A high performance approach for parallel computing of fibre Bragg grating strain profiles using graphics processing units

L.H. Negri, H.S. Lopes, M. Muller and J.L. Fabris*

Graduate Program in Electrical and Computer Engineering, Federal University of Technology – Paraná, Brazil
Email: lucashnegri@gmail.com
Email: hslopes@utfpr.edu.br
Email: mmuller@utfpr.edu.br
Email: fabris@utfpr.edu.br
*Corresponding author

A.S. Paterno
Department of Electrical Engineering, Santa Catarina State University, Brazil
Email: aleksander.paterno@udesc.br

Abstract: This work proposes an efficient approach to recover the mechanical strain profile applied on fibre Bragg grating sensors. The proposed method is based on differential evolution and uses only the sensor reflectivity, without requiring phase information. The method has been shown to be highly parallelisable, with the fitness evaluation procedure implemented on graphical processing units. Experiments were performed to evaluate the performance of the method on three distinct graphics processing units (GPU), under a series of increasing loads. An enhancement up to three orders of magnitude in performance was obtained in respect to other evolutionary method proposed in the literature for the same purpose. Furthermore, it was observed that, for smaller problem sizes, the GPU clock rate was more significant than the number of cores of the GPU.

Keywords: graphical processing unit; GPU; differential evolution; performance evaluation; sensors; parallel computing; computational intelligence; fibre Bragg grating; FBG; optical fibre; sensing.


Biographical notes: L.H. Negri is currently working towards his PhD in Electrical Engineering and Industrial Informatics at the Federal University of Technology – Paraná. His interests include machine learning, computational intelligence, and fibre optic sensors.

H.S. Lopes is a Titular Professor at the Federal University of Technology Paraná, in Curitiba, Brazil. He holds a PhD and MSc in Electrical and Biomedical Engineering. His current research interests are bioinformatics, evolutionary computation, deep learning and high-performance computing.

M. Muller received her BSc in Physics from Federal University of Paraná in 1985, MSc from the Fluminense Federal University in 1989 and PhD from the University of São Paulo in 1994. She is currently a Full Professor at Federal University of Technology in Brazil. Her current main research area is photonics, with special interest in optical fibre grating-based sensors.

J.L. Fabris received his BSc in Physics from the Federal University of Paraná in 1986, MSc from the Fluminense Federal University in 1989 and PhD from the University of São Paulo in 1994. He is currently a Full Professor at Federal University of Technology in Brazil. His current main research area is photonics, with special interest in optical fibre grating-based sensors and spectroscopy.
1 Introduction

Fibre Bragg gratings (FBGs) are devices produced within the core of optical fibres. An FBG reflects light at certain wavelength bands, according to its structure. Changes in the grating structure result in the modification of its reflection spectrum, allowing the device to be inherently sensitive to temperature variations and mechanical deformation (Hill et al., 1978; Kersey et al., 1993). Also, FBGs are immune to external electromagnetic fields, they have intrinsic wavelength multiplexing capabilities and are convenient for integration in optical links over long distances. These characteristics make FBGs suitable for a wide range of sensing applications.

FBG transducers can be interrogated by illuminating them with a broadband light source and detecting the spectral position of the reflected band, which is related to the measurand. This principle of operation of an FBG sensor is depicted in Figure 1. Changes in the measurand are then determined by traditional methods (Kersey et al., 1997) used to detect spectral shifts in the central wavelength of the reflected band. These methods are efficient when the external parameters affects uniformly the whole device. Nevertheless, such methods cannot be efficiently employed when the FBG sensor is subject to non-uniform perturbations along its length.

Figure 1 Principle of operation of an FBG sensor: a broadband light source illuminates the FBG, that reflects light at a given wavelength band

Note: The reflection and transmission spectra are complementary.

A method based on differential evolution (DE) (Storn, 1996; Price et al., 2005) was proposed by Negri et al. (2014) to circumvent these issues, enabling the recovery of the mechanical deformations profiles applied to the FBG. This method requires only information about the magnitude of the band reflected by the FBG sensor, with no need for phase information. This method showed a reduced computational time when compared to another evolutionary computation solution (Cheng and Lo, 2004), but it is still not enough for high demand applications such as real time monitoring.

The present paper proposes a new implementation of the DE method to take advantage of the parallel processing capabilities of modern graphical processing units (GPU). The performance of the resulting central processing unit (CPU) and GPU hybrid method was assessed for a series of problem instances with increasing loads under three different GPUs.

2 Related works

Methods to determine the periodicity of FBGs from their reflection signal were already proposed in the literature. Such methods are useful mainly for distributed strain sensing and for FBG synthesis. Each method have its own advantages and limitations.

The method proposed by Leblanc et al. (1996) allows measuring the distributed strain on an FBG by means of its reflectance. Nevertheless, the method works only for monotonic strain profiles. A subsequent work by Ohn et al. (1997) introduced an interferometric method to determine arbitrary strain profiles. In this method, the strain profile is determined from the complex reflection spectrum of the FBG. This requirement prevents the application of the method when only the FBG reflectance is available. On the following years, other methods that also required complex spectra were proposed (Muriel et al., 1998; Azaña et al., 2001; Skaar and Feced, 2002).

Later, Cheng and Lo (2004) proposed a method to determine the strain profile on an FBG by using its reflectance. The method employed a genetic algorithm (GA) to evolve a population of strain profiles, selecting the profile that resulted in a reflectance as close as possible to the measured reflectance. However, the method required a computational effort of hours and also required the use of two FBGs that must be subject to the same strain profile simultaneously, thus limiting its applicability.

The method proposed by Negri et al. (2014) showed a strategy similar to the GA method (Cheng and Lo, 2004), replacing the GA by a DE method and introducing a
smoothness constraint. The computational time was reduced from hours to tens of seconds, and the drawback of using two FBGs was dropped. However, the performance of the method was still not enough for real-time applications. In a subsequent work (Negri et al., 2015), the usage of GPU was proposed to reduce the computational time. Initial results showed a reduction of the computational time from tens to tenths of seconds.

3 Methodology

The process of determining the mechanical deformation profile for a given FBG sensor starts with the acquisition of the FBG reflectance spectrum, that can be routinely obtained with interrogation methods (Kersey et al., 1997). The spectrum obtained is, henceforth, referenced as target spectrum, and contains the information used by the DE method to determine the deformation profile.

The DE method proposed maintains a population of candidate solutions, where each individual corresponds to a deformation profile. By using DE, this population is improved until a suitable deformation profile is found. A deformation profile is deemed suitable when it results in a light reflectance spectrum that matches the target spectrum, with the reflectance spectrum being simulated by using the transfer matrix method (Yamada and Sakuda, 1987). This evolutionary process is guided by evaluating the fitness of each individual, composed of the difference between the obtained and target spectra plus a term to avoid convergence to undesired solutions.

Simulating the reflectance spectrum for the whole population requires more time than the other steps of the evolutionary method due to the sequence of operations performed in the transfer matrix method. To compute the reflectance of an FBG at a given wavelength, the transfer matrix method requires a series of trigonometric operations and matrix multiplications with complex numbers (floating point).

To reduce the computation time required by the method, GPUs with capacity of running multiple parallel threads were used. By taking advantage of the massive parallelisation capabilities of modern GPUs, the fitness evaluation can be divided into multiple threads, each one computing the reflectance of a candidate profile for a given wavelength, thus achieving a fine granularity.

3.1 Transfer matrix method

As shown by Huang et al. (1995), the transfer matrix method can be used to compute the reflectance of a FBG in a simple way. A non-uniform FBG can be analysed as a sequence of uniform sections. Each uniform section can be modelled as a $2 \times 2$ matrix (transfer matrix) that provides the relationship between the incident and reflected power of that section. The resulting transfer matrices are then sequentially multiplied to obtain the response of the whole FBG.

The transfer matrix for a FBG section of length $l$, for a specific wavelength $\lambda$, is given by equation (1), where $b$ corresponds to the forward signal ($b(0)$ is the input, and $b(l)$ the output) and $a$ corresponds to the backward signal ($a(0)$ is the input and $a(l)$ the output). Here, we are interested in the reflectance $R = |a(0)/b(0)|^2$.

$$
\begin{bmatrix}
  a(0) \\
  b(0)
\end{bmatrix} =
\begin{bmatrix}
  T_{11} & T_{12} \\
  T_{21} & T_{22}
\end{bmatrix}
\begin{bmatrix}
  a(l) \\
  b(l)
\end{bmatrix}.
$$

(1)

The transfer matrix coefficients $T_{11}$ and $T_{22}$ are given by equation (2), and the coefficients $T_{12}$ and $T_{21}$ are given by equation (3):

$$
T_{11} = T_{22}^* = \frac{\Delta\beta \sinh(sl) + is \cosh(sl)}{is} \exp(-i\beta_0 l),
$$

(2)

$$
T_{12} = T_{21}^* = \frac{k \sinh(sl)}{is} \exp(i\beta_0 l),
$$

(3)

where $k = \pi \Delta n_0 / \lambda$, $s = (|k|^2 - \Delta \beta^2)^{1/2}$ and $\Delta \beta = 2\pi n / \lambda = \pi / \lambda$ (Huang et al., 1995).

The mean value of the core refractive index is given by $\pi$. This refractive index is assumed to be sinusoidal modulated, with amplitude $\Delta n_0$ and modulation period (pitch) $\Lambda$.

When a FBG is subject to a dilatational deformation gradient, its structure changes. These modifications are associated with the photoelastic effect (Huang et al., 1995) and the longitudinal dilation. The non-uniform FBGs analysed here are decomposed into 20 uniform sections, where each section is subject to an uniform strain $\epsilon$. The new pitch $\Lambda'$ is given by equation (4) and the new mean refractive index value $\pi'$ is given by equation (5):

$$
\Lambda' = \Lambda + \Lambda \epsilon,
$$

(4)

$$
\pi' = \pi - 0.5\pi^2 (p_{12} - \nu (p_{11} + p_{12})) \epsilon,
$$

(5)

where $p_{11}$ and $p_{12}$ are the fibre photoelastic coefficients and $\nu$ is the Poisson coefficient.

An FBG can present a variation of the pitch along its axis (chirp). The pitch is modelled by equation (6), that computes the new pitch $\Lambda'$ as function of the position $z$ and pitch factor $\delta \Lambda$:

$$
\Lambda'(z) = \Lambda_0 + (\delta \Lambda / \Lambda_0) z.
$$

(6)

The modulation of the core refractive index of a grating can present a Gaussian envelope due to the writing process or to achieve the apodisation of the reflection spectrum. A Gaussian envelope is described here by equation (7):

$$
\Delta n_0(z) = \Delta n_{oc} \exp \left[ -\alpha \left( \frac{z - l/2}{l} \right)^2 \right],
$$

(7)

where $\alpha$ is a control parameter and $n_{oc}$ is the amplitude corresponding to the central point of the envelope.
3.2 Differential evolution

The DE is a metaheuristic optimisation method initially proposed by Storn (1996). Due to its characteristics, the method has been used successfully in many engineering optimisation problems (Kalegari and Lopes, 2013; Krause and Lopes, 2013). In this application of the DE algorithm, a population $NP = 96$ individuals is evolved. Each individual corresponds to a deformation profile, discretised in $D = 20$ points.

The individuals are randomly initiated. At each iteration of the method, a new trial population is built from the current population by using the crossover and mutation operators. The new individuals are evaluated by the fitness function, replacing the old individuals if a higher fitness value is found. This iterative process runs for 2,000 iterations (value experimentally chosen), when the individual with highest fitness is selected as solution. Figure 2 shows an algorithmic description of the DE method, with the mutation and crossover operators described in Figures 3 and 4.

**Figure 2** Differential evolution (rand/1/bin scheme)

```
1: for $d$ ← 1 to $D$ do
2: \[ mut^d_i \leftarrow x^d_i + F \times (x^a_i - x^b_i) \]
3: end for
4: return \( mut \),
```

**Figure 3** Mutation operator (rand/1/bin scheme)

```
1: for $d$ ← 1 to $D$ do
2: \[ mut^d_i \leftarrow x^d_i + F \times (x^a_i - x^b_i) \]
3: end for
4: return \( mut \),
```

**Figure 4** Crossover operator (rand/1/bin scheme)

```
1: fixed ← random integer \( \in \{1, \ldots, D\} \)
2: for $d$ ← 1 to $D$ do
3: if random(0 to 1) < CR or $D = fixed$ then
4: \( trial^d_i \leftarrow mut^d_i \)
5: else
6: \( trial^d_i \leftarrow x^d_i \)
7: end if
8: end for
9: return \( trial \),
```

The DE/rand/1/bin (Price et al., 2005) scheme was employed with a crossover rate $CR = 0.95$ and mutation factor $F = 0.7$. This classic scheme was preferred here for its simplicity, as newer variations (Zhang and Sanderson, 2009; Cai et al., 2014) of the DE method did not result in a significant enhancement for this application. Each individual in the current population (base-individual) is used to construct a new individual in the trial population by means of the mutation and crossover operators. For each base-individual, the mutation operator chooses another three distinct individuals from the base population and combines them, generating a new mutant individual. By using the crossover operator, the mutant individual is recombined with the base-individual, generating a new individual in the trial population. The trial population is evaluated and its individuals replaces their respective base-individual if a higher fitness is reached.

The fitness evaluation is performed by simulating the reflectance spectrum of the FBG when applying the deformation profile represented by the individual. The mean squared error (MSE) between the simulated reflectance and the target spectrum is computed, with the fitness value being inversely proportional to the MSE. Thus, the algorithm searches for a deformation profile that minimises the MSE.

However, distinct deformation profiles can result in similar reflectance spectra. To avoid this ambiguity, we employ FBGs with linear chirp. Also, it is assumed that there are no abrupt changes in the deformation profile. This smoothness is enforced by penalising solutions which sections showed a strain differing more than 0.2 mC from the strain in the previous section. This penalty was added to the fitness evaluation and corresponds to the excess difference between the strain of each consecutive section, multiplied by 10 and divided by the total number of sections, as seen in Negri et al. (2015).

3.3 Parallel computation using GPU

Evaluating the fitness of the individuals is the most computationally expensive procedure of the proposed method. The fitness evaluation of a single individual requires computing the FBG reflectance for a sequence of 64 wavelengths in a predefined range. For each wavelength, the transfer matrix method is employed to compute the corresponding reflectance, requiring the calculation of a sequence of matrices that are then multiplied to compose the grating response. However, this problem can be easily divided into independent problems, taking advantage of a hardware capable of executing parallel threads.

In the fitness evaluation procedure, the evaluation of each individual can be seen as an independent process that depends only on the FBG structural parameters and the information carried by the individual. The evaluation of an individual can be further divided into independent threads, each one being responsible for computing the reflectance at a given wavelength. The difference between the resulting reflectance and the target spectrum is computed point by point each thread, and the total error for an individual is computed by using a parallel sum reduction operator.

The characteristics of the proposed problem make modern GPUs a very suitable hardware for performing the fitness evaluations, due to the high number of independent threads and low requirements on memory transfer and sharing. The compute unified device architecture (CUDA®) platform was used in this work with compatible GPU boards. The DE method was implemented in CPU, with the exception of the fitness evaluation that was implemented as a *kernel* for parallel execution in a GPU. This strategy of running the fitness evaluation of an
evolutionary method on GPU, while the overall process is run on CPU has been successfully employed in the literature (Harding and Banzhaf, 2008).

An algorithmic representation of the kernel is shown in Figure 5. This kernel is initially executed with a total of 96 thread blocks, each block corresponding to one individual. Each block is composed by 64 threads, where each thread simulates the FBG response for a given wavelength. The kernel creates a total of $96 \times 64 = 6,144$ threads at each iteration, with the number of actual parallel threads being dependent of the GPU hardware used. Each process block shares the information about the individual analysed, with the FBG parameters stored in static read-only global block. Each process block consists of individuals (process blocks) and wavelengths (process per block) were increased to the correct solution by using three deformation profiles (process per block). Two experiments were performed, $E1$ and $E2$. The first experiment ($E1$) evaluated the convergence of the method with the number of segments, $n_s$, is the number of segments (20), $\text{target}^i$ is the target strain in the $i^{th}$ segment and $\text{result}^i$ is the resulting strain in the $i^{th}$ segment. Each profile was evaluated 1,000 times, since the method is stochastic.

$$MAE = \sum_{i=1}^{n_s} |\text{target}^i - \text{result}^i|$$

In experiment $E1$, the mean absolute error (MAE) between the resulting deformation profile and the target deformation profile was used as error metric. Equation (8) was used to compute the MAE, where $n_s$ is the number of segments (20), $\text{target}^i$ is the target strain in the $i^{th}$ segment and $\text{result}^i$ is the resulting strain in the $i^{th}$ segment. Each profile was evaluated 1,000 times, since the method is stochastic.

The second experiment ($E2$) evaluated the performance of the method on three distinct GPUs ($\text{GPU}1$, $\text{GPU}2$, and $\text{GPU}3$), by varying the number of individuals to 96, 384, and 1,536 and the number of distinct wavelengths, from 16 to 1,024 (in steps of power of two). The employed GPUs were an NVIDIA GeForce® GTX480 ($\text{GPU}1$), an NVIDIA GeForce® GTX660 ($\text{GPU}2$), and an NVIDIA Tesla®K40C ($\text{GPU}3$). The specifications of the GPUs are shown in Table 1. Profile $P3$ was used as the target profile in experiment $E2$. Each wavelength evaluation was repeated five times in order to measure the mean running time and to increase precision of the timing.

The timing was done for the entire process and not only for the time spent executing the kernels in the GPUs. This timing strategy includes the initialisation time, that can be significant for some applications.

Both the experiments $E1$ and $E2$ used the FBG structural parameters shown in Table 2.

### Table 1 Specifications of the evaluated GPUs

<table>
<thead>
<tr>
<th>GPU</th>
<th>Clock rate</th>
<th>CUDA cores</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{GPU}1$ (GTX480)</td>
<td>1.40 GHz</td>
<td>480</td>
</tr>
<tr>
<td>$\text{GPU}2$ (GTX660)</td>
<td>1.06 GHz</td>
<td>960</td>
</tr>
<tr>
<td>$\text{GPU}3$ (K40C)</td>
<td>0.75 GHz</td>
<td>2,880</td>
</tr>
</tbody>
</table>

### Table 2 Parameters used to simulate the spectral response of the FBGs

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{11}$</td>
<td>0.113</td>
</tr>
<tr>
<td>$p_{12}$</td>
<td>0.252</td>
</tr>
<tr>
<td>$v_f$</td>
<td>0.17</td>
</tr>
<tr>
<td>$l$</td>
<td>1 cm</td>
</tr>
<tr>
<td>$\Lambda$</td>
<td>532.4 nm</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>1,552 nm to 1,555 nm</td>
</tr>
<tr>
<td>$\pi$</td>
<td>1.457</td>
</tr>
<tr>
<td>$\Delta n$</td>
<td>$2.5 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\delta \Lambda$</td>
<td>$5 \times 10^{-8}$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>5</td>
</tr>
</tbody>
</table>
4.1 Results

An example of the evolutionary process in experiment $E1$ with target profile $P3$ is shown on Figure 7, where the best solution found is seen at different iterations (1, 250, and 2,000).

Figure 7  Example of evolution of the best individual in experiment $E1$, using profile $P3$ as target (see online version for colours)

A summarisation of the results of experiment $E1$ is shown in Table 3, while the complete results (MAE computed for every repetition of all experiments) are shown in Figure 8. These results corresponds to the ones already presented by Negri et al. (2015).

Table 3  Results from experiment $E1$, according to the target profile

<table>
<thead>
<tr>
<th>Experiment</th>
<th>MAE [mε]</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P1$</td>
<td>0.0006937 ± 0.078082</td>
</tr>
<tr>
<td>$P2$</td>
<td>0.0009022 ± 0.050273</td>
</tr>
<tr>
<td>$P3$</td>
<td>0.003254 ± 0.040156</td>
</tr>
</tbody>
</table>

Figure 8  MAE obtained in experiment $E1$ (see online version for colours)

5 Discussion

According to the results of experiment $E1$ shown in Table 3, the MAE for all experiments was in the order of $\mu\varepsilon$. Results from Table 3 also presented a standard deviation higher than the mean value. This high deviation shows that some repetitions did not converge to the expected solution. This can also be inferred from the results shown in Figure 8, where a small number of repetitions differs in orders of magnitude from the others.

Figure 9 shows that for 96 individuals (from 1,536 to 98,304 threads), the number of threads is small enough for the performance to be directly related to the GPU clock rate of the evaluated GPUs. When using 96 individuals, GPU1 (the GPU with the highest clock rate) had the best performance for all tests, while GPU3 (lowest clock rate) showed the worst performance.

For 384 individuals, Figure 10 shows that the number of CUDA cores starts to be more important for the performance when the number of threads is high enough (turning point at 512 wavelengths, meaning 196,608 threads).

The tests performed with 1,536 individuals (Figure 11) show a greater difference in the performance of the GPUs for a higher thread count ($1,536 \times 1,024 = 1,572,864$...
threads at 1,024 wavelengths). For 1,536 individuals, it is clear that the higher number of CUDA cores of GPU3 (in comparison to GPU1 and GPU2) compensates for its lower GPU clock.

Figure 11  Running time for each GPU when using 1,536 individuals (see online version for colours)

Experiment E2 shows that for a lower thread count the GPU clock rate may be more significant than the number of CUDA cores, whereas a higher thread count changes the situation. Different applications may have different results than the ones obtained in this work, depending on the memory requirements of the application. Since a relatively low amount of memory needed to be shared between the threads, memory transfer speed and the amount of shared memory per block were not critical issues.

Also, results from experiment E2 suggest that a single GPU can evaluate the fitness for multiple populations (for a multi-population DE), allowing for detecting the strain profile on multiple FBG sensors. This can be seen when comparing the results of the GPUs for 64 wavelengths for 96, 384, and 1,536 points. Taking GPU1 as example, the running time for 64 wavelengths is multiplied by 4.3 when the population was multiplied by 16 (from 96 to 1,536 individuals).

6 Conclusions

Previous results (Negri et al., 2015) showed that the GPU implementation had a decrease in orders of magnitude when compared to existing methods. The results obtained in this work complements the previous results, showing that GPUs still have capacity for more parallel load.

The performance evaluation with distinct GPUs showed that the GPU clock rate is more significant than the number of cores for smaller problem sizes. Accordingly, a higher number of cores compensated lower clock rates for larger problem sizes (higher thread count). This information is useful as a criteria for choosing an adequate GPU for parallel applications.

Results showed that, by using the proposed evolutionary method in conjunction with high performance GPUs, it is possible to recover the mechanical deformation profile applied on an FBG in less than a second. The achieved processing rate greatly enhances the applicability of the method when compared to previous (Cheng and Lo, 2004; Negri et al., 2014) CPU implementations.

Acknowledgements

The authors acknowledge the collaboration of Dr. Hugo A. Perlin from the Paraná Federal Institute of Education, Science and Technology – Paranaguá and the financial support received from the CAPES, CNPq, FINEP, and Fundação Araucária Brazilian organisations.

References


