a comparison with direct error signal generation based on synthesis results. Finally, we present a toolchain to implement these techniques on an FPGA.

A. Pipelining

PBED is based on two modules: error detection and error handling. The error detection module itself consists of many smaller error detection clusters and additional circuitry which reduces all the error signals output by the clusters to a single error signal. Figure 5 shows the structure of an error detection cluster $\text{cluster}_{\text{ED}}$. One cluster consists of $k$ application FFs $\text{FF}_{a}$, one parity FF $\text{FF}_{p}$ and two XORs: one for parity
that a bit needs to be visible at PO. For example, a FF whose
word is written to memory. In this work, the approach is called
five cycles later. In other words, in the same cycle when this
signals are activated, and the data word is written, then it
instruction takes five cycles before corresponding memory
detection on the module level. For example, if a memory write
This latency introduced by a circuit can be exploited for error
detection. Generally, a PBED-hardened circuit contains many of these
clusters, whose error outputs can be reduced to a single error
output is a PO of the circuit has \( d_{\text{seq}} = 0 \). FFs with \( d_{\text{seq}} = d \) belong to a particular error detection stage, which is named
stage \( d_{\text{ED}} \). These stages are visualized in Figure 7.

The inner structure of a stage is shown in Figure 8. Analogous to direct error detection, the FFs are grouped in clusters within a stage. Stages contain an error FF \( \text{FF}_{e} \), which stores the error signal that is coming from the previous stage, with the exception of the leftmost stage with the greatest \( d_{\text{seq}} \). The error signal of stage \( d_{\text{ED}} \), \( \text{error}^{d} \), is generated by ORing the buffered error signal from the last stage and the error signals from the clusters within the stage.

The error signal generated by direct and pipelined error detection (error and error)\(^{d} \) are fed to the error handling module. The error module is responsible for (1) recovering the circuit from the unknown state and (2) isolating the circuit from the rest of the system while it is in the unknown state. An example implementation is shown in Figure 7. Isolation is realized by masking the appropriate nets from the POs of the bare circuit (PO\(_{\text{bare}} \)), which can alter the state of the system, e.g., the control signals of a memory interface. The masking starts in the same cycle when the error signal is visible at PO and remains active until the circuit is reset. As long as the circuit is in isolated state, the circuit can be asynchronously reset with the help of a shift register. The number of FFs in the shift register must be chosen such that all FFs in the circuit are guaranteed to be reset after the respective number of reset cycles.

B. Algorithm

PBED can be implemented on-top of a technology-level netlist using an automatic tool\(^{1} \). The pseudo code of the PBED application program is shown in Algorithm 1. Before processing, the netlist needs to be parsed, for which we used Verilog-Perl\(^{9} \). Then, all the FFs with enable input must be replaced with a basic FF and a multiplexer, which is normally also

\(^{1}\)Tool available at https://gitlab.informatik.uni-bremen.de/goekce/pbed
Fig. 9. Example implementation of error handling module. When the error signal is active then the control signals are masked to isolate the circuit in the same cycle. In subsequent cycles, the reset signal is hold active and the circuit falls back to reset state.

done in LTMR. The multiplexer emulates the enable behavior by switching between the output of the FF and the input data which must be fed to the FF when enable signal is active. This is crucial, because an enable FF is not updated in every cycle, but only when the enable input is activated. If a soft error happens on enable FFs, these errors can eventually accumulate and are undetectable for even number of bitflips in a cluster.

If pipelined error detection is used, \( d_{seq} \) for every FF must be determined. For this purpose, a FF-only dataflow graph is generated by setting the POs as sink vertexes and exploring the design using breadth-first search and only adding the FFs to the graphs. While traversing, the FFs are annotated with \( d_{seq} \) to each particular PO. Subsequently, the minimum of these \( d_{seq} \)'s is determined, which corresponds to \( d_{seq} \) to the output.

In the next step, the FFs are put to clusters according to cluster size, and clock and reset signals of FFs. The FFs in a cluster must be sensitive to the same edge. Furthermore, all the FFs in a cluster must have the same reset type: all active-low or -high. These countermeasures enable the connection of the parity FF to the same clock and reset signal of the application FFs in the cluster possible. After the clusters are generated, the generation of stages is done according to Section III-A.

In case of direct error detection, \( d_{seq} \) does not play a role, therefore only clusters are created and the cluster errors are reduced.

Finally, the netlist must be recompiled along with the error handling module. Before that, the synthesizer must be instructed to not optimize the redundancies away, because the tools normally presume no external effects (e.g. bitflips) during optimizations. The error handling module is designed manually using HDL. After compilation, the module is ready to be verified under error injection.

\section*{C. Overall toolflow}

In previous sections, we explained the concept of pipelined- and direct error detection in detail. The implementation of these techniques requires further attention regarding the design flow to follow. In this subsection, we show the overall toolflow that we used with some remarks.

Figure 10 shows an overview of our toolflow, which is partly analogous to classic FPGA design flow. Often, verification of the technology netlist is skipped in the FPGA design, and

\begin{algorithm}
\begin{algorithmic}
\State \textbf{Data:} technology netlist
\State \textbf{Result:} PBED applied technology netlist
\State \textbf{foreach FF do}
\State \quad \textbf{if} \hspace{1mm} \text{has enable input} \hspace{1mm} \textbf{then}
\State \quad \quad \text{eliminate enable input;}
\State \quad \textbf{end}
\State \textbf{end}
\State \textbf{if pipelined error detection then}
\State \quad \textbf{foreach primary output (PO) do}
\State \quad \quad build a FF dataflow graph with this PO as sink vertex;
\State \quad \textbf{end}
\State \textbf{end}
\State \textbf{foreach FF do}
\State \quad \textbf{if pipelined error detection then}
\State \quad \quad determine \( d_{seq} \) to output by using the FF dataflow graphs;
\State \quad \quad categorize according to clock-, reset-signal and \( d_{seq} \) to output;
\State \quad \else
\State \quad \quad categorize according to clock- and reset-signal;
\State \quad \textbf{end}
\State \quad \textbf{push to a cluster according to category;}
\State \textbf{end}
\State \textbf{foreach cluster do}
\State \quad add parity-generation and -check circuitry;
\State \textbf{end}
\State \textbf{if pipelined error detection then}
\State \quad \textbf{for} \hspace{1mm} \text{sequential distance} \hspace{1mm} (\( d_{seq} \)=max to 0) \hspace{1mm} \textbf{do}
\State \quad \quad put clusters with \( d_{seq} \) to a new stage;
\State \quad \quad reduce cluster error signals to a single error signal;
\State \quad \quad merge the error signal from the previous stage;
\State \quad \quad add an error FF to the stage;
\State \quad \textbf{end}
\State \else
\State \quad \textbf{reduce cluster error signals to a single error signal;}
\State \textbf{end}
\end{algorithmic}
\caption{Application of PBED to a technology netlist}
\end{algorithm}

the design is tested directly on an experimental board or real hardware. This also applies to LTMR, because it is a straightforward approach to harden a circuit against soft errors. But verification of complex redundancy techniques needs a verification on the technology level under error injection. This approach is also used for PBED, in which error handling module can be implemented differently for different target circuits.

To verify the hardened circuit against soft errors time-efficiently, the testbench should be designed in such a way that:

\begin{itemize}
\item the RTL-design under test can be easily replaced with a technology netlist, which is mostly achieved by operating on the primary inputs and outputs of the design under test,
\end{itemize}
IV. SYNTHESIS RESULTS AND COMPARISON WITH LTMR

In this section, we compare pipelined error detection to our previous approach and LTMR using experimental results.

A. Comparison of pipelined- with direct-error detection

To show the positive effect of pipelined error detection on the critical path of a PBED-applied design, we synthesized a generic shift register without any combinatorics using following input parameters:

- data width \( w_d \) = \{8, 16, 32, 64\}
- stage count \( c_{st} \) = \{4, 8, 16, 32\}

Then, we applied PBED with direct (PD) and pipelined (PP) error detection on the netlists with cluster size \( s_{cl} = 3 \) and gathered the following output parameters for both approaches:

- combinational area overhead \( A_{comb+} \)
- FF area overhead \( A_{FF+} \)
- total area overhead \( A_+ = A_{comb+} + A_{FF+} \)
- critical path overhead \( t_{c+} \)

Using the \( t_{c+} \), we additionally determined how much of the critical path overhead of PD can be saved by using PP by calculating \( 1 - \frac{t_{c+}}{t_{c+PP}} \), which we will refer as \( t_{c+} \) saving.

The synthesis was done for the ProASIC3 FPGA A3PE3000L using Precision (Mentor Graphics) for synthesis and Designer (Microsemi) for placing and routing. The results are shown in Table I.

<table>
<thead>
<tr>
<th>( c_{st} )</th>
<th>( w_d )</th>
<th>( A_{comb+} )</th>
<th>( A_{FF+} )</th>
<th>( A_+ )</th>
<th>( t_{c+} ) (ns)</th>
<th>( 1 - \frac{t_{c+}}{t_{c+PP}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>8</td>
<td>51</td>
<td>18</td>
<td>7.84</td>
<td>33.21%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>94</td>
<td>34</td>
<td>8.79</td>
<td>22.32%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>32</td>
<td>177</td>
<td>66</td>
<td>9.89</td>
<td>21.75%</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>346</td>
<td>128</td>
<td>21.75%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>94</td>
<td>34</td>
<td>7.84</td>
<td>43.06%</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>177</td>
<td>66</td>
<td>9.89</td>
<td>41.01%</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>346</td>
<td>128</td>
<td>21.75%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>64</td>
<td>693</td>
<td>256</td>
<td>40.85%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>8</td>
<td>177</td>
<td>66</td>
<td>9.28</td>
<td>35.74%</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>16</td>
<td>346</td>
<td>128</td>
<td>41.01%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>32</td>
<td>693</td>
<td>256</td>
<td>40.85%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>64</td>
<td>1369</td>
<td>512</td>
<td>40.48%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>8</td>
<td>346</td>
<td>128</td>
<td>8.80</td>
<td>65.10%</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>16</td>
<td>693</td>
<td>256</td>
<td>63.03%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>32</td>
<td>1369</td>
<td>512</td>
<td>56.06%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>64</td>
<td>2735</td>
<td>1024</td>
<td>49.40%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE I**

PBED SYNTHESIS RESULTS FOR DIRECT (PD) AND PIPELINED (PP) ERROR SIGNAL PROPAGATION WITH FIXED CLUSTER SIZE \( s_{cl} = 3 \),

overwritten in the next cycle.

For error injection in our simulation flow, we used the Foreign Language Interface [10] for the simulation software Questa (Mentor Graphics), which allows access to the netlist elements and allows to simulate bitflips. In reality, a soft error can normally be overwritten in the next cycle by the new data. To reproduce this behavior in the simulation, we flipped the target bit by simply setting the signal to the inverted value (instead of forcing and releasing a signal), which can be visualized in Figure 11.
The following parameters are shown in Table II. In addition to the known parameters from Table I, configurations from Subsection IV-A. The results are shown in Table II.

---

**TABLE II**  
SYNTHESIS RESULTS FOR LTMR AND PIPELINED PBED WITH CLUSTER SIZE \( s_{cl} = 3 \)

<table>
<thead>
<tr>
<th>( s_{cl} )</th>
<th>( A_{\text{comb}+} )</th>
<th>( A_{\text{FF}+} )</th>
<th>( A_{+} )</th>
<th>( t_{c+} ) (ns)</th>
<th>( 1 - \frac{A_{+} + \frac{A_{\text{FF}+}}{A_{\text{comb}+}}}{A_{\text{FF}+}} )</th>
<th>( t_{c+} - t_{c+} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>369</td>
<td>265</td>
<td>634</td>
<td>6.38</td>
<td>17.45%</td>
<td>6.79</td>
</tr>
<tr>
<td>3</td>
<td>349</td>
<td>137</td>
<td>486</td>
<td>6.16</td>
<td>36.72%</td>
<td>6.54</td>
</tr>
<tr>
<td>4</td>
<td>312</td>
<td>97</td>
<td>409</td>
<td>6.55</td>
<td>46.74%</td>
<td>6.96</td>
</tr>
<tr>
<td>5</td>
<td>350</td>
<td>73</td>
<td>423</td>
<td>6.34</td>
<td>44.92%</td>
<td>6.74</td>
</tr>
<tr>
<td>6</td>
<td>348</td>
<td>65</td>
<td>413</td>
<td>5.73</td>
<td>46.22%</td>
<td>6.09</td>
</tr>
<tr>
<td>7</td>
<td>311</td>
<td>57</td>
<td>368</td>
<td>6.63</td>
<td>52.08%</td>
<td>7.04</td>
</tr>
<tr>
<td>8</td>
<td>345</td>
<td>49</td>
<td>394</td>
<td>6.05</td>
<td>48.70%</td>
<td>6.43</td>
</tr>
<tr>
<td>9</td>
<td>343</td>
<td>41</td>
<td>384</td>
<td>5.65</td>
<td>50.00%</td>
<td>6.96</td>
</tr>
<tr>
<td>10</td>
<td>326</td>
<td>41</td>
<td>367</td>
<td>5.37</td>
<td>51.21%</td>
<td>6.92</td>
</tr>
<tr>
<td>11</td>
<td>331</td>
<td>41</td>
<td>372</td>
<td>5.37</td>
<td>51.21%</td>
<td>6.77</td>
</tr>
<tr>
<td>12</td>
<td>321</td>
<td>33</td>
<td>354</td>
<td>6.36</td>
<td>53.91%</td>
<td>6.76</td>
</tr>
</tbody>
</table>

---

**TABLE III**  
Pipelined PBED SYNTHESIS RESULTS FOR STAGE COUNT \( c_{st} = 8 \), DATA WIDTH \( w_{d} = 32 \), AND VARIOUS CLUSTER SIZES \( s_{cl} \).

B. Comparison with LTMR

We have shown that PP delivers better results than PD regarding \( t_{c+} \). In this section we compare PP with the state-of-the-art hardening approach LTMR (LT) using the same shift register configurations from Subsection IV-A. The results are shown in Table III. In addition to the known parameters from Table I, the following parameters are shown:

- critical path of the unhardened circuit \( t_{c} \)
- area overhead per user FF \( \frac{A_{+}}{A_{\text{FF}+}} \) (BA: bare circuit)
- area overhead saving if PP is used instead of LTMR \( 1 - \frac{A_{+} + \frac{A_{\text{FF}+}}{A_{\text{comb}+}}}{A_{\text{FF}+}} \)
- critical path overhead factor of PP \( \frac{t_{c+} - t_{c}}{t_{c+}} \)

According to the results, at least 29% of the area overhead caused by LTMR can be saved by using PP. If \( w_{d} \) increases, PP can save more area, but it has a negative impact on \( t_{c+} \). PP has up to 6.5% critical path overhead reduction. Note that we compared not the absolute critical path, which determines the maximum frequency, but only the overhead, which adds up to the critical path of the bare circuit.

We gathered additional results by setting \( c_{st} = 8 \) and \( w_{cl} = 32 \) and varying \( s_{cl} \) between 2 and 12. PP saves more area with increasing \( s_{cl} \), but \( s_{cl} \) does not have a significant effect on \( t_{c+} \).

---

V. CONCLUSION

We have presented a new approach for PBED by exploiting pipeline structures in a circuit, and shown that it can save from 22% to 65% of the critical path overhead caused by the direct error detection approach. LTMR has still better timing performance and the pipelined PBED increases the critical path overhead by a factor of 2 to 7. As future work, PBED will be applied on a soft-processor core to get results on a more complex design. Formalization of transaction-based processing is required to safely utilize PBED in a system; this is also planned as future work.

---

REFERENCES