Modeling Common Cause Failures in Systems with Triple Modular Redundancy and Repair

Matthew J. Cannon, Ph. D., Brigham Young University
Andrew M. Keller, Brigham Young University
Andrés Pérez-Celis, Brigham Young University
Michael J. Wirthlin, Ph. D., Brigham Young University

Key Words: triple modular redundancy (TMR), systems with repair, common cause failure, continuous time Markov chains

SUMMARY & CONCLUSIONS

Triple modular redundancy (TMR) is commonly employed to increase the reliability and mean time to failure (MTTF) of a system. This improvement can be shown by using a continuous time Markov chain. However, typical Markov chain models do not model common cause failures (CCF), which is a singular event that simultaneously causes failure in multiple redundant modules.

This paper introduces a new Markov chain to model CCF in TMR with repair systems. This new model is compared to the idealized models of TMR with repair without CCF. The fundamental limitations that CCF imposes on the system are shown and discussed. In a motivating example, it is seen that CCF imposes a limitation of 51× on the reliability improvement in a system with TMR and repair compared to a simplex system, (i.e., without TMR). A case study is also presented where the likelihood of CCF is reduced by a factor of 18× using various mitigation techniques. Reducing the CCF compounds the reliability improvement of TMR with repair and leads to a overall system reliability improvement of 10,000× compared to the simplex system as supported by the proposed model.

1 INTRODUCTION

Triple modular redundancy (TMR) with repair is a common fault mitigation strategy for increasing the reliability of a system. Applying TMR allows the system to tolerate failures limited to one of the redundant modules. If multiple modules fail or are in a failure state at the same time, then TMR is defeated and the system is no longer protected. Provisioning a repair mechanism allows the system to correct itself. The system will operate correctly as long as the repair mechanism prevents failures from accumulating in multiple modules.

TMR with repair is very effective at increasing the reliability of a system; but when a single event causes multiple modules to fail, then no amount of repair can prevent the system from failing. TMR is often thought of as a catch all to protect any system from failure, and repair is often believed to monotonically improve the effectiveness of TMR overtime as the repair rate increases with respect to the failure rate. In truth, if only one module could fail at a time, then TMR with an increasing repair rate would improve the reliability of the system without bound. But single events can affect failure in multiple modules and thereby thwart TMR with repair as a system-level protection scheme.

This paper proposes a reliability model for common cause failure (CCF) in systems with TMR and repair, it examines the implications of CCF on such a system, and it presents an insightful case study that highlights the impact that CCF can have on a TMR system with repair and the benefits that can be obtained from mitigating CCF. When the repair rate in a TMR system with repair is much larger than the failure rate, then even a small likelihood of CCF can have significant impact on the reliability of the system. In fact, CCF imposes a fundamental limit on the reliability improvement that can be obtained by protecting a system with TMR and repair. This paper explores all of these facets and contributes novel insights into understanding the impact of CCF on systems with TMR and repair.

Using the proposed reliability model, the impact of CCF on the reliability of a system with TMR and repair can be quantified. CCF impacts the overall system reliability and it places a limit on the improvement in reliability that can be gained from increasing the repair rate of the system. Both aspects can be quantified using the proposed model.

2 MOTIVATION

TMR with repair has traditionally been modeled using a Markov chain [1]. Markov chains can be used to derive the theoretical continuous time reliability and mean time to failure (MTTF) of the system. The theoretical equations for the reliability and MTTF of TMR with repair show that both metrics should improve as the repair rate increases [2]. In fact, as the repair rate approaches infinity, the estimated MTTF also approaches infinity. This makes TMR with repair an attractive fault mitigation technique for systems where the repair rate is relatively much higher than the fault rate.

A limitation in the traditional TMR Markov chain model is that it assumes that a single fault will affect only one redundant module, thus there must be two separate module failure events

978-1-7281-3690-5/20/$31.00 ©2020 IEEE

to cause a system failure. Generally, this is how TMR fails, but under some circumstances it is possible for a single event to cause system failure or for a single fault to simultaneously affect multiple redundant modules. These types of events are often referred to as common cause failure (CCF) [3]. A Markov chain reliability model can be constructed for TMR and repair that would also take CCF into account by allowing the system failure due to a single fault. This paper seeks to adapt current Markov chain reliability models for systems with TMR and repair so that they also take CCF into consideration.

Mathematical models are often used to represent potential fault tolerant mitigation techniques. Markov chains are useful because metrics such as reliability as a function of time and the MTTF can be derived and analyzed. The typical TMR with repair system can be modeled using the Markov chain shown in Figure 1.

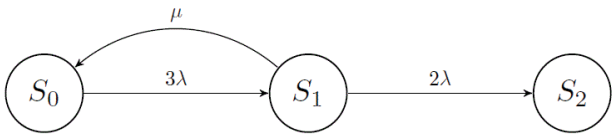

Figure 1 – TMR with Repair Markov Chain

There are three states in this Markov chain. The first state, S_0, is the normal operation state where all three TMR modules are operating correctly. The second state, S_1, is the impaired operation state where one of the TMR modules has failed. The third state, S_2, is the failed state where two or more of the TMR modules have failed. The states are connected by three arcs. The first arc transitions from S_1 to S_2 and represents a single module failing. This occurs at three times the module failure rate, λ. The second arc is from S_2 to S_1 and represents the module repair rate, μ. The third is from S_2 to S_3 and represents another module failure. This occurs at two times the module failure rate since there are only two correctly functioning modules in S_1.

Most of the mathematical models for TMR carry an inherit assumption: each of the redundant modules fail independently. This assumption exists since in the typical TMR with repair Markov chain there is no connection between the normal operation state, S_0, and the failed state, S_2. This implies that for the system to fail, it must pass through the impaired operation state, S_1, which requires two separate events for the system to fail, (i.e., a single event cannot cause the system to enter S_3).

This assumption becomes apparent when analyzing the MTTF of the TMR with repair system (see Equation 1). The maximum MTTF for any non-zero failure rate can be found by setting the repair rate, μ, to infinity (see Equation 2). An infinite repair rate suggests that any single TMR module failures are repaired instantaneously. At an infinite repair rate, the system cannot fail because whenever the system transitions into state S_1, it immediately transitions back into state S_0.

$$\mathrm{MTTF}_{\mathrm{Simplex}} = \frac{1}{\lambda};\ \mathrm{MTTF}_{\mathrm{TMR}} = \frac{5\lambda+\mu}{6\lambda^2} \quad (1)$$

$$\lim_{\mu\to\infty}(\mathrm{MTTF}_{\mathrm{TMR}}) = \infty \quad (2)$$

$$\lim_{\mu\to\infty}(\mathrm{Improvement}_{\mathrm{TMR}}) = \frac{\mathrm{MTTF}_{\mathrm{TMR}}}{\mathrm{MTTF}_{\mathrm{Simplex}}} = \infty \quad (3)$$

For many systems (if not all systems), it is possible for multiple modules to fail *simultaneously*. This would represent a single event that causes system failure. Because the modules fail simultaneously, there is no opportunity for a repair element to repair one of the modules before the other module fails. In the Markov chain, this translates to a connection between the normal and impaired operation states, S_0 and S_1, and the failed state, S_2. These types of failures are called common cause failures (CCF). CCF refers to any single event that simultaneously causes multiple TMR modules to fail.

This paper seeks to explore the implications of CCF on system with TMR and repair because real world systems experience this phenomenon and because current models do not adequately emphasis the impact that CCF can have on systems with TMR and repair. Specifically, the interplay between the likelihood of CCF, the failure rate of individual modules, and the repair rate of the system needs to be deciphered. As part of the motivation for this work, an insightful example is presented and related works are discussed.

2.1 Related Work

The concept of common cause of failures (CCF) has been considered on reliability modeling in nuclear and aviation industries for decades [4-5]. In nuclear plants, CCF has been modeled using the beta-factor model introduced in [6]. The beta-factor model assigns a probability of β to an event that causes failures in the remaining components. The β parameter can be seen as the fraction of failures that cause all components to fail. Thus, the system will have a CCF rate of $\lambda_{\mathrm{CCF}} = \beta\boldsymbol{\lambda}$; where $\boldsymbol{\lambda}$ is the failure rate of a single component.

In [3], a discussion around several extensions of the beta model is presented such as the multi beta-factor model, the multiple greek letter model, and the binomial failure rate model. All of the discussed models are presented in the context of power plants and lack the notion of a repair mechanism.

For TMR systems, voters can be seen as an example of a CCF. In [7], the notion of imperfect voters for TMR systems is discussed. The discussion shows that the small area the imperfect voters use compromises the reliability improvement provided by TMR. As for most of the previous work, the work on [7] does not cover the use of a repair mechanism while considering the imperfect voters.

In [8] authors present a method to compute the reliability in the presence of CCF. The method uses a direct modeling approach based on a Venn diagram that yields a linear function of the reliability. The model does not consider repair. Without considering repair, the limits imposed by the CCF failure rate on the reliability of a TMR system with repair cannot be examined.

In [9] the authors proposed a method to incorporate CCF in system analysis. Their method applies Markov modeling in dynamic fault trees. The resulting Markov model is a straightforward TMR model with additional transitions from the working state to the failure state. The model lacks a repair mechanism and the authors do not show an analysis of the proposed model.

Another Markov model for a TMR design with CCF is

presented in [10]. Their model is specific to their design which is a TMR system with active hardware redundancy which has fault-masking and detection. Their Markov model includes a repair mechanism but does not include CCF.

The model we propose is a Markov chain for TMR systems that include both CCF and repair. The proposed general model is compared to an ideal TMR model with and without a repair mechanism. This comparison provides insight into the fundamental limitations that CCF imposes to TMR systems with repair.

2.2 Motivating Example

Our proposed model is applicable to many systems. An interesting example is to apply this model to electronic circuits implemented on a field programmable gate arrays (FPGA). These circuits are subject to faults caused by ionizing radiation. Ionizing radiation can upset the values stored in the devices configuration memory cells, which can change in the functionality of the intended circuit and result in system failure [11]. Applying TMR to FPGA circuits protects the circuit from configuration upsets in radioactive environments such as space-based systems. Upsets in configuration memory can be repaired on-the-fly by continuously checking for and correcting upsets as they are encountered. The repair rate can be set very high on these systems compared to the upset rate, (e.g., one-hundred thousand checks to one upset or higher). In theory, this configuration should yield a system that is extremely reliable in the presence of harsh radiation.

Based on the traditional Markov models for TMR with repair, the reliability of an FPGA system using TMR with repair should be relatively high. However, the improvements in system failure rate measured in testing are much lower than expected [12]. After carefully analyzing the behavior of the system with a variety of artificial upsets, (i.e., purposeful corruption of configuration memory), it was found that some upsets cause two or more of the circuit modules to fail, (i.e., a CCF), which violated the assumption that a single fault can cause only one module to fail. TMR defeat has also been observed in situations where a single energetic atomic particle causes multiple configuration memory cells to upset at the same time [13].

Another study of a TMR circuit on an FPGA showed the limitations imposed by CCF [14]. The authors tested the circuit using a method called fault injection, where single faults are intentionally introduced into the circuit to observe the circuit behavior. A fault injection study essentially tests at an infinite repair rate, (as only one fault is ever present in the system at a given time). The ideal model of TMR suggests that no failures should be observed in the TMR circuit, and there should be an infinite improvement over the unmitigated circuit. Instead, the results in Table 1 show that the circuit only saw a 51x improvement in design sensitivity over the unmitigated circuit, and there were single faults that could cause TMR failure. This motivated us to create a way to model the behavior of CCF so we could more accurately estimate the improvement offered by TMR.

Table 1 – CCF in a TMR circuit on an FPGA [14]

Circuit/Metric	Unmitigated	TMR
Faults	1,831,859	29,443,885
Failures	6,501	2,037
Sensitivity	.355%	.00692%
Improvement	1.0x	51.3x

3 MODELING COMMON CAUSE FAILURE IN TMR

Modeling the impact of CCF on TMR systems requires adaptation of existing models. In this section, two different models are explored. First, a model is presented that considers the impact of CCF on a TMR system without repair. Second, a model is presented that considers the impact of CCF on a TMR system with repair. Both models aid the understanding of the impact that CCF has on the reliability improvement of a TMR system.

To model CCF in TMR systems, two additional arcs can be added to the Markov chain. The first arc is added from state S_0 to S_2 which represents direct TMR system failure from a single event. This can model any event that simultaneously affects two or more TMR modules. The second arc that needs to be added is from S_1 to S_2. Even when one TMR module has failed (which is the case in S_1), there are still events that can affect multiple domains, which needs to be accounted for. Both of these arcs have the CCF failure rate, i.e., λ_{CCF}.

In a more general context, λ_{CCF} can be analyzed with relation to the mode failure rate, λ. This can be done by employing a simple ratio using the variable ρ,

$$\lambda_{\mathrm{CCF}} = \rho\lambda. \tag{4}$$

Four different values of ρ were chosen for the analysis to explore how different rates of CCF affect the system. The four values are $\rho = (1, .1, .01, .001)$. The next two subsections explore how CCF affects TMR systems with and without repair.

3.1 TMR Without Repair

Figure *2* shows the Markov chain for TMR without repair, but with CCF. As previously explained, two arcs have been added to the traditional model, from states S_0 and S_1 to state S_2, with the CCF failure rate, λ_{CCF}.

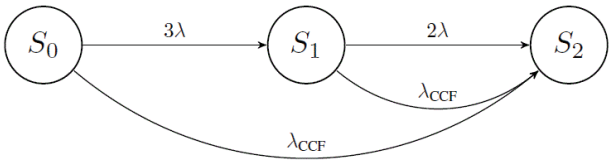

Figure 2 – TMR with CCF and No Repair

The Markov chain can then be used to derive the reliability functions [2], which have been plotted in Figure *3*.

This chart has three main takeaways:

- For values of $\rho \leq .1$, TMR with and without CCF are nearly identical;
- For values of $.1 \leq \rho < 1$, TMR with CCF *may* be better or worse than the Simplex system;

- For values of $\rho \geq 1$, TMR makes the system worse than Simplex.

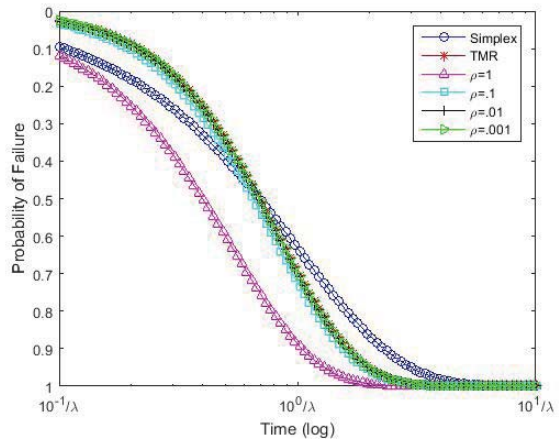

Figure 3 – Plots of TMR with CCF and No Repair

3.2 TMR With Repair

Similar to the TMR without repair model, the Markov chain for TMR with repair can be altered to account for CCF, as shown in Figure 4. Compared to the model in Figure 1, only the two new arcs are added from states S_0 and S_1 to S_2. Figure 5, shows the effects that CCF has on the reliability of a system with a high repair rate.

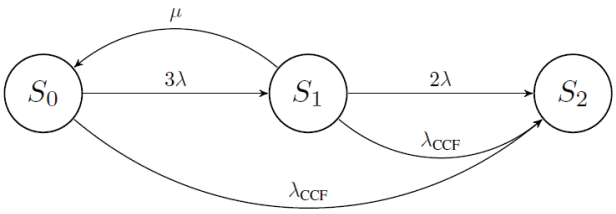

Figure 4 – TMR with CCF and Repair

There are a few trends that can be observed from these charts. One observation is the effect CCF has on the system as the CCF rate becomes larger. When $\rho = 1$ the system digresses back into the Simplex system. This trend is clearly observed in Figure 4 where the plots for the Simplex system and the TMR system with $\rho = 1$ are nearly identical. TMR will not be beneficial to the system if the CCF rate is too high.

As the CCF rate λ_{CCF} becomes lower than the module failure rate λ, the reliability over time of the TMR system increases. This varies according to how λ_{CCF} compares to λ. As $\rho \to 0$ the system approaches the reliability of the ideal TMR system with repair and no CCF. How fast it approaches the ideal TMR system depends on the repair rate μ relative to the failure rate λ.

Equations for the MTTF of a system with TMR, repair and CCF are also derived from the Markov model. Equation 5 gives the MTTF of such a system with respect to CCF ratio, ρ, the single module failure rate, λ, and the system repair rate, μ. Equation 6 gives the MTTF limit as the repair rate approaches infinity and Equation 7 shows how the improvement of a system with TMR and repair is limited by CCF rate.

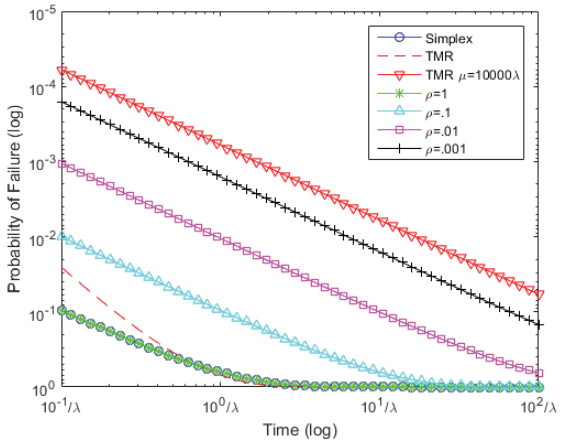

Figure 5 – Plots of TMR with CCF and a High Repair Rate

Without CCF, the reliability of a system with TMR and repair can be improved without bound by increasing the reliability rate. With CCF, improvement is limited. The limitation imposed is the inverse of the CCF rate.

$$\text{MTTF}_{\text{TMR with CCF}} = \frac{5+\rho+\frac{\mu}{\lambda}}{6\lambda+5\lambda\rho+\lambda\rho^2+\mu\rho} \quad (5)$$

$$\lim_{\mu\to\infty}\left(\text{MTTF}_{\text{TMR with CCF}}\right) = \frac{1}{\rho\lambda} = \frac{1}{\lambda_{CCF}} \quad (6)$$

$$\text{Improvement}_{\text{TMR with CCF}} = \frac{\text{MTTF}_{\text{TMR with CCF}}}{\text{MTTF}_{\text{Simplex}}} = \frac{1}{\rho} \quad (7)$$

At high repair rates, even low values of λ_{CCF} can have a significant impact on the system. This is not to say that the system digresses back into the Simplex system, but the difference in reliability between the TMR system with CCF and the ideal system grows. In Figure 5 where $\mu = 10{,}000\lambda$, all three of the TMR with CCF systems are significantly different from the ideal system. As the repair rate increases in order of magnitude, the CCF rate must be reduced by the same orders of magnitude in order to realize the full benefits of TMR with repair.

4 APPLICATION EXAMPLE

In a previous work [12], we set out to improve the TMR circuit reliability by reducing the CCF rate. We did this by implementing a mitigation technique called PCMF. We can use the results of the fault injection test reported in that paper to theoretically analyze the improvements, using the equations presented in this paper. The results of the fault injection test for the unmitigated, TMR and PCMF circuits are reported in Table 2. The PCMF circuit is the TMR circuit with an additional CCF mitigation technique applied.

From the table, the sensitivity for the unmitigated circuit would be the module failure rate, i.e., λ=1.34x10^{-2}. The sensitivities for the TMR and PCMF circuits would be the CCF failure rates, i.e., λ_{CCF}=1.83x10^{-5} and 1.25x10^{-6}, for the TMR and PCMF circuits, respectively. Using the values for λ and λ_{CCF} the values of ρ for each of the circuits can be calculated using Equation 6. This would result in ρ_{TMR}=1.36x10^{-3} and

$\rho_{PCMF}=9.32x10^{-5}$. Then using the values for ρ combined with Equation 7, the improvements can be calculated. This would result in $I_{TMR}=7.34x10^{2}$ and $I_{PCMF}=1.07x10^{4}$, which are the improvements that are reported (before rounding). This application shows that by reducing the CCF rate, or lowering ρ, the reliability of the TMR system can be greatly improved.

Table 2 – CCF Mitigation in a TMR circuit on an FPGA [12]

Circuit/Metric	Unmitigated	TMR	PCMF
Faults	2,193,073	2,351,568	2,396,265
Failures	29,436	43	3
Sensitivity	$1.34x10^{-2}$	$1.83x10^{-5}$	$1.25x10^{-6}$
Improvement	1x	730x	11,000x

5 CONCLUSION

This paper has proposed a new Markov chain to model systems that employ TMR with repair but are also susceptible to CCF. By using the new model we have shown that the reliability and MTTF of these systems is limited by the CCF rate. The system cannot improve past the CCF rate even when the repair rate of the system is set very high.

For future work we plan on continuing to explore the theoretical limits imposed by CCF. We plan on exploring the tradeoffs between increasing the repair rate and decreasing the CCF rate. We would also like to extend the model for other systems, such as systems with partial TMR and systems that employ partitioning.

ACKNOWLEDGEMENTS

This work was supported by the Utah NASA Space Grant Consortium, and by the I/UCRC Program of the National Science Foundation under Grant No. 1738550 as part of research conducted in the NSF Center for Space, High-Performance, and Resilient Computing (SHREC).

REFERENCES

1. D. McMurtrey *et al.*, "Estimating TMR reliability on FPGAs using Markov models," 2008, [Online] http://scholarsarchive.byu.edu/facpub/149/
2. S. McConnel and D. Siewiorek, "Evaluation criteria," *Reliable Computer Systems: Design and Evaluation,* 3rd ed., 1998, ch. 5, pp 334-336.
3. P. Hokstad and M. Rausand, "Common cause failure modeling: status and trends," *London: Springer London,* 2008, pp 621-640.
4. "Reactor safety: An assessment of accident risk in U.S. commercial nuclear power plants," *NUREG-75/014*, U.S. Nuclear Regulatory Commission, Washington D,. 1975.
5. S Hauge *et al.*, "Reliability prediction method for safety instrumented systems," *PDS method handbook*, Report STF50 A06031, 2006.
6. K. Fleming, "A reliability model for common mode failures in redundant safety systems," *Report GA-A13284*, General Atomic Company, 1975.
7. M. Shooman, "Reliability of computer systems and networks: fault tolerance, analysis and design," John Wiley & Sons, 2003, pp. 176-178.
8. K. Chae and G. Clark, "System reliability in the presence of common-cause failures," *IEEE Trans. On Reliability,* vol. 35, no. 1, (April) 1986, pp 32-35.
9. Z. Tang and J. Dugan, "An integrated method for incorporating common cause failures in system analysis," *Annual Symposium on Reliability and Maintainability*, (Jan.) 2004, pp 610-614.
10. H. Kim, "The design and evaluation of all voting triple modular redundancy system," *Annual Reliability and Maintainability Symposium,* 2002, pp. 439-444.
11. R. Katz *et al.*, "Radiation effects on current field programmable technologies," *IEEE Trans. On Nuclear Science*, vol. 44, no. 6, Dec. 1997, pp. 1945-1956.
12. M. Cannon *et al.*, "Strategies for removing common mode failures from TMR designs deployed on SRAM FPGAs," *IEEE Trans. On Nuclear Science,* vol. 66, no. 1, Jan. 2019, pp 207-215.
13. H. Quinn *et al.*, "Domain crossing errors: Limitations on single device triple-modular redundancy circuits in Xilinx FPGAs," *IEEE Trans. On Nuclear Science*, vol. 54, no. 6, (Dec.) 2007, pp 2037-2043.
14. M. Wirthlin *et al.*, "SEU mitigation and validation of the LEON3 soft processor using triple modular redundancy for space processing," *Proceedings of the 2016 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays,* 2016.

BIOGRAPHIES

Matthew J. Cannon
NSF SHREC
Department of Electrical and Computer Engineering
Brigham Young University
450 Engineering Building
Provo, UT 84602 USA

e-mail: mcannon@sandia.gov

Matthew Cannon earned his Ph.D. from Brigham Young University in the Electrical and Computer Engineering department where he also received his undergraduate degree. His research focused on the use of field programmable gate arrays (FPGA) in radioactive environments. His research started by exploring radiation effects on FPGAs and has recently focused on mitigation strategies to reduce the radiation susceptibility of FPGAs. He has developed several computer-aided design (CAD) algorithms to provide automated radiation mitigation for FPGA system. He has also explored mathematical models to theoretically analyze the impacts of his proposed mitigation strategies. Since graduation, Matthew works as an R&D engineer at Sandia National Labs.

Andrew M. Keller
NSF SHREC
Department of Electrical and Computer Engineering
Brigham Young University

450 Engineering Building
Provo, UT 84602 USA

e-mail: andrewmkeller@byu.edu

Andrew M. Keller is a PhD candidate in the Configurable Computing Lab in the Electrical and Computer Engineering Department of Brigham Young University. He has been with the lab since the last year of his Undergraduate studies, September 2014. His research focuses on the automated application of fault-tolerance techniques to improve the reliability of FPGA designs in radiation environments. Specifically, he is developing new techniques for the application of partial circuit replication. His interests include configurable computing, wireless communications, and embedded systems.

Andrés Pérez-Celis
NSF SHREC
Department of Electrical and Computer Engineering
Brigham Young University
450 Engineering Building
Provo, UT 84602 USA

e-mail: pcelis@byu.edu

Andrés Pérez-Celis is a Ph.D. student at Brigham Young University in the Electrical and Computer Engineering department. He received his master's degree in Computer Science from the National Institute of Astrophysics, Optics, and Electronics (INAOE). He has been working with FPGAs for over five years. His research focuses on the identification of anomalies that occur in FPGAs within radiation environments. He has developed several statistical techniques to extract multi-cell upsets from accelerated radiation testing. He has also developed tools to help in the analysis of accelerated radiation test data from FPGAs. His interests include reliability analysis and simulation, radiation effects on electronic devices, and reconfigurable computing.

Dr. Michael J. Wirthlin
NSF SHREC
Department of Electrical and Computer Engineering
Brigham Young University
450 Engineering Building
Provo, UT 84602 USA

e-mail: wirthlin@byu.edu

Dr. Mike Wirthlin is a Professor in the Department of Electrical and Computer Engineering at BYU in Provo, Utah and the BYU site director of the National Science Foundation Center for Space, High-Performance and Resilient Computing (SHREC). He has been actively involved in FPGA design and research for over 23 years. His research interests include reliable FPGA design, fault tolerant computing, and Configurable Computing Systems. He has led the development of a number of tools and techniques for improving the reliability of FPGA designs. He is a senior member of the IEEE and a member of the ACM.