Computers and Electrical Engineering 46 (2015) 157–175

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

The hardware accelerator debate: A financial risk case study using many-core computing ☆

Blesson Varghese

School of Computer Science, University of St Andrews, UK

ARTICLE INFO

Article history:
Received 15 October 2014
Received in revised form 15 January 2015
Accepted 15 January 2015
Available online 7 February 2015

Keywords:
Hardware accelerators
Many-core computing
Secondary uncertainty
Financial risk
Catastrophic risk
Risk analysis

ABSTRACT

The risk of reinsurance portfolios covering globally occurring natural catastrophes, such as earthquakes and hurricanes, is quantified by employing simulations. These simulations are computationally intensive and require large amounts of data to be processed. The use of many-core hardware accelerators, such as the Intel Xeon Phi and the NVIDIA Graphics Processing Unit (GPU), are desirable for achieving high-performance risk analytics. In this paper, we set out to investigate how accelerators can be employed in risk analytics, focusing on developing parallel algorithms for Aggregate Risk Analysis, a simulation which computes the Probable Maximum Loss of a portfolio taking both primary and secondary uncertainties into account. The key result is that both hardware accelerators are useful in different contexts; without taking data transfer times into account the Phi had lowest execution times when used independently and the GPU along with a host in a hybrid platform yielded best performance.

1. Introduction

Risk analytics [1] has become an integral part of a business process in a range of domains including structural inspections [2], nuclear power plants [3] and radioactive waste disposal [4]. Large datasets are consumed by a risk model and hundreds of thousands or even millions of time consuming simulations are performed. Here the application of parallel and high-performance computing techniques are attractive.

Interestingly, in the financial risk domain, specifically insurance and reinsurance, where data sizes are as large or even larger than what is employed in the above domains, relatively fewer parallel and high-performance computing techniques have been applied. Given the dependencies of the insurance and reinsurance setting on volatile markets, simulations that can be performed in a timely manner are essential.

1.1. Background

Companies hold portfolios of contracts that cover risks associated with catastrophic events such as earthquakes, hurricanes and floods. In order to have a marketplace for such risk it is critical to be able to efficiently quantify individual risks and portfolios of risks. The analytical pipeline of the modern quantitative insurance or reinsurance company typically consists of three major stages, namely risk assessment [5], portfolio risk management and pricing [6], and enterprise risk management [7].

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. Jesus Carretero.

E-mail address: varghese@st-andrews.ac.uk
URL: http://www.blessonv.com

http://dx.doi.org/10.1016/j.compeleceng.2015.01.012
0045-7906/

In the first stage, catastrophe models [8] are used to provide loss estimates by taking two inputs. Firstly, stochastic event catalogues which are a mathematical representation of the natural occurrence patterns and characteristics of catastrophes. Secondly, exposure databases that describe thousands or millions of buildings to be analysed, their construction types, location, value, use, and coverage. Each event-exposure pair is then analysed by a risk model that quantifies the hazard intensity at the exposure site, the vulnerability of the building and resulting damage level, and the expected loss, given the customer's financial terms. The output of a catastrophe model is an Event Loss Table (ELT) which specifies the probability of occurrence and the expected loss for every event in the catalogue.

In the second stage of the analysis pipeline, portfolio risk management and pricing of portfolios of contracts necessitates a further level of stochastic simulation, called Aggregate Risk Analysis or referred to as ARA in this paper [9,10]. ARA is a Monte Carlo like simulation in which each trial represents an alternative view of which catastrophic events occur and in which order they occur within a predetermined period or a contractual year. In order to provide actuaries and decision makers with a consistent lens through which to view results, a pre-simulated Year Event Table (YET) containing a million alternative views of a single contractual year is employed. The output of ARA is a Year Loss Table, in which the results are highly aggregated.

From the output of ARA, a reinsurer can derive important portfolio risk metrics such as the Probable Maximum Loss (PML) [11] and the Tail Value-at-Risk (TVaR) [12]. The output is then interpreted by actuaries for key internal decision making, planning a financial year and reporting to regulators and rating agencies. Furthermore, these metrics then flow into the final stage in the risk analysis pipeline, namely Enterprise Risk Management, where liability, asset, and other forms of risks are combined and correlated to generate an enterprise wide view of risk.

1.2. Problems

There are two problems that need to be addressed for achieving high-performance risk analytics. Both problems can be solved if the data, memory and computational challenges of the analysis can be efficiently addressed. The first problem is related to developing methods for applying uncertainties. ARA presented above accounts for only 'Primary Uncertainty', which is the uncertainty associated with whether an event occurs or not in a simulated year. However, there is 'Secondary Uncertainty', which captures the uncertainty in the level of loss due to the use of simplified physical models and limitations in the available data.

There are many sources of this uncertainty that need to be taken into account when considering catastrophic risk, including unknown exposure and hazard parameters and their interaction. For example, the exposure data which describes the buildings, their locations, and construction types may be incomplete, lacking in sufficient detail, or may be simply inaccurate. Also the physical modelling of hazard, for example an earthquake, may naturally generate a distribution of hazard intensity values due to uncertainty associated with the energy attenuation functions used or driving data such as soil type. Lastly, building vulnerability functions are simplifications of complex physical phenomenon and are therefore much better at producing loss distributions than accurate point estimates. Hence, there is a need to develop methods to not only capture primary uncertainty but also quantify secondary uncertainty in risk analysis.

The analysis uses mean loss values when only primary uncertainty is accounted for. Using such discrete values is an oversimplification, because in reality for any event there is a multitude of possible loss outcomes resulting in a distribution of potential losses. A simulation taking a distribution of losses requires statistical tools, for example, the beta probability distribution to estimate the loss using inverse beta cumulative density function which are both data intensive and computationally intensive. Such an analysis will need to accept as input complete event loss distributions represented by the event rate, mean loss, independent standard deviation, and correlated standard deviation, and better account for the range of possible outcomes.

The second problem is related to implementing parallel risk analysis methods for achieving timely results. From a computational perspective the ARA simulation differs from other Monte Carlo simulations since trials are pre-simulated, rather than randomly generated on-the-fly. This provides millions of alternate views of a contractual year comprising thousands of events which are pre-simulated as a YET. From an analytical perspective, a pre-simulated YET lends itself to statistical validation and to tuning for seasonality and cluster effects. However, there are significant challenges in achieving efficient parallelisation. The extremely large YET must be carefully shared between processing cores if the computation is to achieve good speed-up when there is limited memory bandwidth.

1.2.1. Addressing the problems

In this paper, we investigate hardware acceleration platforms for ARA. Parallel algorithms for ARA that initially take primary uncertainty into account is implemented. Further, a methodology that considers secondary uncertainty is presented. The algorithms are implemented on the Intel Phi Coprocessor and an NVIDIA Graphics Processing Unit (GPU). Experimental studies evaluate how ARA performs on the hardware accelerators independently and along with a host processor.

Both the Phi Coprocessor and the GPU are competing hardware accelerators that offer alternative machine architectures to that of a regular CPU. While both hardware accelerators are significantly different, they provide in common, firstly, lots of cycles for independent parallelism, secondly, fast memory access under the right circumstances, and finally, fast mathematical computations. The parallel algorithms implemented in the paper take full advantage of the high levels of parallelism, fast shared memory access and fast numerical performance. In this research, the algorithm exploits the machine architecture of