Workload-Aware DRAM Error Prediction using Machine Learning


Published in:
2019 IEEE International Symposium on Workload Characterization: Proceedings

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
© 2019 IEEE
This work is made available online in accordance with the publisher’s policies. Please refer to any applicable terms of use of the publisher.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
Workload-Aware DRAM Error Prediction using Machine Learning

Lev Mukhanov*, Konstantinos Tovletoglou†, Hans Vandierendonck‡, Dimitrios S. Nikolopoulos§ and Georgios Karakonstantis¶
School Of Electronics, Electrical Engineering AND Computer Science, ECIT, Queen’s University Belfast
Email: *l.mukhanov@qub.ac.uk, †ktovletoglou01@qub.ac.uk, ‡h.vandierendonck@qub.ac.uk,
§d.nikolopoulos@qub.ac.uk ¶g.karakonstantis@qub.ac.uk

Abstract—The aggressive scaling of technology may have helped to meet the growing demand for higher memory capacity and density, but has also made DRAM cells more prone to errors. Such a reality triggered a lot of interest in modeling DRAM behavior for either predicting the errors in advance or for adjusting DRAM circuit parameters to achieve a better tradeoff between energy efficiency and reliability. Existing modeling efforts may have studied the impact of few operating parameters and temperature on DRAM reliability using custom FPGAs setups, however they neglected the combined effect of workload-specific features that can be systematically investigated only on a real system.

In this paper, we present the results of our study on workload-dependent DRAM error behavior within a real server considering various operating parameters, such as the refresh rate, voltage and temperature. We show that the rate of single- and multi-bit errors may vary across workloads by 8×, indicating that program inherent features can affect DRAM reliability significantly. Based on this observation, we extract 249 features, such as the memory access rate, the rate of cache misses, the memory reuse time and the data entropy, from various compute-intensive, caching and analytics benchmarks. We apply several supervised learning methods to construct the DRAM error behavior model for 72 server-grade DRAM chips by using the memory operating parameters and extracted program inherent features. Our results show that, with an appropriate choice of program features and supervised learning method, the rate of single- and multi-bit errors can be predicted for a specific DRAM module with an average error of less than 10.5 %, as opposed to the 2.9x estimation error obtained for a conventional workload-unaware error model. Our model enables designers to predict DRAM errors in advance for less than a second and study the impact of any workload and applied software optimizations on DRAM reliability.

I. INTRODUCTION

The worsening of parametric variations in deep nanometer technologies and aggressive scaling of circuit parameters for low power operation made memory cells more prone to errors, the number of which may vary significantly across different memory chips [38]. The manifestation of such errors, that depends on various factors [38] related to circuit parameters, temperature, as well as system architecture, and workloads, threaten the availability of computing systems and quality of service of sensitive storage components in data centers [62] and supercomputers [6], [21], [66]. The increased risks have triggered few research studies on prediction of DRAM errors in advance [5], [19], [33], [37], [53], [61], [77]. However, these studies were performed only for DRAM operating under nominal circuit parameters and typical environmental conditions. Moreover, even though they tried to consider other workload/architecture related factors, this was limited due to the constrained access to only specific features, like percentage of utilized memory, average CPU utilization and characteristics of the hardware [43]. The joint consideration of more features may reveal new non-linear behaviors that cannot be captured by linear regression models [43] or traditional workload-agnostic statistical models [29]. In addition, all these studies lacked an adequate number of samples because of the rare manifestation of errors for DRAM operating under nominal circuit parameters, which may result in contradictory observations [43], [62].

In the past, there have been several experimental studies that tried to predict the error behavior of DRAM operating under non-nominal circuit parameters [38], such as the refresh period ($T_{REFP}$) and the supply voltage ($V_{DD}$), and even under various temperatures [20], [25], [38]. However, the main goal of these studies was to improve DRAM performance and energy efficiency by scaling $T_{REFP}$ or $V_{DD}$, rather than model DRAM errors. Although, some of these works have indicated the fact that the certain program features, such as the pattern of data stored in memory [1], [4], [18], [39], [56], [58], [69], [73], [75], [75], [76], may change the number of manifested errors, none of them attempted to jointly consider the impact of DRAM circuit parameters and various program inherent features on DRAM reliability. Beside the data pattern, program inherent features encapsulate features that can be extracted using hardware program counters, e.g. the processor utilization, the rate of memory and cache misses, IPC. The program counters may have been used in the past for power and performance modelling [8], but were never used for DRAM reliability modeling in conjunction with various circuit parameters and temperature. Modeling the joint impact of such a wide range of features requires a novel experimental framework implemented on a real system with a complete software stack. This framework, unlike the custom FPGA setups used in prior studies [23], [39], [58], should be capable of running real workloads under different DRAM temperatures and provide a mechanism to measure errors and hardware program counters.

The main goal of this work is to systematically investigate
the effect of various program inherent features on DRAM reliability and develop a DRAM error model that takes into consideration the combined effect of these features, as well as the reliability variation across chips, DRAM circuit parameters and temperature. This model enables designers to predict DRAM errors based on few workload-specific features for a given set of DRAM circuit parameters and temperature. Such a prediction does not require long-running DRAM characterization campaigns that may take hours or even days on complex experimental setups. The error behavioral model facilitates: i) evaluating how prone to errors are specific workloads; ii) evaluating the implicit impact of applied software optimizations (e.g. compiler, or thread level parallelism) on DRAM reliability; iii) predicting maintenance cycles, as aimed by recent works [21], [43]; iv) guiding the adjustment of the circuit DRAM parameters for saving energy [40], [58].

Our contributions can be summarized as follows:

- We develop a novel experimental framework for characterizing DRAMs under relaxed refresh period and lowered supply voltage within a state-of-the-art 64-bit ARM based server. In order to experiment under different DRAM temperatures, we implement a thermal testbed that allows us to fine tune the temperature of each DIMM on the server.
- We perform a characterization of 72 server-grade DRAM chips under scaled refresh period and lowered supply voltage running compute-intensive, caching, and analytics benchmarks. Our study shows that the rate of single- and multi-bit errors may vary across workloads and DRAM chips by 8× and 188×, respectively.
- To quantify the dynamically changing data and access patterns of a running program, we introduce new metrics, namely the DRAM reuse time and the data entropy. We extract these program features along with 247 features measured using hardware performance counters during the execution of each workload and correlate them with DRAM errors, identifying features that are more likely to affect DRAM reliability.
- We apply three different Machine Learning methods to train a workload-aware DRAM error model based on the extracted program features, DRAM circuit parameters and temperature. In particular, we investigate the accuracy of Support Vector Machines (SVM), K-nearest neighbors (KNN) and Random Decision Forests (RDF). We compare these models on 4 different DRAM devices considering various sets of program features used for training. Our study shows that the highest accuracy of DRAM error estimates is achieved by KNN, which enables us to predict DRAM error rates within 300 ms with an average error that does not exceed 10.2 %, as opposed to the 2.9× estimation error obtained for a conventional workload-unaware error model. We make the DRAM error behavioral model (KNN-based) publicly available, which will be periodically updated based on new characterization results [3].

Fig. 1: Interaction between workloads and DRAM; Internal structure of DRAM.

II. BACKGROUND

A. DRAM Basics

DRAM is an essential component in any modern computing system, used to realize the memory subsystem. Beside the data caches, the memory subsystem includes several channels (Memory Channel Units, MCUs) which are used to transfer data and commands between the processor and DRAM. Each channel is connected to a number of Dual In-line Memory Modules (DIMMs). A DIMM usually has two ranks that contain DRAM chips. Within each chip, DRAM cells are organized into banks, which are two-dimensional arrays that can be accessed in parallel based on rows and columns (see Figure 1 on the right). The basic storage element of a DIMM is a cell, consisting of a transistor and a capacitor. When a row of cells is accessed, the peripheral circuitry of a DIMM senses the data stored in this row via amplifiers and sends it to the processor.

B. DRAM Error Behavior: Main Operating Parameters

The main drawback of the DRAM technology is the limited retention time [38] of a cell’s charge. To avoid any error induced by the charge leakage, DRAM employs an Auto-Refresh mechanism that recharges the cells in the array periodically [38]. Conventionally, all DDR technologies adopt a refresh period, $T_{REFP}$, of 64 ms for refreshing each cell. Other critical parameter that affects DRAMs’ power and reliability behavior is the supply voltage, $V_{DD}$. Similar to $T_{REFP}$, $V_{DD}$ of DRAM chips is chosen conservatively by vendors to ensure that each chip operates correctly under a wide range of conditions. In addition to the above circuit parameters, one of the main environmental conditions that affect DRAM reliability is temperature ($T_{DRAM}$). In fact, it has been reported that the retention time of DRAM cells decreases exponentially with increasing temperature [20].

C. DRAM Error Behavior: Workload-Dependent Parameters

The use of DRAM depends on executed instructions that access the memory in a certain way. In particular, the data read and written by a program (data pattern) from/to memory and the order in which the program refers to this data (access pattern) vary across workloads. Note that access pattern also
encapsulates the rate of memory accesses and the average time between accesses to DRAM cells.

Previous studies have demonstrated that the data pattern of a running program may affect DRAM errors [25], [38]. Meanwhile, the frequency of read and write accesses (i.e. the memory access pattern) may reduce the number of manifested errors, since each read/write naturally refreshes DRAM [1], [69]. We demonstrate such accesses in Figure 1 where 3.load and n.load instructions from the t-th workload refresh data in the m4 DRAM line. By contrast, if a row is accessed many times, then some cells from neighbouring rows may leak charge due to the DRAM cell-to-cell interference [30]. This effect has been exploited widely for “row hammer” attacks [50], [72]. Specifically, data in the m3 and m5 DRAM rows (see Figure 1) may be compromised when the m4 row is accessed too often. Thus, by increasing the memory access frequency to the same row, we reduce the number of errors manifested in this particular row, while inducing errors in neighborhood rows due to the DRAM cell-to-cell interference. Accordingly, inherent program features that change memory data and access patterns of a running workload may have an important effect on DRAM reliability. However, to the best of our knowledge, none of the previous studies have systematically investigated the combined effect of data and access pattern on DRAM reliability under relaxed DRAM parameters and varying DIMM temperature.

Failing to identify the combined effect of program features on real server deployments may limit or nullify the efficacy of existing approaches. For example, several previous studies have proposed fine-grained methods to control DRAM parameters based on the retention time measured for each cell [39], [57]. To measure the retention time, authors use micro-benchmarks that implement the worst-case data pattern manifesting errors in the vast majority of error-prone memory locations [4], [20], [23], [25]. However, our study shows that real applications may trigger errors in many more memory locations than the conventional data pattern micro-benchmarks. Figure 2 depicts the rate of single-bit errors per 64-bit word (WER) observed for DRAM operating under relaxed parameters when running two different benchmarks (memcached and backprop), and the most stressful data pattern micro-benchmark (the random data pattern micro-benchmark [38]). We see that the WER incurred by backprop is 3.5× higher than the rate observed for random. As a result, the cell retention time measured using this data pattern micro-benchmark may be inaccurate, which, in turn, may lead to uncertain hardware behavior or even hardware crashes when the proposed methods are applied in practice. On the other hand, the proposed methods may be too pessimistic about the retention time and thus ineffective, since real applications, such as memcached, may trigger errors in fewer memory locations than the micro-benchmark. These results indicate that designers should take into account the combined effect of workload-dependent factors on DRAM reliability when designing error mitigation techniques.

D. DIMM-to-DIMM Variation

Apart from the above circuit and workload-dependent parameters, DRAM reliability may vary across DIMMs from different vendors [27], [38], and even across DIMMs manufactured by the same vendor. This variation is due the manufacturing process [29] and the internal design of DRAM modules, such as true-anti cell organization [38], address scrambling [27], [71] and the remapping of faulty cells [26]. Our study indicates that the rate of single-bit errors per 64-bit word may vary by 188× across different DRAM chips.

E. Challenges

According to the above discussion, there are various cross-layer parameters, at the circuit (e.g. \( V_{DD} \), \( T_{REFP} \)), micro-architecture (i.e. cache organization and DRAM architecture), application (i.e. data and DRAM access patterns) layers, which in combination with environmental parameters (i.e. the DRAM temperature) can significantly influence DRAM reliability. Predicting the potential failures early at design or operation cycle by considering all the combined cross-layer effects is an extremely challenging problem.

III. DRAM Error Prediction

A. Mathematical Formulation of the Problem

Let us assume that a workload, having a specific set of program features \( F_{trs} = (f_1, f_2, ..., f_K) \) where \( f_i \) is the i-th feature), allocates data on a DRAM device \( D_{ev} \) when this device operates under \( T_{REFP} \) and \( V_{DD} \) at a certain temperature \( T_{DRAM} \). Then, to predict a target DRAM error metric \( M_{err} \) for this workload, we need to model a prediction function \( M \) such that:

\[
M_{err} = M(F_{trs}, D_{ev}, T_{REFP}, V_{DD}, T_{DRAM}) \tag{1}
\]

It is evident that building such a model is extremely challenging due to the number of possible parameter combinations. To address this challenge, we propose to use a supervised Machine Learning (ML) technique, since we believe it is hard to find an analytical model that predicts DRAM error behavior accurately considering the DIMM-to-DIMM variation and all the parameters.
TABLE I: Types of DRAM errors that can be corrected or detected with ECC SECDED.

<table>
<thead>
<tr>
<th>Num. of corrupted bits</th>
<th>Type of errors</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>corrected</td>
<td>CE</td>
</tr>
<tr>
<td>&gt; 1</td>
<td>uncorrected/detected</td>
<td>UE</td>
</tr>
<tr>
<td>&gt; 2</td>
<td>uncorrected/undetected</td>
<td>SDC</td>
</tr>
</tbody>
</table>

B. ML Models

In our study, we investigate the accuracy of the following Machine Learning models: Support Vector Machines (SVM), K-nearest neighbors algorithm (KNN) and Random Decision Forests (RDF), as it was previously shown that the models have a high accuracy for both linear and non-linear prediction problems [16]. This is essential for our study, as the type of correlation between some parameters and DRAM error metrics can be linear and non-linear for others parameters. For example, it was previously shown that the DRAM bit-error rate is exponentially proportional to \( T_{REFP} \) and the DRAM temperature [25]. Finally, the training data are expensive to obtain in our study, since each program run requires at least 2 hours to collect DRAM errors. It was previously shown that the chosen models have a high accuracy when trained using small data [16], [34].

C. DRAM Error Metrics

There are several types of errors that may manifest in DRAM chips [2], [52], [59], [65]. Vendors implement a special hardware (ECC, Error Correction Codes) in server-grade chips to automatically correct such errors. In this study, we use hardware that supports ECC SECDED, which is implemented in the majority of commercial servers. There are three types of memory errors that may occur when ECC SECDED is enabled (see Table I): single-bit errors (or correctable errors, CE); detected errors where more than one bit in a 64-bit word is corrupted (or uncorrectable errors, UE); and errors where more than 2 bits are corrupted per word, which are not corrected and not detected by ECC. The last types of errors manifest so-called Silent Data Corruption (SDC), since such errors are invisible for hardware.

Correctable errors: To characterize DRAM in terms of CEs, we measure the rate of single-bit errors per 64-bit, \( WER \), for the amount of memory used by an application as:

\[
WER = \frac{N_{CE}}{MEMSIZE} \tag{2}
\]

where \( N_{CE} \) is the number of unique 64-bit word locations where CEs have manifested and \( MEMSIZE \) is the size (in 64-bit words) of memory allocated by the application. \( WER \) shows the probability of a word being erroneous regardless of the size of memory allocated by the application.

Uncorrectable Errors: To characterize DRAM in terms of UEs, we estimate the probability of an UE, triggered by a running application as:

\[
P_{UE} = \frac{N_{UE}}{N_{EXP}} \tag{3}
\]

where \( N_{UE} \) is the number of experiments with the application that resulted in an UE, and \( N_{EXP} \) is the total number of experiments with the application.

D. Program Inherent Features

To investigate software-level factors that may affect DRAM reliability, we extract the following program features.

The DRAM Reuse Time: The DRAM reuse time \( (T_{reuse}) \) is the average time between memory accesses to the same 64-bit word (or a DRAM location). This metric is important for our study, as memory accesses inherently refresh the stored charge [1], [69], while \( T_{REFP} \) denotes the average period between accesses to the DRAM cells, and thus, the average refresh period of cells incurred by memory accesses. If \( T_{reuse} < T_{REFP} \) for a running program, then the number of DRAM errors induced by the charge leakage will decrease. We estimate \( T_{reuse} \) by averaging the DRAM reuse time over all memory accesses, i.e., \( T_{reuse} = \frac{\sum_{i=1}^{N_{mem}} T_{i}}{N_{mem}} \), where \( T_{i} \) is the reuse time for the \( i \) memory access instruction with reference to some address. In turn, we calculate \( T_{reuse} \) as:

\[
T_{i}^{reuse} = CPI \times D_{i}^{reuse} \tag{4}
\]

In this equation, \( CPI \) is the average number of clock cycles per instruction measured for an entire program, \( D_{i}^{reuse} \) is the number of instructions executed since the last reference to the address accessed by the \( i \) instruction. We extract \( D_{i}^{reuse} \) using a dynamic binary instrumentation tool, DynamoRIO [9]. We validated \( T_{reuse} \) estimates using micro-benchmarks where we can control and measure \( T_{reuse} \) for specific memory accesses, and found that the approximation is accurate.

The Data Entropy: To quantify the varying data patterns (DPs) stored in memory across different time instances, we introduce a new metric, the DP entropy, \( H_{DP} \). To estimate \( H_{DP} \), we profile all workloads with DynamoRIO and take samples of the data for each write memory access that is ultimately stored in DRAM. We then estimate \( H_{DP} \) based on the sampled data as:

\[
H_{DP} = -\sum_{i=0}^{2^{32}-1} P(x_i) \times \log_2(P(x_i)); P(x_i) = \frac{N_{WR}(x_i)}{N_{TOT}} \tag{5}
\]

where \( N_{WR}(x_i) \) is the number of writes operations with data \( x_i \) in a word and \( N_{WR} \) is the total number of writes.

Performance Counters: Another important parameter that may affect DRAM reliability is the number of memory accesses executed per cycle, as the cell-to-cell interference grows with the rate of memory accesses [46], [59]. We measure this number, along with 247 program metrics, such as L1/L2/memory accesses (writes and reads) per cycle, and IPC and the SoC utilization, using existing hardware performance counters (perf) to investigate the potential effect of other architecture-level parameters on DRAM error behavior.

E. Data Collection

To collect data for training of the ML models, we run a set of representative benchmarks (workloads) under varying
DRAM operating parameters, such as $T_{REFP}$, $V_{DD}$ and temperature, and measure $WER$ and $P_{UE}$, as shown in Figure 3. We additionally run each benchmark to collect all the inherent program features using DynamoRio and the perf tool (Profiling phase). Then, we combine collected program inherent features with the $WER$ or $P_{UE}$ measurements.

F. Accuracy Evaluation of ML Models

We evaluate accuracy of the ML models using the cross-validation technique [31] by partitioning the collected data into a test set and a training set. We use the Leave-One-Out [49] partitioning as shown in Figure 3. In particular, for each benchmark we create a test set that consists of samples taken only for a specific benchmark, whereas the training set contains all other samples. We train the model (Training phase) and test its prediction accuracy (Testing phase) for each pair of training and testing sets (see Figure 3). Finally, we average the prediction accuracy over all testing experiments, the number of which is equivalent to the total number of benchmarks.

IV. EXPERIMENTAL SETUP

To enable DRAM characterization, we developed a unique experimental setup which we discuss in this section.

A. Experimental Framework

The basis of our experimental framework is a state-of-the-art commodity 64-bit ARMv8-based server, the X-Gene2 Server-on-a-Chip. The X-Gene2 SoC consists of eight 64-bit ARMv8 cores running at 2.4GHz. The X-Gene2 has four DDR3 Memory Controller Units (MCUs). In our campaign, we are experimenting with 4 Micron DDR3 8GB DIMMs at 1866 MHz [44], with one DIMM per MCU. In total, we are characterizing 72 chips of 4Gb x8 DDR3 [45], since each DIMM includes 16 and 2 DRAM chips for data storage and ECC, respectively.

DRAM Thermal Testbed on a Server. To perform the experiments under controlled temperatures, we implement a temperature-controlled testbed using heating elements [23] for DRAMs on a server. Figure 5 shows the X-Gene2 board with four DIMMs fitted with our custom adapters. Each adapter consists of a resistive element, with thermally conductive tape transferring the heat of the element to all the chips in a DIMM in a uniform way, and a thermocouple to measure the temperature. The temperature of each element is controlled by a controller board, as shown in Figure 6, which contains a Raspberry Pi 3 [17] and four closed-loop PID controllers [10].

B. DRAM Parameters and Error Accounting

The X-Gene2 provides access to a separate, scalable, and light-weight intelligent processor (SLIMpro), which is a special management core that is used to boot the system and provide access to the on-board sensors to measure the temperature and the power of the SoC and DRAM. The SLIMpro also reports all memory errors corrected or detected by SECDED ECC to the Linux kernel, providing information about the DIMM, bank, rank, row and column in which the error occurred. Finally, SLIMpro allows the configuration of the parameters of the MCUs, such as $T_{REFP}$ and $V_{DD}$. Specifically, $T_{REFP}$ may be changed from the nominal 64 $ms$ to 2.283 $s$, which is the maximum on the X-Gene2 server. The server runs a fully-fledged OS based on CentOS 7 with the default Linux kernel 4.3.0 for ARMv8 and support for 64KB pages.

C. Benchmarks

In our study, we use Rodinia and Parsec benchmark suites, specifically the backprop, nw, srad, kmeans and fmm benchmarks, which represent a variety of compute-intensive algorithms [7], [12]. To evaluate how parallelism and processing power affect the characterization, we run these benchmarks with 1 and 8 threads. To investigate the effect of popular caching and analytics workloads on DRAM reliability, we run the memcached benchmark [55], the pagerank algorithm...
(pagerank), the betweenness centrality algorithm (bc) and the breadth-first search algorithm (bfs) [63]. Finally, we run each benchmark allocating 8 GB of data to exclude the effect of the data size factor on DRAM errors.

V. DRAM CHARACTERIZATION AND DATA COLLECTION

In this section, we characterize DRAM error behavior when running real workloads under lowered $V_{DD}$, different levels of $T_{REFP}$ and the selected DRAM temperature range.

Temperature. We characterize DRAM at three temperature levels: $50^\circ C$, $60^\circ C$ and $70^\circ C$. We use this temperature range to follow previous studies [38] and the DIMM specification [44], in which the vendor reports the maximum operating temperature of $70^\circ C$. Note that this temperature range is common for dense server environments [39], [47], [48].

DRAM Circuit Parameters. We experimentally determine the lowest operating DRAM $V_{DD}$ as 1.428 V, after which the circuitry of the DRAM is likely to stop working. We execute all the benchmarks with the memory operating under the minimum $V_{DD}$ (1.428 V) discovered in our experiments; however, the benchmarks have not manifested errors for DRAM operating at $50^\circ C$. Moreover, we discover only a few CEs by running benchmarks at $60^\circ C$ and $70^\circ C$. Thus, reducing $V_{DD}$ from the nominal 1.5 V down to 1.428 V (or by 5%) has a negligible effect on DRAM reliability.

The maximum power gain is achieved when both $T_{REFP}$ and $V_{DD}$ are scaled. To achieve this gain, in the rest of this paper, we set the minimum $V_{DD}$ (1.428 V) and run all the benchmarks under different $T_{REFP}$.

A. Correctable errors

In our experiments with all the benchmarks for DRAM operating under scaled $T_{REFP}$ and $V_{DD}$, we encounter only CEs at $50^\circ C$ and $60^\circ C$, and no UEs or SDCs.

Previously, it was discovered that the memory cell leakage may change over time due to a phenomenon called variable retention time ($VRT$) [60]. As a result, DRAM error behavior may vary across runs of the same application, and thus, it is essential to run each application several times until a target DRAM error metric converges to a specific value. To this end, we run each application for 2 hours with DRAM operating under the maximum $T_{REFP}$ (i.e. 2.283 s) and lowered $V_{DD}$ (1.428 V) at $50^\circ C$. Figure 4 shows how the rate ($WER$) of single-bit errors detected in 64-bit words changes over time for each benchmark. Note that labels with abbreviation (par) correspond to the parallel version of the computer-intensive benchmarks. We see that after 2-hour runs $WER$ achieves a certain value for each benchmark: the average change in the $WER$ for the last 10 minutes of each experiment does not exceed 3 % at $50^\circ C$. We observe the same results for DRAM operating at $60^\circ C$. These observations imply that 120 minutes is sufficient for identifying the vast majority of error-prone memory locations and characterize DRAM behavior when running a specific benchmark.

WER: Further, we investigate how $WER$ varies across benchmarks when DRAM operates under different $T_{REFP}$ at $50^\circ C$ and $60^\circ C$. We run benchmarks for DRAM operating under $0.618\ s$, $1.173\ s$, $1.727\ s$, $2.283\ s$ $T_{REFP}$ and lowered $V_{DD}$. Figure 7 illustrates how $WER$ changes with scaling $T_{REFP}$ at $50^\circ C$ and $60^\circ C$. Our first observation is that $WER$ varies across benchmarks significantly; for example, the difference achieves almost $8\times$ for memcached and backprop(par) when DRAM operates under $0.618\ s$ at $70^\circ C$. Our second observation is that the benchmark that incurs the highest $WER$ may change with the DRAM temperature and $T_{REFP}$; for example, the highest $WER$ is obtained for $srad$ when DRAM operates under $1.173\ s$ at $50^\circ C$, while for DRAM operating under the same $T_{REFP}$ at $70^\circ C$ the highest $WER$ is observed for the $bfs$ benchmark. This indicates that DRAM operational and environmental parameters may change the ratio of $WER$ between different workloads which is hard to capture with an analytical model. Our third observation is that $WER$ grows exponentially with $T_{REFP}$ (see Figure 7f).

Finally, we see that the $WER$ incurred by the parallel version of some benchmarks differs from the $WER$ obtained for the single-threaded version of these benchmarks. For example, the $WER$ measured for backprop is almost 30 % greater than the $WER$ obtained for backprop(par) when DRAM operates under $2.283\ s$ $T_{REFP}$ at $50^\circ C$ and $60^\circ C$. The same difference is also observed in the case of the $srad$ benchmark. Importantly, parallel and single-threaded versions of the same workload have different memory access scenarios, but a similar data pattern. Thus, these observations imply that the memory access pattern of a running program may also significantly affect DRAM error behavior.

To investigate the difference in $WER$ for parallel and single-threaded benchmarks, we calculate $T_{reuse}$ for each workload, as shown in Table II. We see that the $T_{reuse}$ of the parallel backprop and $srad$ is less than the $T_{reuse}$ estimated for the single-threaded version of backprop and $srad$, respectively. As follows, in the case of backprop and $srad$, the parallel benchmarks implicitly refresh data in the memory more frequently than the single-threaded benchmarks do by generating more accesses to the same regions of memory per cycle. As a result, we observe a low error rate for these parallel benchmarks. Nonetheless, in the case of kmeans, the parallel version has a higher $T_{reuse}$ (0.50 s) than the serial version (0.17 s) due to a better data locality in caches obtained for the parallel kmeans. Respectively, the parallel version generates fewer references to the same memory per cycle than does the single-threaded version, resulting in a higher $T_{reuse}$ and therefore a higher $WER$. Lastly, memcached incurs the lowest $WER$ and has the lowest $T_{reuse}$ for DRAM operating under different $T_{REFP}$ and temperatures among all workloads at the same time, which confirms that there is a correlation between

<table>
<thead>
<tr>
<th>Workload</th>
<th>$T_{reuse}$ (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>memcached</td>
<td>0.09</td>
</tr>
<tr>
<td>pagerank</td>
<td>0.48</td>
</tr>
<tr>
<td>bfs</td>
<td>0.61</td>
</tr>
<tr>
<td>bc</td>
<td>0.56</td>
</tr>
</tbody>
</table>

TABLE II: The average DRAM reuse time.
Fig. 7: \textit{WER} for DRAM operating under 0.618 s, 1.173 s, 1.727 s, 2.283 s at 50°C (a,b), 60°C (c,d) and 70°C (e). The \textit{WER} averaged over all benchmarks for DRAM operating at 50°C and 60°C (f).

Fig. 8: \textit{WER} per DIMM/rank obtained for DRAM operating under 2.283 s \textit{T_{REFP}} at 50°C.

Fig. 9: a) \textit{P_{UE}} and b) The probability to obtain an UE on a specific DIMM/rank when DRAM operates under 1.450 s, 1.727 s and 2.283 s \textit{T_{REFP}} at 70°C.

\( T_{REFP} \): and DRAM error behavior.

To investigate how \textit{WER} varies across different DIMMs and ranks, we grouped all the collected errors by a source DIMM/rank. Figure 8 shows \textit{WER} measured on different DIMMs and ranks when DRAM operates under 2.283 s \textit{T_{REFP}} at 50°C. We see that \textit{WER} varies across DIMMs/ranks by up to 188x; in particular, \textit{WER} incurred by the \textit{bc} benchmark on DIMM2/rank0 and DIMM3/rank1 are \( 1.75 \times 10^{-7} \) and \( 9.31 \times 10^{-10} \), respectively. Therefore, to enable accurate DRAM error predictions, a model should take into consideration the error behavior of a specific DIMM. These results motivate the use of Machine Learning models for DRAM error prediction, as such models can be trained specifically for a certain device.

\textbf{B. Uncorrectable Errors and System Crashes}

In our experiments with DRAM operating at 50°C and 60°C, we have discovered no Silent Data Corruptions (SDCs) or uncorrectable errors (UEs). However, we encounter UEs and system crashes when raising the DRAM temperature to 70°C and scaling \textit{T_{REFP}} up to 1.45 s under lowered \textit{V_{DD}}. Note that in our framework, any UE triggered by the Linux kernel or a user-level program, once detected by ECC, will result in a system crash.

Figure 9a shows \textit{P_{UE}}, the likelihood to observe an UE, measured across all benchmarks for DRAM operating under 1.450 s, 1.727 s, 2.283 s \textit{T_{REFP}} and lowered \textit{V_{DD}} at 70°C. To estimate this probability, we repeat each 2-hour experiment with a specific benchmark 10 times. We see that \textit{P_{UE}} varies significantly across benchmarks for DRAM operating under 1.450 s \textit{T_{REFP}}; it achieves 0.8 for \textit{fmm(par)}, whereas it equals to 0 for \textit{memcached} and \textit{pagerank}. We also observe that \textit{P_{UE}} is greater than 0 only for the parallel compute-intensive benchmarks, while it is 0 for all the single-threaded benchmarks except for \textit{srad}. The \textit{P_{UE}} averaged over all benchmarks
for DRAM operating under 1.450 s $T_{REFP}$ is below 0.4. However, when we increase $T_{REFP}$ up to 1.727 s, then the likelihood of crashing averaged over benchmarks grows by 2.15× (see Figure 9). Moreover, for DRAM operating under this $T_{REFP}$, there is no benchmark with $P_{UE} = 0$. Finally, all the benchmarks trigger UEs in 100% of the experiments when we use the maximum $T_{REFP}$ (2.283 s) at 70°C. These results shows that $T_{REFP}$ and the DRAM temperature have a dominant effect on the likelihood of an UE.

Figure 9b depicts the probability to obtain an UE on a specific DIMM/rank when ECC detects an UE. We see that the vast majority of UEs are triggered by DIMM0/rank1 and DIMM2/rank0, while DIMM3/rank1 do not trigger UEs at all. Thus, DRAM reliability varies significantly from DIMM-to-DIMM not only in terms of WER but also the probability to obtain an UE.

Importantly, we have discovered no SDCs when running experiments under different $T_{REFP}$ at 50°C, 60°C, and 70°C.

VI. ACCURACY EVALUATION OF ML MODELS

In this section, we present the results of the feature selection process and accuracy evaluation of ML models.

A. Feature selection

The accuracy of an ML model strongly depends on the set of features used for training of the model. If the model is trained using the set of features that are not correlated with a metric that we target to predict, then the model may overestimate the significance of some features [13]. As a result, a low prediction accuracy will be obtained for this model. To identify those features that may affect DRAM reliability, we extract 249 program features, including $T_{reuse}$ (the average memory reuse time) and $H_{DP}$ (the data entropy, see Section III), for each benchmark, and correlate them with both WER and $P_{UE}$ metrics.

WER: We build the correlation of WER and program features using the combined measurements taken under different levels of $T_{REFP}$ (0.618 s, 1.173 s, 1.727 s, 2.283 s) at 50°C, 60°C and 70°C, where we observe no UEs or system crashes.

To identify and quantify any dependency between program features and the DRAM error rates, we use the Spearman’s rank correlation coefficient ($r_s$). This correlation coefficient allows us to detect both linear and non-linear relationships [51]. Coefficient values lie in a range $[-1, +1]$ in which -1 or +1 occurs when there is a perfect monotonic relationship between two variables.

Figure 10 shows the correlation coefficients for 249 program features and WER on the Y-axis, whereas the correlation coefficients for these features and $P_{UE}$ are shown on the X-axis. We see that the number of memory accesses per cycle is highly correlated with WER, as $r_s$ is above 0.57, indicating a positive direction of the correlation; in other words, WER grows with the number of accesses per cycle. We also observe that the group of performance indicators that reflects the number of issued memory read and write commands per cycle in different MCUs is also highly correlated with WER.

However, the number of such commands is determined by the number of memory read and write instructions executed by the processor per cycle.

Another inherent program feature that is strongly correlated with WER is wait cycles ($r_s = 0.4$). This feature reflects the ratio of the number of cycles spent on waiting for data to the total number of program cycles. Nonetheless, wait cycles is implicitly determined by the number of memory accesses per clock cycle, as it encapsulates idle cycles due to memory access stalls which explains its correlation with WER.

We attribute the correlation of the memory access rate and WER to disturbance errors induced by the cell-to-cell interference [46], [59]. Previously, it was shown that, if a row is accessed many times, then some cells from neighbouring rows may leak charge quickly [30]. Thus, by accessing the memory with a high rate, we increase the probability of the interference errors for DRAM operating under scaled $T_{REFP}$ and $V_{DD}$. By contrast, under a higher memory access rate, each cell may be implicitly refreshed more frequently, resulting in a lower WER. However, this effect occurs only for those benchmarks in which $T_{reuse} < T_{REFP}$. Therefore, a high memory access rate may have negative or positive effects on DRAM reliability, which depends on $T_{reuse}$ and $T_{REFP}$.

Notably, $T_{reuse}$ is greater than the maximum $T_{REFP}$ (2.283 s) available on our platform for almost 30% of the benchmarks. Thereby, $T_{reuse}$ does not have any effect on DRAM error behavior in these benchmarks. This lack of an effect explains why $T_{reuse}$ ($r_s = 0.23$) is less correlated with WER than the rate of memory accesses.

Our experiments show that $H_{DP}$, which reflect the data pattern of a running application, is also correlated with WER as the $r_s$ is 0.39, see Figure 10. Although it is higher than the $r_s$ obtained for $T_{reuse}$, it is by 51% lower than the $r_s$ observed for the memory access rate.

The probability of an UE: Similar to WER, we discover a correlation between $P_{UE}$ and the memory access rate, the number of issued memory read and write commands per cycle in different MCUs, $H_{DP}$, and wait cycles. However, the level of this correlation is lower than in the case of WER; for example, the $r_s$ for the memory access rate and $P_{UE}$ is 0.43, which is 35% less than the same $r_s$ for WER. It is noteworthy that unlike previous studies, which have indicated a strong impact of $T_{reuse}$ or $H_{DP}$ [25], [68], we obtain the highest $r_s$ for the memory access rate among all the program features when correlating it with WER and $P_{UE}$ metrics.
the maximum error of $WER$ estimated per application is about 55% when we use the third input set for the RDF model, see Figure 11f (the fmm benchmark). Meanwhile, the average error of $WER$ estimates provided by SVM and KNN per application do not exceed 30% and 24%, correspondingly, when we use the first input set. Thus, we may conclude that RDF has the lowest accuracy among the considered models when predicting $WER$.

The probability of an UE: Figure 12 depicts the mean percentage error of $P_{UE}$ estimates averaged over all benchmarks and DIMMs. Similar to our experiments with $WER$, we see that the first set incurs the lowest error (12.3%) when we use SVM. While the average error obtained by this model for the second and third sets is above 15%. However, KNN and RDF demonstrate the lowest average error when we use the second input feature set. Notably, this error is only 4.1% and 5.5% for KNN and RDF, respectively, which is almost 3x lower than the lowest error (12.3%) achieved by SVM.

To conclude, our study shows that the highest accuracy of $WER$ estimates is achieved by the K-nearest neighbors algorithm (KNN) when we train it using the first input set of parameters (i.e., the memory access rate, $wait$ cycles, $H_{DP}$ and $T_{reuse}$, $TEMP_{DRAM}$, $T_{REFP}$ and $V_{DD}$). The highest accuracy of $P_{UE}$ estimates is also demonstrated by KNN when we use the second input set, which contains only the memory access rate, $wait$ cycles, $TEMP_{DRAM}$ and $T_{REFP}$.

### C. Workload-Aware Modeling vs Conventional Modeling

Many studies have proposed to model DRAM errors for investigation either hardware design efficiency [40] or software fault tolerance [35], [36], [41], [42]. However, all those studies use constant DRAM error rates extracted on real DRAMS when running the data pattern micro-benchmarks [4], [20], [23], [39], [57]. Our model can be used to improve those studies and proposed techniques by considering workload-aware DRAM error behavior. For example, Figure 13 depicts the measured $WER$ over all DIMMs when DRAM is operating under 0.618 $s$ $T_{REFP}$ at 70°C for the lulesh benchmark and a data pattern micro-benchmark that implements a random data pattern [25]. This figure also shows the $WER$ which has been predicted by the KNN-based DRAM error behavioral model. In this experiments, we use two versions of lulesh to illustrate the implicit effect of compiler optimizations on DRAM reliability: the benchmark compiled with $-O2$ (default optimizations) and $-F$ (aggressive optimizations). We see that the model correctly predicts the $WER$ incurred by both versions of the benchmark; the error is less than 3%. Such a high accuracy enables us to correctly predict the difference in $WER$ between these benchmarks, which is about 29%. At the same time, we see that the random micro-benchmark incurs the $WER$ which is higher than the $WER$ obtained for lulesh by 2.9×. Thus, the conventional DRAM error modeling based on the constant rates may be inaccurate and lead to incorrect conclusions about the effectiveness of hardware or software techniques.
Moreover, the vast majority of research studies have considered only hardware-level techniques to mitigate errors for DRAM operating under scaled [4], [20], [23], [57], which introduce additional power and chip area overheads. However, as we see, even compiler optimizations may implicitly affect DRAM error behavior. To systematically study the effect of compiler optimizations, it is essential to build a model, since such a study may take months or even years if it is conducted using DRAM characterization campaigns. While our models predict DRAM errors within 300 ms, which opens new avenues for research.

VII. RELATED WORK

Scaling of $T_{REFP}$ and $V_{DD}$: Many studies [24], [28], [39], [54], [57], [73] tried to improve DRAM performance and energy efficiency by adopting a low refresh period for "weak" cells. The main idea of such an approach is to split memory cells into groups based on their retention time and relax the refresh rate for those groups where cells have small leakage. Other works [1], [15], [18] suggested to skip refresh operations for those memory segments that have been implicitly refreshed by memory accesses. Several studies [4], [22] proposed to extend this technique and refresh selectively only rows with valid data allocated by running applications or OS. Chang et al. [11] provided the results of their study on reduced-voltage operation in DDR3L memory devices. However, even though the latest study [27] tried to capture the effect of varying data patterns on DRAM reliability when running real applications, all these studies ignored the combined effect of data and memory access patterns on DRAM errors. To the best of our knowledge, none of previous works have systematically investigated the combined impact of these patterns on memory errors on a real server. Understanding of such an impact is crucial for facilitating the co-design of software and hardware techniques to improve DRAM energy efficiency. At the same time, modeling this impact enables us to predict the relaxed DRAM circuit parameters which do not trigger errors for a specific workload, and thus, reduce the DRAM power without compromising reliability. In our future research, we target to investigate governors based on the developed models that change dynamically the DRAM circuit parameters. Other research studies proposed various fine-grained schemes to reduce the number of refresh operations and thus improve DRAM energy efficiency [18], [22], [32], [67], [70], [74]. Although some of these studies utilize workload inherent features, such as the memory reuse time, they are orthogonal to our work.

Predictive maintenance and statistical prediction of errors: Considerable research has been done on statistical prediction of different types of hardware faults, including DRAM errors, in supercomputers [5], [19], [33], [37], [43], [53], [61], [77]. The majority of these studies proposed different techniques, based either on rules [37] or Machine Learning [61], for prediction of failures that may happen in various hardware components using history of errors. Other research studies tried to systematically investigate factors, including workload-dependent factors, that may affect DRAMs in data centers and supercomputers [43], [62], [64], [66]. Nonetheless, all these studies tried to investigate and predict errors for hardware operating under nominal parameters.

Hardware error prediction becomes extremely important in production lines for identifying maintenance cycles or faulty components (predictive maintenance) [19], [62]. However, any study of failures for hardware operating under nominal parameters may require years [19], while a reliability characterization of hardware that operates under relaxed parameters is much faster. In our future research, we aim to investigate...
how characterization and modeling of errors for DRAM operating under relaxed parameters can be applied to identify maintenance cycles or any abnormal hardware behavior.

VIII. CONCLUSION

In this work, we present the results of a study on characterization and prediction of the error behavior for DRAM operating under scaled parameters within a real server. Our results indicate that the rate of single- and multi-bit errors may vary across workloads and DRAM chips by 8× and 188×, respectively. We quantify the effect of inherent program features that may significantly affect DRAM errors by correlating 249 features extracted from various benchmarks with DRAM errors. We train three ML models to predict DRAM failure rates and compare the accuracy of the models using different sets of program features. We demonstrate that, with the correct choice of program features and an ML model, the word-error-rate for single-bit failures and the likelihood of a system crash triggered by uncorrectable errors can be predicted for a specific DRAM device with an average error of less than 10.5%. Finally, we release the developed DRAM error behavior model online.

REFERENCES


