

Interactive comment on “Porting marine ecosystem model spin-up using transport matrices to GPUs” by E. Siewertsen et al.

Anonymous Referee #2

Received and published: 3 September 2012

General

This paper presents a method of spinning up ocean ecosystem models using GPGPU infrastructure. The authors show good performance compared to traditional CPUs, with a single GPGPU performing as well as 17 2.93Ghz Gainestown processors. Whilst this paper presents nothing groundbreaking it shows the performance gains that can be had with GPGPU software and as such is worthy of publication.

However, more attention needs to be paid on the performance comparisons. From the paper it is not clear that "one cycle" of timings include data transfer to/from GPGPU. More detailed analysis of the performance difference is required, such as % of peak FLOP, for example.

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

[Discussion Paper](#)

Some of the language used in the paper is not suitable for publication as it contains a number of informal constructs, e.g.: pg2180, line 17: "...that a consumer CPU can beat a significant number..." pg2181, line 17 "(if there is any)" pg2194, line 22 "...which basically looks like this:"

There are a number of similar examples throughout the paper.

Specific:

- pg2181 line: 17. There must be a better reference for MITGCM than a webpage - line 25: conform to, not with - use of the word "predestined". As per the first reviewers comments, please explain /why/ GPGUs are suitable for such operations. References for this? - "...big performance gain" -> "...performance gain"
- pg 2182, line 5-10. Hard to make sense of this sentence - line 19 - sentence is not complete - line 27 - "three space dimensions" -> "three spatial dimensions"
- pg2182-2183 line 27-line 2 - Not sure what this sentence means - pg2185, line 21 - Better reference for MPI - line 23, same for PETSc. There are better references than the websites!
- pg2186. A description of CSR format is not required.
- pg2191. Might be worth mentioning if the GPGPU can handle double floats, or just floats.
- pg2192. I don't think the file extensions are necessary - line 12. Again, better reference than the website, please
- pg2193. Reference for CUSP
- pg2194. line 7-9. If this paragraph/sentence required? If you are going to say you didn't use an "official" release, then say why and which features were missing at the time this work was done

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

[Discussion Paper](#)

- pg2197. Does the description of the hardware really need a table?
- pg 2198. Worth noting that 1 degree resolution is now fairly low resolution and that this might well be an issue for using GPGPUs in this way - line: 12. "Somehow 'fair'". Is it fair? The CPU compiler performs multiple optimisations? - line 10-12. If you have 100 runs, then report the fastest, not the mean - that is the top speed the code runs at. Also, please report the mean, and std dev too - why not? Most speed-up comparisons consist of 2-5 runs only, so this it's very good to have so much data available here.
- Pg2200, line 5-8. Remove the (!). Comparing speed only is not fair, despite the comment comment two pages ago. GPGUs have thousands of cores, compared to the 8 of a CPU. Whilst, I agree it is very hard to compare performance directly, I think this deserves more attention than simply raw wallclock time. What about comparing cost or watts? % of peak FLOP performance? Do the times include the data transfer to/from the GPGPU - the paper isn't clear on this.
- Pg2206. Table 3. "totally 480 CUDA" -> "a total of 480 CUDA"
- Pg 2207. Table 4 - this table could contain more info, like the max, min and std dev.
- Pg 2210. Fig 2. Can the names be made more generic? e.g. metos3d.exe -> executable. I think this would make the figure clearer.
- Pg 2211-12. Fig 3, 4. Can't tell the difference between colours when printed in greyscale. Please alter accordingly.
- Pg 2213. Fig 5. Why a block size 166. I got to this figure and realised the block size had been mentioned, but the reason for 166 was not explained. Is there an explanation?
- Pg2214. Fig 6. It took me a while to understand this figure. The vertical dashed lines look just like the GeForce line. The other processors look identical in greyscale.