

DISS. ETH NO. 22206

COMMUNICATION EFFICIENT ALGORITHMS  
FOR NUMERICAL PROBLEMS ON  
FULL AND SPARSE GRIDS

A thesis submitted to attain the degree of  
DOCTOR OF SCIENCES of ETH ZURICH  
(Dr. sc. ETH Zurich)

presented by

PHILIPP HUPP

Dipl.-Math. Univ., Technische Universität München, Munich  
MS in Mathematics, Georgia Institute of Technology, Atlanta

born 16.06.1987

citizen of Germany

accepted on the recommendation of  
Prof. Dr. Peter Widmayer, examiner  
Riko Jacob, PhD, co-examiner  
Prof. Dr. Ulrich Meyer, co-examiner

2014

## ABSTRACT

---

As the gap between peak performance and bandwidth on modern computers is already large and increasing further, the design of communication-efficient algorithms becomes more and more important. This effect is reinforced further, as the problems that we solve also continuously grow in size. The problems considered in this thesis stem from the field of solvers for partial differential equations (PDEs). For low-dimensional problems, stencil computations as they appear in finite difference methods are studied. For high-dimensional problems, this thesis focuses on the sparse grid combination technique, an hierarchical discretization scheme that reduces the “curse of dimensionality”. Based on extrapolation, the combination technique additionally reduces the need for global communication. For time dependent problems, e.g., time dependent PDEs, each time step is computed on several independent and much smaller discretizations, so called component grids, and the global communication shrinks to a reduce/broadcast step in between. This work analyzes this remaining communication bottleneck of the combination technique as well as hierarchization, one of the fundamental sparse grid algorithms.

In this work, theoretical considerations are used to identify the most important communication aspects of the examined problems. Communication, in this thesis, refers to both data transfer through the memory hierarchy as well as communication via message passing. For all considered problems, novel algorithmic approaches are derived and it is shown that these approaches are, with respect to communication, either optimal or within a small constant factor of the optimum. In two cases, the theoretical insights also immediately lead to efficient implementations of the respective algorithms. In detail, the contributions of this work are the following:

**STENCIL COMPUTATIONS.** To solve problems numerically, the solution space is first discretized, e.g., by being mapped to a full grid space, and then the numerical problem is solved in this discretized space. For PDEs this happens, for example, using stencil computations on grids. Stencil computations update the values of the grid points according to a stencil. The stencil describes which neighbors of a grid point are used for this update. This work derives new algorithms for stencil computations and proposes data layouts that exactly match the access patterns of these algorithms. The improved lower bounds are proven by carefully applying an isoperimetric result to the rounds into which an arbitrary algorithm can be split. In combination, the existing gaps between lower

and upper bounds reduce significantly, and the new bounds match for the 2-dimensional case.

**UNIDIRECTIONAL HIERARCHIZATION.** Hierarchization describes the base change from the full grid basis to the hierarchical basis of sparse grids. As such, hierarchization is one of the fundamental sparse grid algorithms, and furthermore it is an important preprocessing step for the communication schemes discussed later. The unidirectional principle is the dominating design pattern for sparse grids. It splits the  $d$ -dimensional problem into  $d$  distinct phases and solves each phase working on 1-dimensional subproblems only. This work describes an implementation of the unidirectional hierarchization algorithm for the component grids of the combination technique that runs within a factor of 1.5 of the lower bound given by the unidirectional principle and the memory bandwidth of the system used.

**DIVIDE AND CONQUER HIERARCHIZATION.** The unidirectional principle guides the development of many sparse grid algorithms as it simplifies their computations. It is, however, a bad choice with respect to data transfer. Due to the  $d$  phases of the unidirectional principle, any unidirectional algorithm has to load each grid point at least  $d$  times. This work develops an alternative approach for isotropic component grids: a divide and conquer algorithm that avoids the unidirectional principle globally but applies it recursively to smaller subproblems. Let  $M$  denote the size of the internal memory and let  $B$  denote the cache line size. The derived algorithm is cache-oblivious and, assuming a tall cache of size  $M \in \omega(B^d)$ , optimal, as it brings the leading term of the cache misses down to the compulsory scanning complexity. Furthermore, the algorithm is complemented with a non-trivial lower bound for the leading term of the non-compulsory cache misses.

**COMMUNICATION SCHEMES FOR THE COMBINATION TECHNIQUE.** The combination technique breaks the global communication requirements of conventional discretization approaches by splitting the problem into subproblems that can be solved independently. Global synchronization, however, is still necessary but shrinks to a reduce/broadcast step. This work presents the first systematic study of this remaining communication bottleneck and derives two communication schemes to solve it. One of the schemes exploits the hierarchical structure of sparse grids. The communication schemes are evaluated experimentally on two supercomputers, demonstrating their performance for different scenarios. Furthermore, the schemes are analyzed and a theoretical model is built to predict their performance. The model is validated using the experiments and then applied to predict scenarios that are as yet out of scope on current supercomputers due to their high computational demands.

## ZUSAMMENFASSUNG

---

Da die Kluft zwischen Prozessorleistung und Speicherbandbreite moderner Computer bereits groß ist und immer weiter wächst, wird die Entwicklung kommunikationseffizienter Algorithmen immer wichtiger. Dieser Effekt wird noch dadurch verstärkt, dass die Probleme, die wir lösen, auch stetig größer werden. Die Probleme, die in dieser Arbeit behandelt werden, kommen aus dem Gebiet der Löser für partielle Differentialgleichungen (PDG). Bezüglich niedrig-dimensionaler Probleme untersucht diese Arbeit Stencil-Berechnungen, wie sie in der Finite-Differenzen-Methode zur Anwendung kommen. Für hoch-dimensionale Probleme untersucht diese Arbeit die Kombinationstechnik für dünne Gitter. Die Kombinationstechnik ist ein hierarchisches Diskretisierungsschema, welches erlaubt den "Fluch der Dimension" zu reduzieren. Da die Kombinationstechnik auf Extrapolation basiert, kann sie darüber hinaus die notwendige globale Kommunikation verringern. Bei zeitabhängigen Problemen, z. B. bei zeitabhängigen PDGs, wird jeder Zeitschritt auf mehreren, unabhängigen und wesentlich kleineren Diskretisierungen, so genannten Teilgittern, berechnet. Da die einzelnen Teilgitter unabhängig voneinander berechnet werden können, schrumpft die globale Kommunikation auf eine Reduce/Broadcast-Phase zwischen den Zeitschritten. Die vorliegende Arbeit analysiert zum einen diesen verbleibenden Kommunikationsengpass der Kombinationstechnik, sowie den Hierarchisierungsalgorithmus, einen der grundlegenden Algorithmen für dünne Gitter.

In dieser Arbeit werden theoretische Überlegungen genutzt, um die wichtigsten Kommunikationsaspekte der untersuchten Probleme zu identifizieren. Sowohl der Datentransfer durch die Speicherhierarchie als auch der Nachrichtenaustausch auf verteilten Systemen wird in dieser Abhandlung unter dem Begriff "Kommunikation" zusammengefasst. Für alle untersuchten Probleme werden neue algorithmische Konzepte hergeleitet. Zudem wird bewiesen, dass diese Konzepte in Bezug auf die Kommunikation entweder optimal oder nur einen kleinen konstanten Faktor vom Optimum entfernt sind. In zwei Fällen führen die theoretischen Einsichten darüber hinaus direkt zu effizienten Implementierungen der entsprechenden Algorithmen. Die Beiträge dieser Arbeit sind im Detail:

**STENCIL-BERECHNUNGEN:** Um Probleme numerisch zu lösen wird der Lösungsraum zunächst diskretisiert. Dies passiert beispielsweise indem der Lösungsraum auf einen Vollgitterraum abgebildet wird. Das numerische Problem kann dann anschließend in diesem diskretisierten Raum gelöst werden. Bei PDGs kann die numerische Lösung des Pro-

blems beispielsweise durch Stencil-Berechnungen erfolgen. Stencil-Berechnungen aktualisieren die Werte der Gitterpunkte anhand von einem Stencil. Der Stencil gibt an, welche Nachbarn eines Gitterpunktes benutzt werden, um den neuen Wert des Gitterpunktes zu berechnen. Diese Arbeit leitet neue Algorithmen für Stencil-Berechnungen her und schlägt Datenformate vor, die die Zugriffsmuster der Algorithmen exakt abbilden. Die unteren Schranken werden verbessert, indem ein isoperimetrisches Resultat sorgfältig auf die Runden angewandt wird, in die ein beliebiger Algorithmus unterteilt werden kann. Insgesamt werden die bisher existierenden Diskrepanzen zwischen oberen und unteren Schranken wesentlich verringert und für den 2-dimensionalen Fall beseitigt.

**UNIDIREKTIONALES HIERARCHISIEREN:** Das Hierarchisieren beschreibt den Basiswechsel zwischen der Vollgitterbasis und der hierarchischen Basis der dünnen Gitter. Als solches ist das Hierarchisieren einer der grundlegenden Algorithmen für dünne Gitter. Darüber hinaus ist das Hierarchisieren ein wichtiger, vorbereitender Schritt für die Kommunikationsschemata, die später diskutiert werden. Das unidirektionale Prinzip ist das dominierende Entwurfsschema für Dünn-Gitter Algorithmen. Es zerlegt das  $d$ -dimensionale Problem in  $d$  verschiedene Phasen. In jeder dieser Phasen werden ausschließlich 1-dimensionale Teilprobleme gelöst. Diese Arbeit beschreibt die Implementierung eines unidirektionalen Hierarchisierungsalgorithmus für die Teilgitter der Kombinationstechnik, die die untere Schranke, welche durch das unidirektionale Prinzip und die Speicherbandbreite des benutzten Systems gegeben ist, bis auf einen Faktors von 1.5 erreicht.

**DIVIDE AND CONQUER HIERARCHISIEREN:** Das unidirektionale Prinzip lenkt die Entwicklung vieler Dünn-Gitter Algorithmen, weil es die Berechnungen vereinfacht. In Bezug auf den Datentransfer ist das unidirektionale Prinzip jedoch eine schlechte Wahl. Wegen der  $d$  Phasen des unidirektionalen Prinzips, muss jeder unidirektionale Algorithmus jeden Gitterpunkt mindestens  $d$  mal laden. Diese Arbeit entwickelt einen alternativen Ansatz für die isotropen Teilgitter der Kombinationstechnik. Der neue Algorithmus beruht auf dem "Divide and Conquer" (teilen und herrschen) Paradigma und vermeidet das unidirektionale Prinzip auf globaler Ebene. Stattdessen wird das unidirektionale Prinzip rekursiv auf kleinere Teilprobleme angewendet. Im Folgenden bezeichnet  $M$  die Größe des Caches und  $B$  die Größe einer Cache-Zeile. Der hergeleitete Algorithmus ist Cache-Oblivious und optimal unter der Annahme eines großen Caches, d. h. falls  $M \in \omega(B^d)$ , da er den führenden Term der Cache Ladevorgänge auf das minimal notwendige, einmalige Lesen der Eingabe reduziert. Des Weiteren wird der Algorithmus durch

eine nicht triviale untere Schranke für den führenden Term der nicht obligatorischen Cache Ladevorgänge ergänzt.

#### KOMMUNIKATIONSSCHEMATA FÜR DIE KOMBINATIONSTECHNIK:

Die Kombinationstechnik verringert die globalen Kommunikationsanforderungen herkömmlicher Diskretisierungsmethoden, indem sie ein Problem in mehrere Teilprobleme zerlegt, die unabhängig voneinander berechnet werden können. Eine globale Synchronisation ist nach wie vor notwendig, beschränkt sich jedoch auf eine Reduce/Broadcast-Phase. Diese Arbeit ist die erste systematische Untersuchung dieses verbleibenden Kommunikationsengpasses und leitet insbesondere zwei Kommunikationsschemata her, die diese Reduce/Broadcast-Phase durchführen. Eines der beiden Schemata nutzt dabei die hierarchische Struktur der dünnen Gitter. Beide Kommunikationsschemata werden durch Experimente auf zwei Supercomputern untersucht und zeigen dabei ihre jeweilige Leistungsfähigkeit in unterschiedlichen Szenarien. Darüber hinaus werden beide Schemata analysiert und es wird ein theoretisches Modell entworfen, welches ihre Laufzeit vorhersagen soll. Nachdem das Modell experimentell validiert wurde, wird es benutzt um Szenarien vorherzusagen, die auf Grund ihres hohen Berechnungsbedarfs momentan außerhalb der Möglichkeiten von Supercomputern liegen.