Geosci. Model Dev. Discuss., 8, C2931–C2934, 2015
www.geosci-model-dev-discuss.net/8/C2931/2015/

Geoscientific
Model Development
Discussions
Open Access

Interactive comment on "Improving data transfer for model coupling" *by* C. Zhang et al.

Anonymous Referee #1

Received and published: 17 November 2015

The authors describe an approach to optimise compute and communication time for the data transfer between a pair of domain decomposed numerical models. They describe the implementation of a combination of two existing algorithms for a data exchange in such settings, peer-to-peer and butterfly. Novel in this publication is their flexible approach to select an appropriate combination of the two algorithms to minimise the wall clock time required to perform the data exchange.

The general idea, the combining of several algorithms to achieve the best performance depending on the number of hardware resources, message size, and number of messages, is not new and is - to the referees knowledge - at the heart of several MPI implementations. However, a combination of algorithms has not yet been used or investigated in the context of data exchange and climate model coupling.

Even though the approach the authors take is valid and worth to be investigated from

an algorithmic point of the authors fail to elaborate on possible benefits for existing coupled climate models. At the end, this leaves me with the question about whom the authors would like to address with their publication.

Grammar and syntax needs to be considered much more carefully.

The software does not contain any version number.

Introduction

In the introduction the authors raise the impression that they address high-resolution climate model applications running on modern high-performance compute systems where a single model employs several thousand processors or cores. Later on the algorithmic approach is investigated with test cases at very coarse resolution o(2 degrees) on a comparatively low number of cores (192) and leave the reader alone with any guess about the scalability of their approach.

P8983,L1: Is it the number of coupled models or the number of coupled model configurations the authors have in mind?

P8984,L27: Do you believe or are you convinced?

Section 2

The main - if not the only - purpose of section 2 is to provide the reader with an overview about the communication algorithms which are used in existing coupling software. In essence this section is telling us that all existing coupling software products use P2P communication. I wonder why I have to read approx. 85 lines to arrive at this. An overview of existing coupler software has already been published elsewhere - among the GMD - and the author should be able to reference those rather than providing another overview.

In section 4 the headline raises the expectation that we can learn how the butterfly algorithm works. The reader is not really guided through this section. Is the numbered list in sec 4.1 based on findings by the authors? In this case some piece of information is missing which guides the reader to this statement. In case it is not based on the authors findings a reference is missing. Fig. 6 (and likewise Fig 8) does not help me at all to learn how the butterfly algorithm works. If each of the 8 processes P0 to P8 already has all data D0 to D8 I cannot see any necessity for communication. What is the information that shall be transported to the reader with the colours?

I would have loved to be guided through Fig 7 in the text a little bit. If this Figure is not important at all it should be removed.

In section 5 it remains unclear (to me at least) how the adaptive process works and I would appreciate if this was clarified in a revised version. Does this work as a kind of self-learning algorithm where the optimal path is determined of the first n data exchanges of a model integration or is this part of the initialisation procedure beforehand and made available already for the first data exchange?

The first sentence of section 6 does not make sense to me. Having read the previous sections the authors put the focus of the reader to the adaptive transfer library. Now the authors propose the butterfly implementation as well. Later we learn that the butterfly approach can be outperformed. At the end of the section the authors show that for coupled climate models the P2P communication is as good as the adaptive transfer library, probably because the adaptive transfer library completely switches to P2P in the latter case. I think that this is an important finding and should be emphasised. It tells us that the P2P which is used in existing coupler software is not that bad. But is also tell me that the paper is severely suffering from a clear structure. If my conclusion (P2P is sufficient) is wrong the authors will need to put more effort in getting the reader onto the correct track.

Table 1 and Fig 10 are not really addressed. Are they required to understand the adaptive data transfer library? These can be removed of shifted to the user guide.

Could Fig 9 be replaced by a real flow chart rather than providing pseudo code?

In section 6 the performance of the data transfer is evaluated by using a coupled climate model with roughly 2 degree grid horizontal grid spacing using 192 processes. As there are 8400 cores available Tansuo100 I would have expected to see an evaluation of the performance at least with a toy model and exploring the scalability of the adaptive data transfer library up to several thousand cores. Unless there are sound arguments why this cannot be done this raises the impression that the authors are trying to hide something. The dynamical core sets an upper limit to the number of cores that can reasonably be employed - when the communication starts dominating over the computing part (MPI messages required for the boundary exchange required for advection and diffusion operators versus the time for the forward integration of the less and less points left on a single core). With roughly 2 degree resolution we have probably reached this point with 192 processes. Here it would be nice to know how much percentage of the overall compute time is consumed by the data exchange, and how much wall clock time can be gained for a single run of the coupled model. Last but not least, how important is the load imbalance between the processes as the boundary exchange between the model components (atmosphere and ocean) provides a synchronisation point, either explicitly or implicitly, where the components have to wait for each others.

The conclusions are weak if not misleading. Fig. 17 does not really confirm the last statement, that “the adaptive transfer library can effectively improve the performance of data transfer in model coupling. What can we conclude or expect for model with higher resolution than those investigated in this study?

Interactive comment on Geosci. Model Dev. Discuss., 8, 8981, 2015.