Adaptive Multispectral GPU Accelerated Architecture for Earth Observation Satellites

R. L. Davidson, C. P. Bridges
Surrey Space Centre
University of Surrey
Guildford, Surrey, GU2 7XH, UK
rebecca.davidson@surrey.ac.uk, c.p.bridges@surrey.ac.uk

Abstract—In recent years the growth in quantity, diversity and capability of Earth Observation (EO) satellites has enabled increase's in the achievable payload data dimensionality and volume. However, the lack of equivalent advancement in downlink technology has resulted in the development of an onboard data bottleneck. This bottleneck must be alleviated in order for EO satellites to continue to efficiently provide high quality and increasing quantities of payload data.

This research explores the selection and implementation of state-of-the-art multidimensional image compression algorithms and proposes a new onboard data processing architecture, to help alleviate the bottleneck and increase the data throughput of the platform. The proposed new system is based upon a backplane architecture to provide scalability with different satellite platform sizes and varying mission’s objectives. The heterogeneous nature of the architecture allows benefits of both Field Programmable Gate Array (FPGA) and Graphical Processing Unit (GPU) hardware to be leveraged for maximised data processing throughput.

Keywords—Onboard, Image, Compression, Processing, Earth Observation, GPU, FPGA

I. INTRODUCTION
Earth Observation (EO) satellites are often operated in a store and forward mode, whereby payload data is stored onboard until it can be scheduled for transmission to a suitable ground station. High resolution optical images, collected from panchromatic and multispectral sensors, are among the most common payload data products to be downlinked. Advancements in instrument and sensor technologies have resulted in an increase in the achievable spatial, spectral, temporal and radiometric resolutions. Consequently, exponentially increasing data volumes are creating new data processing challenges for the EO and satellite industry. This issue is compounded onboard due to the fact that comparable advances in data downlink technologies have not occurred in recent years. The performance of current downlink systems is inhibited by limitations on antenna size, pointing abilities, power and the restricted availability and increasing demand for transmission frequencies. As a result, many current and future missions will experience a growing onboard data bottleneck. This must be alleviated in order for EO satellites to deliver the quality and quantity of payload data expected by reliant applications.

To help relieve this bottleneck, onboard image compression can be performed. Image compression is achieved by exploiting correlations in the data to achieve a more efficient representation of the image information. The first known satellite to implement onboard image compression was SPOT-1, launched in 1986 [1]. Since then, many satellites have featured onboard compression capabilities [2]. Historically very simple compression was performed in software on the main onboard computer (OBC). However, as algorithm sophistication and data sizes have increased dedicated hardware solutions are now often used. Hardware solutions have to conform to a number of satellite environment-induced constraints such as minimised size, mass, power consumption and tolerance to radiation effects. In recent years, the field programmable gate array (FPGA) has emerged as the standard choice for onboard image processing [3]. FPGA’s have gained wide spread adoption due to their relatively small size, mass, power consumption, radiation hardness and re-configurability. However, radiation-hardened, space qualified devices often lag several generations behind their terrestrial counterparts with regards to computational resources. As the data volume and bottleneck are now increasing exponentially, more sophisticated compression and processing algorithms will need to be implemented onboard. Thus onboard processing resources are becoming a growing priority. A recent ESA study concluded that the “challenging requirements for future onboard payload data processing systems cannot be met with space qualified processors available today” [4].

The performance of any onboard data processing system will be dependent upon the selection of hardware and processing algorithms and also the effectiveness of the algorithm implementation to harness the full capabilities of the hardware architecture. This paper will focus on the image compression algorithm selection and proposal of a hardware architecture system for onboard data processing. Efficient hardware specific algorithm implementation will be briefly introduced, but is the focus of ongoing research.

II. LOSSLESS IMAGE COMPRESSION LITERATURE REVIEW
A vast number of image compression algorithms have been proposed in literature to date. The theoretical basis, computational complexity and achieved compression ratios of these algorithms can vary greatly. Compression can be performed in either a lossless or lossy manner, however as high
data fidelity is of critical importance to many EO data applications. Lossless algorithms are greatly preferred. Lossless image compression algorithms can be categorised based on the types of redundancy they exploit. The classification used in this study is detailed in Figure 1, where traditional algorithms are those that utilise spatial and statistical redundancy reduction techniques, whilst multidimensional algorithms additionally exploit spectral redundancies.

An extensive study covering both traditional and multidimensional lossless image compression algorithms has been conducted. Compression ratio performance data was gathered from a wide range of published literature and the results from this survey can be found in Figure 2. These results show that there has been negligible improvement in the compression ratio performance achieved by traditional algorithms proposed in recent years. Multidimensional algorithms on the other hand show an increasing trend in compression ratio performance and also increasing research into the development of new theoretical compression techniques. To highlight the compression advantages of multidimensional algorithms as a whole it has been calculated that the average compression ratios of all multidimensional algorithms surveyed is 3.04, this is 52% higher than the 2.0 average of all traditional algorithms.

Current onboard EO data processing systems often utilise traditional predictive image compression algorithms such as JPEG-LS achieving typical compression ratios between 2.0 – 2.5. The ever increasing spectral dimensionality of EO imagery provides a great opportunity to select multidimensional algorithms for onboard implementation to increase image compression performance. Therefore, a review of key multidimensional algorithms identified from this study will be given in the following paragraphs. It is essential algorithm selection takes into account the resource requirements of algorithms in addition to the achievable compression ratio. This is important as onboard computational resources and memory are often limited due to design constraints induced by the space environment.

Several of the first multidimensional algorithms proposed in literature were predictive based adaptations of existing traditional algorithms. In 1999, a multidimensional version of JPEG-LS was proposed, D-JPEG-LS [5]. A comparison between these algorithms shows a considerable increase in average compression performance, approximately 35% for the multidimensional version. D-JPEG-LS achieves this by employing simple differential band encoding after the standardised predictive JPEG-LS encoding.

Several novel techniques such as Look Up Table (LUT), Distributed Source Coding (DSC) and Vector Quantisation (VQ) based algorithms have also been proposed in literature for multispectral compression. LUT algorithms are similar in principle to predictive schemes [6]. However, the prediction equation is replaced by a search, conducted in the preceding image band, for the nearest located pixel that is equal in value to the current pixel to be encoded. Once found the pixel of the same spatial location, but in the current band, is used as the estimated value. To reduce computational complexity, the search logic is replaced by a simple lookup table in which the co-located pixel value is used as an index and the LUT returns
the nearest matching pixel value in the current band. Two extended versions have also been proposed. The first includes a preliminary locally averaged inter-band scaling (LAIS) calculation to perform outlier rejection, and the second uses quantised LUT values to reduce the size of the necessary LUT, improving the memory requirements of the technique [7][8]. LUT based algorithms have been shown to achieve 15-20% reduced compression ratio on uncalibrated data when compared to calibrated data performance. As calibration processes are performed on the ground, as post-processing, the LUT compression performance for raw uncalibrated onboard data is expected to be significantly less than alternative predictive based schemes [9].

Several DSC based multidimensional algorithms have been proposed in recent years [10][11][12]. DSC is an information coding concept in which independent encoders can be used to encode multiple correlated sources at a reduced entropy. By utilising independent encoders, it is possible to encode a first source at its entropy level and perform conditional encoding of a second source at a rate lower than its entropy. The algorithms were designed specifically to evaluate the potential of applying DSC principles in an effective compression algorithm. The average compression results show that substantial further research will be required before DSC image compression is able to achieve competitive results against more mature predictive coding techniques. The DSC-CALIC algorithm is able to achieve a compression ratio 19% greater than the original CALIC algorithm but 22% less than the early multidimensional predictive version IB-CALIC [12].

In addition to the exploration of new techniques, several novel predictive based multidimensional algorithms have also been proposed in literature. The most significant is the 2005 Fast Lossless (FL) algorithm, designed specifically to provide an appropriate trade-off between compression complexity and performance for onboard utilisation [13]. In the FL algorithm both spectral and spatial redundancy are exploited using a small 3D neighbourhood of previously encoded pixels from a user-defined number of preceding spectral bands. The key novelty of this scheme is the use of the sign algorithm. The sign algorithm is a low complexity variation of the least mean square algorithm and is used in FL to produce optimised predictor weightings. The final predictor is a combination of a weighted sum, determined using both current and previous band information, and a local mean subtraction used to improve processing throughput. This algorithm has proved extremely popular in the community, it has recently been adopted by the Consultative Committee for Space Data Systems (CCSDS) as the basis of the CCSDS-123 image compression standardisation title “Lossless Multispectral and Hyperspectral Data Compression” [14]. The FL and CCSDS-123 algorithms achieve an impressive average compression ratio of 3.24. Of all the algorithms surveyed, the CCSDS-123 standardised algorithm is the most suitable algorithm for onboard implementation. It is able to achieve a compression ratio approximately 54% greater than JPEG-LS, providing a great trade-off between minimising the required computational resources and maximising compression performance for raw data.

As well as low complexity specific algorithms, several high computational complexity algorithms that feature advanced pre-processing have been proposed [15][16][17][18]. These algorithms are able to achieve state-of-the-art levels of compression due to the incorporation of image segmentation to determine areas of homogenous pixels for increased prediction accuracy. Clustered Differential Pulse Code Modulation (C-DPCM) is an example of one such algorithm whereby error optimised linear predictors are calculated for each cluster utilising collocated pixels from previously encoded bands [17]. C-DPCM-APL (Adaptive Predictor Length), is a variation of this algorithm that uses a brute force approach to determine the optimum number of previously encoded bands to use in the linear predictor calculation [18]. C-DPCM-APL is able to achieve an average compression ratio of 3.47, the highest compression ratio of all the algorithms surveyed. However, implementing classification and segmentation as part of the compression algorithm results in a substantial increase in computational complexity, resource requirements and processing time. Whilst these algorithms are currently seen as unsuitable for onboard implementation they do show that increased knowledge and statistics for an image can be used to further increase the achievable compression ratio.

III. ONBOARD DATA PROCESSING ARCHITECTURE

A. Current Onboard Processing Systems

Current onboard payload data processing systems are commonly based upon space qualified FPGA hardware and feature cross-strapped hard redundant systems to maximise the protection against space radiation effects. They will likely be comparable in structure to the block diagram given in Figure 3.

![Payload Data Handling and Storage System Diagram](image)

Fig. 3. A current typical onboard data processing system design.

However, the computational performance of the radiation hardened processor’s used, lag several generations behind their terrestrial counterparts [19]. With ever increasing image resolutions, these devices will unlikely be able to provide the computational resources necessary to process multiple spectral band data concurrently. The data from each spectral band will therefore be processed independently in its own hardware - meaning multispectral compression will not be able to be utilised. The scaling of this traditional system, with increasing
data volumes, will likely see exponential increases to the area, mass and power consumption of the system.

As demand for onboard processing capabilities will only continue to increase in priority; ESA findings suggest that it is becoming even more essential than ensuring low electrical power consumption [20]. Therefore, this work proposes a new onboard data processing architecture which provides increased computational resources to facilitate the implementation of state-of-the-art multidimensional image compression to help alleviate the onboard data bottleneck.

### B. Proposed New Onboard Data Processing System

The proposed new heterogeneous onboard data handling architecture, shown in Figure 4, has been designed to exploit the high performance computing capabilities that can be leveraged from the latest GPU hardware. Whilst GPU’s and FPGA’s are often compared against each other as hardware accelerators and are rarely utilised in the same system, interest in the area of GPU-FPGA heterogeneous computing and its potential application is a growing field of research [21].

![Fig. 4. The proposed new onboard data processing system design.](image)

The proposed new hardware architecture is based around a physical backplane, which connects different hardware devices chosen to handle specific stages in the onboard data processing chain. The use of a backplane will also provide a level of scalability to the system. Thereby the number of processing and or memory devices can be adapted in order to suit a specific satellite platform and mission.

1) **FPGA - Data interfacing and formatting**

In this new architecture the FPGA has been chosen to provide the main data interface to the payload and perform data manipulation tasks that would be costly in GPU hardware. Utilising an FPGA allows the flexible logic cells to provide a fully customisable data path for efficient implementation of a streaming data interface and additional functionality such as image tiling and pixel reordering. Image tiling helps increase image processing throughput by exposing an additional dimension of parallelism. Another advantage of image tiling is increased data error resilience, as error propagation is limited to a single independently processed tile. Pixel reordering will help ensure aligned memory operations occur on the GPU for more efficient hardware utilisation.

2) **GPU – Image processing**

GPU’s provide large computational resources, in the region of several thousand GFLOPS (Giga Floating Point Operations Per Second) for single precision arithmetic, and high memory bandwidth, of around a couple hundred GB/S. These characteristics in addition to the highly parallel computing architecture make the GPU well suited for the implementation of image processing tasks, and thus a good candidate for the computing platform of a new onboard processing system. The GPU is centred around a single instruction multiple data (SIMD) architecture, utilising high levels of multithreading to achieve high data throughput [22]. A current trend in terrestrial computing devices, such as GPUs, is toward lower power, mass and device area. The requirements of terrestrial technologies and space applications are becoming increasingly aligned, creating greater opportunities to employ novel commercial off the shelf (COTS) based solutions in space systems.

In the proposed onboard data processing system the GPU will be utilised to enable the implementation of parallelised state-of-the-art multidimensional image compression and advanced data processing algorithms. To demonstrate the advantages of leveraging the latest GPU technology for onboard image processing, a hardware optimised software processing chain, including an implementation of the CCSDS-123 algorithm, will be developed.

Currently, GPU’s are not manufactured to provide radiation tolerance at the silicon gate level. This poses an obstacle in the practical deployment of a GPU device in an onboard system as it conflicts with the traditional ethos within the space industry, to strictly utilise only radiation hardened processors. Therefore Radiation Hardening by Software Design (RHBSW) principles will be adapted and implemented in the proposed system to ensure radiation effects are mitigated [19][23][24]. Some RHBSW techniques which have been proposed in literature to date include: heartbeat monitoring, watchdog timers, control flow assertions, progress monitoring and checkpoint and rollback functionality. The highly parallel nature of the GPU could also allow for the adaption of traditional space modular redundancy techniques, such as triple modular redundancy (TMR), for GPU software.

### IV. GPU Optimised Onboard Image Processing Software

Details of several different CCSDS-123 implementations have been published in recent years [25][26][27][28]. Two of these papers detail findings, from work contracted by the JPL NASA, on the efficient GPU implementation of CCSDS-123 for airborne hyperspectral imagers [27][28]. Although the CCSDS-123 algorithm was originally designed for efficient FPGA hardware implementation, it was shown that an effective parallel GPU implementation of the algorithm can be devised. The initial implementation employed a greedy optimisation approach in which buffering was used to remove serial data dependencies and increase algorithmic parallelism. The second implementation restricted the parallelism of the algorithm to be the same as the least parallel stage. This allowed the large memory buffers to be removed and data to be passed between stages using low latency registers. The cost of
synchronisation was overcome by the benefits of greater data reuse and leveraging the high performance caches. Freed resources, from the greedy optimisation, also allowed for parallel processing of several image tiles simultaneously, further increasing data throughput.

There are several key differences between multispectral and hyperspectral data and onboard processing for airborne and spaceborne platforms. As previous implementations address both hyperspectral and airborne platform options this new research will focus on the development of a GPU implementation optimised multispectral image compression for onboard a spaceborne platform. This implementation will utilise the latest technology to provide new features not covered under previous research:

- The number of spectral bands in a multispectral data set are significantly less, typically in the order of 10 bands, than hyperspectral images, which can contain over 200 bands. The spectral the axis is the greatest source of parallelism for the CCSDS-123 algorithm therefore with less images bands there will be a reduced level of inherent parallelism.

- Previous GPU implementations have conformed to a flat parallel programming model, in which parallelism is determined at each sequential kernel launch. The latest GPU hardware are now able to leverage dynamic parallelism where individual threads are now equipped with the capability to launched new threads dynamically during a kernel execution [29]. This provides an opportunity to exploit new hardware to replace some of the parallelism lost for multispectral data sets.

- The correlation between spectral bands will likely be reduced for multispectral imagery, when compared to hyperspectral data, due to the increased spectral range per band, potentially reducing the achievable compression ratio. This has been proved experimentally and the results are shown in Table 1. It can be seen that typically there are increased compression ratios for hyperspectral images over multispectral images.

<table>
<thead>
<tr>
<th>Image Type</th>
<th>Image Name</th>
<th>Image Dimensions (x,y,z)</th>
<th>Compression Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multispectral</td>
<td>Landsat Agriculture</td>
<td>1024x1024x6</td>
<td>4.50</td>
</tr>
<tr>
<td>Multispectral</td>
<td>Landsat Coast</td>
<td>1024x1024x6</td>
<td>5.73</td>
</tr>
<tr>
<td>Multispectral</td>
<td>Landsat Mountain</td>
<td>1024x1024x6</td>
<td>4.20</td>
</tr>
<tr>
<td>Hyperspectral</td>
<td>AVIRIS Hawaii</td>
<td>614x512x224</td>
<td>6.20</td>
</tr>
<tr>
<td>Hyperspectral</td>
<td>AVIRIS Maine</td>
<td>680x512x224</td>
<td>5.74</td>
</tr>
</tbody>
</table>

- In order to ensure the onboard compression ratios is maximised, excess computational resource provided by GPU hardware will be utilised to perform lossless pre-processing functions such as band and image calibrations and band reordering. Through preliminary testing it has been shown that the order in which the bands are used for multispectral compression has an impact on the achievable compression ratio. The results from this testing are shown in Figure 5.

![Fig. 5. Compression ratio and band ordering results using the Emporda CCSDS-123 algorithm implementation.](image-url)

- In order to be successfully deployed in space the onboard data processing system should be able to tolerate a certain level of radiation upsets.

- Dynamic parallelism and the highly threaded nature of the GPU will be utilised to incorporate RHBSW principles to provide a GPU algorithm implementation suitable for use onboard spacecraft.

V. CONCLUSION

This paper has discussed three areas key to the design of any new data processing system: the selection of an appropriate processing algorithm, the choice of appropriate system architecture and hardware computing platform and finally the proposed implementation of the algorithm to leverage the advantages of the chosen hardware.

Firstly, the findings from an in depth literature review of lossless image compression algorithms was discussed. On average multidimensional algorithms were able to achieve 52% higher compression ratio than traditional algorithms. Due to the higher achievable compression performance coupled with the natural trend in EO imagery towards greater dimensionality, it is thoroughly recommended that the EO satellite community move toward employing multidimensional algorithms. In particular, the CCSDS-123 algorithm was identified in this study as the current state-of-the-art algorithm most suited for onboard implementation. CCSDS-123 is able to achieve an average compression ratio 54% higher than JPEG-LS.

Due to the demands imposed by the ever increasing raw data volumes new hardware architectures, designed to provide increased computational and memory resources, will be
necessary. The solution proposed in this work is a novel backplane hardware architecture, designed specifically to provide scalable provisions for state-of-the-art image compression and processing functionality. Data interfacing, buffering and formatting will be performed on a separate FPGA whilst computationally intense processing functions will be implemented on independent GPU hardware. The FPGA will provide a highly optimised data path for interfacing to the payload, which provide flexibility and configurability to the mission and platform. The GPU's will provide a high performance and throughput platform for intensive processing functionality. This architecture follows an increasing trend in which satellite platforms begin to mimic design principles from terrestrial computing platforms.

The new GPU implementation will need to be carefully designed to exploit new dynamic parallelism functionality to increase processing throughput and implement RHBSW design principles to make the architecture suitable for onboard implementation, this is currently part of ongoing work.

REFERENCES
[30] CCSDS MHDC - Space Link Services Image Compression Testing Corpus http://cwe.csed.ucl.ac.uk/docs/Forms/AllItems.aspx?RootFolder=%2Fsls%2FFoldes%2FSLSL%2DDDC%2F123%2DE0%2DDB%2DInfo%2FTestData &View=%7B16ACDA38%2DFFA3%2D4657%2D8F27%2D166C23C24A%2D7D& (June 2016)