

Abschlussbericht

Verbundprojekt ScalES: Entwickeln von Methoden und Softwaremodulen
mit dem Ziel einer hohen und effizienten Skalierung komplexer
gekoppelter Simulationsmodelle auf modernen Parallelrechnern

AP3: Koppler

1.1.2009 – 31.12.2011

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01IH08004E gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt beim Autor.

1. Kurzdarstellung der Ausgangslage und Projektplan

1.1 Aufgabenstellung

In der heutigen Klimamodellierung werden Modelle, die ursprünglich für Rechner mit einem oder wenigen Cores entwickelt wurden, zu komplexen Erd-System-Modellen (ESMs) gekoppelt. Diese Systeme konnten zu Projektbeginn auf Höchstleistungsrechnern mit tausenden von Cores oft noch nicht effektiv genutzt werden. Um mit diesen ESMs möglichst flexibel auf neue wissenschaftliche Fragestellungen eingehen zu können ist es notwendig, separat betreibbare Klimakomponenten in bestehende Systeme möglichst einfach ein- und auskoppeln zu können.

1.2 Voraussetzungen, unter denen das Vorhaben durchgeführt wurde

In Zusammenarbeit mit dem assoziierten Partner NEC-IT wurde beschlossen, dass der Zuwendungsempfänger, das Alfred-Wegener-Institut für Polar- und Meeresforschung, aufbauend auf den PRISM-Koppler OASIS4, eine Erweiterung zur Kopplung unstrukturierter Gitter entwickeln sollte. Im Rahmen von ScaleS sollte so der OASIS4 Koppler eine Schnittstelle erhalten, die die Unterstützung allgemeiner unstrukturierter Gitter erlaubt. Prototypisch sollte hier eine Kopplung mit dem am AWI entwickelten unstrukturierten Ozeanmodell FESOM (finite element sea ice ocean model) realisiert werden. Diese Anforderung stand gleichberechtigt neben der Notwendigkeit, eine effektive Kopplung zwischen verschiedenen Partitionen der gleichen Modellkomponente zu realisieren (Zusammenarbeit mit AP2). Hierzu sollte die Nachbarschaftssuche auf unstrukturierten Gitterkonzepten erweitert bzw. eine effiziente Suche auf unstrukturierten Gittern entwickelt werden. Um zukünftig auch adaptive Gitter geeignet unterstützen zu können, war diese Suche möglichst effizient zu gestalten. Die hierbei zu entwickelnden Methoden wären auch bei der Integration des Parallelisierungs- und Partitionierungsmodul (PPM, s. AP2) in den Koppler von Vorteil.

1.3 Planung und Ablauf des Vorhabens

Das Vorhaben wurde in vier Phasen bzw. Tasks unterteilt: Vorarbeiten und Problemanalyse, Erweiterung der Schnittstellen und Implementierung, Untersuchung der Skalierbarkeit, sowie Erweiterung des Kopplercodes zur Unterstützung unstrukturierter Gitter. Im Allgemeinen wurden die Tasks auch in dieser Weise abgearbeitet, im Detail ergab sich aus technischen Gründen an einigen Stellen die Notwendigkeit, einzelne Unterpunkte in ihrer Reihenfolge zu vertauschen. So konnte die Untersuchung der Skalierbarkeit der Modellkomponenten bereits zu einem frühen Zeitpunkt erfolgen; andererseits verzögerte sich die Arbeit an einer Erweiterung der Kopplerfunktionalität durch fehlerhafte Bibliotheken auf dem zur Entwicklung verwendeten Großrechner einerseits, und dem sich in intensiver Fortentwicklung befindenen Prototypen von OASIS4 andererseits. Die Gründe für die Abweichungen gegenüber der ursprünglichen Planung sind z.T. sehr technisch; auf sie wird im Verlauf detailliert eingegangen.

1.4 Wissenschaftlicher und technischer Stand, an den angeknüpft wurde

Um den Kopplungsanforderungen gerecht zu werden, lag zu Projektbeginn ein Prototyp des im Rahmen der PRISM-Initiative entwickelten OASIS4-Kopplers vor. Das Design dieses Kopplers berücksichtigt dabei die Parallelität der zu koppenden Modell-Komponenten und sollte in einem ersten Arbeitsschritt den sequentiellen OASIS3 Koppler ersetzen, der bis dahin in das COSMOS-Modellsystem (ECHAM5.4

und MPIOM gekoppelt über OASIS3) integriert war. Zu Projektbeginn unterstützte OASIS4 bereits konzeptionell die Kopplung regulärer Gitter, sowie blockstrukturierter Gitter und sogenannte reduzierte Gauß-Gitter, die als Sonderfall eines unstrukturierter Gitters gelten.

1.5 Zusammenarbeit mit anderen Stellen

Die ursprünglich anvisierte Kooperation mit NEC-IT kam nicht zustande, da die eingebundene Abteilung von NEC geschlossen wurde. Durch diese nicht vorhersehbare Entwicklung kam es im Weiteren auch zu Verzögerungen. Mit CERFACS und Dr. Rene Redler konnten jedoch zwei wertvolle Partner gewonnen werden, die uns im Verlaufe des Projektes sehr unterstützt haben.

2. Projektdurchführung und Abschluss

2.1 Erzielte Ergebnisse und Vorgaben aus dem Projektplan

– Task 1: Vorarbeiten, Problemanalyse

Schon zu Beginn des Projektes hat sich eine wichtige Änderung gegenüber der Ursprungsplanung ergeben: der Kooperationsvertrag des Konsortiums konnte vom Industriepartner NEC – IT nicht unterzeichnet werden, da die eingebundene Abteilung geschlossen wurde. NEC - IT stand dem AP 3, Koppler daher als assoziierter Partner nicht zur Verfügung. Die Unterstützung und das Fachwissen, das von NEC – IT in das Projekt eingebracht werden sollte, muss von den verbleibenden Mitarbeitern des Arbeitspakets selber aufgebaut werden. Durch diese nicht vorhersehbare Entwicklung kam es im Weiteren durchaus zu Verzögerungen. Die intensive Zusammenarbeit mit dem CERFACS in Frankreich und Dr. Rene Redler (MPI-M) hat sich glücklicherweise hierbei als sehr fruchtbar erwiesen.

Bevor mit der Entwicklung des Kopplers begonnen werden konnte, mussten wichtige Vorarbeiten die Verwendbarkeit der Komponenten gewährleisten:

1. **FESOM:** Der Einsatz der in FESOM fest implementierten Löserbibliothek PetSc hat sich auf verschiedenen Systemen als durchaus schwierig erwiesen. So wurde im Antrag eine plattformübergreifende Implementierung des gekoppelten Systems mit ECHAM5.4 zugesagt und die Referenzinstallation sollte auf der IBM Blizzard am DKRZ stattfinden. Allerdings war es erst Anfang Juni 2010 möglich PetSc auf der IBM Blizzard zu verwenden, d.h. erst seit diesem Zeitpunkt ist das gekoppelte System am DKRZ lauffähig. Ferner sollte eine Plattformunabhängigkeit demonstriert werden indem es außerdem auf einen NEC Vektorplattrechner portiert wird. Diese Portierung konnte im Rahmen von ScaleS nicht realisiert werden, da sich PetSc als auf Vektorplattformen nicht portierbar herausstellte. Nach Rücksprache mit dem Hardwarehersteller und eigenen Analysen hätte der Zeit – und Personalaufwand für die Portierung in keinem Verhältnis zum Resultat gestanden, besonders da nicht damit gerechnet werden konnte, dass die zu erwartende Performance auch nur näherungsweise zufriedenstellend sein würde. Überlegungen, die verwendeten PetSc-Routinen durch Routinen verfügbarer und nachgewiesener performanter Bibliotheken wie AZTEC zu ersetzen, sind an deren Lizenzbedingungen gescheitert. Erst seit

Ende 2011 ist eine FESOM-Version verfügbar, die auf die Bibliothek PetSc komplett verzichtet, diese kam aber für die Arbeiten in ScaLES zu spät.

FESOM wird im Rahmen anderer Projekte inzwischen großskalig auf den Rechnern des HLRN und DKRZ eingesetzt, auch aufbauend auf den Vorarbeiten in ScaLES.(s. wissenschaftliche Anschlussfähigkeit)

2. **ECHAM5.4** konnte auf den Systemen am AWI (NEC SX8 / IBM) sowie am DKRZ in Hamburg (IBM System blizzard) als „Stand Alone“ Komponente zum Laufen gebracht werden. ECHAM gehört zu den am MPI-M in Hamburg entwickelten Klimasystem-Komponenten, und wird in der gesamten Community intensiv genutzt. Das Modell skaliert problemlos auf mehrere tausend Cores, und liegt inzwischen (nach Abschluss von ScaLES) in der Release-Version 6.1 vor.
3. **OASIS4** liegt als lauffähige Komponente zu Projektende ebenfalls auf allen drei oben genannten Systemen vor. Aufgrund eines Compilerbugs in der IBM - MPI2 Library war es bis zum 2. Halbjahr 2010 jedoch nicht möglich, die ‚spawn – Funktionalität‘ von OASIS4 zu verwenden. Zwischenzeitlich ist daher der Code von OASIS4 insofern überarbeitet worden, dass auch ohne spawning eine Skalierungsanalyse möglich war.

Entgegen Versicherungen der Projektleitung des OASIS4 Kopplers vom CERFACS kam es während der gesamten Laufzeit immer wieder zu starken, strukturellen Modifikationen im Modellcode. Dies machte verlässliche Performance Tests einerseits, wie auch Optimierungen im Code vorläufig und schwierig. Man musste bei Optimierungen im Kopplercode jederzeit das Risiko eingehen, dass diese in einem folgenden Release des Kopplers obsolet geworden wären. Erst im Frühjahr 2011 lag tatsächlich eine OASIS4 – Codeversion vor, die als relativ stabiles Zwischenrelease bezeichnet werden konnte. Dennoch erforderte auch diese Version viel Zeit für die Fehlersuche und Überarbeitung des Kopplerkonzeptes.

Als unvermeidliche Folge ging bereits in den letzten Monaten der Projektlaufzeit das allgemeine Interesse der Nutzerschaft an OASIS4 deutlich zurück. Verschiedene internationale Arbeitsgruppen suchten zunehmend nach Alternativen.

Nach Ende des ScaLES-Projektes wurde dann offiziell, dass das CERFACS den Support und die Weiterentwicklung von OASIS4 einstellen wird. Dennoch hat von ScaLES in OASIS4 investierte Arbeit zu Einsichten in die Kopplerarchitektur - insbesondere in Bezug auf die Kommunikation unstrukturierter Daten - die im Falle eines Wechsels zu einem alternativen Produkt (oder gar der Entwicklung eines eigenen Kopplers) ihre Anwendung finden werden.

- *Task 2: Erweiterung der Schnittstellen und Implementierung*

Das prinzipielle Konzept des Kopplers OASIS4 sieht vor, die Gitterinformationen der Modellkomponenten über eine einheitliche XML-Schnittstelle an den Koppler zu transportieren, der daraufhin die Kommunikation und Verarbeitung ermöglicht. Für strukturierte Gitter ist die Erstellung dieses XML-Files mit Hilfe eines Metadatenstandards (ISO19135 / ISO 19115) möglich. Für die Verkläuserung von unstrukturierten Gittern existiert dieser Standard bisher noch nicht, an der Bildung eines vergleichbaren Standards wird mitgearbeitet.

Um eine Kommunikation der Modelle mit dem Koppler zu ermöglichen, mussten Schnittstellen zwischen ECHAM5.4 und OASIS4 sowie FESOM und OASIS4 im

Rahmen von ScaleS neu entwickelt werden. Hierbei wurde Wert darauf gelegt, die Schnittstellen von OASIS4 mit unstrukturierten Gittern generell flexibel zu halten.

Für den Betrieb mit standardisierten Datenformaten wurde FESOM zudem um eine Schnittstelle erweitert, die den eigenen, unstrukturierten Output als strukturierten NetCDF-Datensatz ausgeben und strukturierte NetCDF-Files als Antriebsdaten einlesen und auf das unstrukturierte Gitter interpolieren kann.

- *Task 3: Untersuchung der Skalierbarkeit*

Die prinzipielle Skalierbarkeit von ECHAM5.4, FESOM und OASIS4 als Einzelkomponenten konnte sichergestellt werden. Bei geeigneter Wahl der Problemgröße konnte keine technische Begrenzung der Skalierbarkeit gefunden werden. Eine Performanceanalyse des Kopplers konnte anhand der mitgelieferten Standardbeispiele auf der NEC SX8-R durchgeführt werden.

- *Task 4: Erweiterung des Kopplercodes zur Unterstützung unstrukturierter Gitter*

Als zentrales Produkt des ScaleS-AP3 war eine Kopplungssoftware geplant, die allgemein unstrukturierte Gitter unterstützt. Als Testcase hierfür war ein gekoppeltes System aus dem Finite-Elemente-Code FESOM, dem Koppler OASIS4 sowie dem auf regulärem Gitter basierenden Atmosphärenmodell ECHAM5.4 geplant. Beide Ziele konnten im Verlauf des Projektes realisiert werden.

Das zentrale Problem bei der Kopplung von Daten, die auf unstrukturierten Gittern repräsentiert sind, ist die Kommunikation der Gitterinformationen. Konkret bedeutet dies, dass die geographischen Koordinaten der Gitterpunkte, sowie der impliziten Interpolationsroutinen von einer Modellkomponente zur anderen kommuniziert werden müssen. Gerade mit Bezug auf die Basisfunktionenräume sind hier prinzipiell beliebige Auswahlen möglich, die in der Regel nicht durch einen endlichen Satz an Zahlen beschrieben werden können. Um dieses Problem zu umgehen, haben wir eine Kopplungsstrategie entwickelt, bei der insgesamt drei Phasen zu unterscheiden sind:

- **Phase 1:** Der Austausch der Daten von FESOM an ECHAM5 über OASIS4 erfolgt in der ersten Realisierung über die Interpolation auf ein reguläres Zwischengitter. Die Gitterinformationen sowie die Informationen über die zu koppelnden physikalischen Größen werden über xml – Files dem System bekannt gemacht. Auch wenn dies nicht der ursprünglich angedachten Lösung entspricht, so hat sie doch einige Vorteile:
 1. Sie ist weniger ressourcenintensiv, als ursprünglich angenommen. Profilings und Tracings des gekoppelten Systems haben eindeutig gezeigt, dass der Zeitverlust durch die globale Kommunikation und Interpolation auf der FESOM-Seite um eine Größenordnung kleiner ist, als die von OASIS4 für den Austausch benötigte Zeit.
 2. Die Verwendung des regulären Zwischengitters ist zwischenzeitlich von den Anwendern als Lösung gerne akzeptiert worden, weil dadurch eine schnelle Visualisierung der interpolierten Daten mit Standardsoftware problemlos möglich ist.
- **Phase 2:** Das regelmäßige Zwischengitter wird obsolet, wenn es gelingt, unregelmäßige Gitter direkt am Koppler anzumelden. Da die Vielzahl der üblichen Gittertypen die Umsetzung erschwert, beschränkten wir uns zunächst auf den Fall linear oder konstant interpolierender Gitter (P0/P1). Für diese ist