Inhalt

Zusammenfassung ... 1
1. Einführung ... 5
2. Multilevel-Methoden ... 6
   2.1. Modellproblem und Notation ... 7
   2.2. Vorkonditionierer ... 8
3. Parallelisierung von Multilevel-Methoden ... 10
   3.1. Theoretische Parallelrechner ... 11
   3.2. Ansätze für Parallelisierungen ... 12
   3.3. Das Kostenfunktional für Punktverteilungen ... 16
   3.4. Abgeleitete Kostenfunktionale ... 18
4. Ein Beispielprogramm ... 19
   4.1. Formulierung ... 20
   4.2. Parallelisierung ... 20
   4.3. Ergebnisse ... 21
5. Adaptive parallele Multilevel-Methoden ... 24
   5.1. Fehlerschätzer ... 24
   5.2. Gittermanipulation ... 25
6. Lastverteilung adaptiver paralleler Multilevel-Methoden ... 26
   6.1. Dynamische Ansätze ... 27
   6.2. Der schlechteste Fall für statische Ansätze ... 28
      6.2.1. Das Problem ... 29
      6.2.2. Ansätze ... 29
      6.2.3. Zusätzliche Voraussetzungen ... 30
   6.3. Übersicht über statische Ansätze ... 30
   6.3. Statistischer Ansatz mit geeigneter Komplexität ... 32
7. Experimentelle Ergebnisse ... 33
   7.1. Ergebnisse auf Intel iPSC/2 ... 34
   7.2. Ergebnisse auf T800 ... 39
   7.3. Ergebnisse auf Alliant FX/2800 ... 41
8. Schlußbemerkungen ... 44
9. Danksagungen ... 45
Literatur ... 46

Adaptive parallele Multilevel-Methoden zur Lösung elliptischer Randwertprobleme

Gerhard W. Zumbusch
mit Unterstützung von Prof. Dr. Ronald Hoppe

Technische Universität München
Arcisstr. 21, D-W-8 München 2

Zusammenfassung

Multilevel Methoden sind die zur Zeit effizientesten Verfahren zur Lösung großer symmetrischer schwachbesetzter linearer Gleichungssysteme, die aus der Diskretisierung selbstadjungierter elliptischer Randwertprobleme mit Finiten-Elementen entstehen. Im folgenden wird die Parallelisierung eines von Bramble, Pasciak und Xu vorgeschlagenen vorkonditionierten Verfahrens der konjugierten Gradienten, eingebettet in ein adaptives Finite-Elemente-Programm wie etwa Kaskade, diskutiert.
Dabei müssen zur effizienten Lastverteilung zusätzliche Forderungen an Triangulierungen, Finite-Elemente-Räume und Gittermanipulationsalgorithmen gestellt werden. Es werden Standardverfahren der Lastverteilung mit einem neuen Aufteilungsverfahren, das auf einem statistischen Ansatz beruht, verglichen. Mit einer hier vorgestellten gemischten Strategie der Aufteilung von Gitterpunkten und Dreiecken kann ein Gesamtverfahren von optimaler Ordnung erreicht werden.
Die experimentellen Ergebnisse auf einigen Parallelrechnern zeigen eine hohe Übereinstimmung mit einem hergeleiteten Kostenfunktional und eine ideale Parallelisierbarkeit des Verfahrens. Unterschiede zu anderen bekannten Multilevelverfahren werden in den einzelnen Abschnitten herausgestellt.

1. Einführung

Multilevel Methoden sind die zur Zeit effizientesten Verfahren zur Lösung großer linearer Gleichungssysteme, die aus der Diskretisierung elliptischer Randwertprobleme entstehen. Der Aufwand, mit Mehrgitterverfahren oder auch mit multilevel vorkonditionierten Verfahren der konjugierten Gradienten (CG) im symmetrisch positiv definiten Fall ein Gleichungssystem bis auf Diskretisierungsgenauigkeit zu lösen, ist unter geeigneten Voraussetzungen proportional zur Zahl der Unbekannten oder nur um einen logarithmischen Term höher.
Nach Standard-Mehrgitter V- und W-Zyklen [Hackbusch] sind in letzter Zeit Verfahren mit hierarchischen Basen, als Mehrgitterverfahren [Bank, Dupont & Yserentant] und als Vorkonditionierer [Yserentant] in den Mittelpunkt des Interesses gerückt, da in den Konvergenzbeweisen auf stark einschränkende Regularitätsannahmen verzichtet werden kann. Für Randwertprobleme in drei Raumdimensionen sind diese Verfahren allerdings nicht mehr von quasioptimaler Ordung wie im zweidimensionalen Fall, so daß sich als Alternative ein ähnlicher Ansatz von [Bramble, Pasciak & Xu] als Vorkonditionierer anbietet, dessen Mehrgittervariante der Standard V-Zyklus ist [Bramble, Pasciak, Wang & Xu]. Das entstehende Verfahren ist im Fall quasiuniformer Hierarchien von Triangulierungen von optimaler Ordung [Oswald], unabhängig von der Raumdimension.
Die Ausführungsgeschwindigkeit kann durch adaptive Verfeinerungstechniken, die die Zahl der notwendigen Unbekannten reduzieren, erhöht werden. Dazu existieren vollständige Programmpakete, wie PLTMG [Bank] und Kaskade [Leinen], [Deuflhard, Leinen, Yserentant], die die Ordnung des eingebauten iterativen Lösers durch Gitterverwaltung, Verfeinerung und Gittermanipulation nicht verschlechtern.
Die Ordnung des Lösungsverfahrens kann nur durch parallele Ausführung gesenkt werden. In Hinblick auf sehr große lineare Gleichungssysteme, wie sie insbesondere auch durch Randwertprobleme in drei Raumdimensionen entstehen, liegt es nahe, beide Techniken zu verbinden. Bei der Lastverteilung adaptiv, also dynamisch erzeugter Strukturen, können allerdings nicht mehr

alle Vorraussetzungen an die Finiten-Elemente-Räume und alle Algorithmen zur Gittermanipulation vom sequentiellen Programm übernommen werden.
Existierende Ansätze, wie [Fox & Otto], [Berger & Bokhari] und [Bastian] führen zu Verfahren, deren Ordnung höher als die des iterativen Lösers ist, und nutzen die Multilevel-Struktur der Gitter nicht aus. Ansätze zur Parallelisierung von Standard-Mehrgitterverfahren wie [Briggs, Hart, McCormick & Quinlan] oder von adaptiven Mehrgitterverfahren wie [Mierendorff] können in dieser Form nicht auf adaptive Verfahren angewendet werden, obwohl sie für regulär verfeinerte Gitter optimale Ergebnisse liefern.
Die Verfahren der Gebietszerlegung, wie [Bramble, Pasciak & Schatz], [Bjorstad & Widlund] oder [Bramble, Pasciak, Wang & Xu], bei denen die Parallelisierung die Konvergenzraten beeinflußt, weil die Kopplung der Daten zwischen den einzelnen Prozessoren anders organisiert wird als die Kopplung der Daten innerhalb der einzelnen Prozessoren, konnten bisher noch nicht die Effizienz von den Verfahren erreichen, die die numerischen Eigenschaften des sequentiellen Programms bei der Parallelisierung erhalten.
Wir werden im folgenden Parallelrechner mit verteiltem Speicher und Message-Passing-Kommunikation und Parallelrechner mit gemeinsamem Speicher und Semaphor-Synchronisation verwenden, um ein multilevel-vorkonditioniertes CG-Verfahren so zu implementieren, daß die Eigenschaften des sequentiellen Programms, soweit möglich, erhalten bleiben, und gleichzeitig eine effiziente Parallelisierung erreicht wird.

2. Multilevel-Methoden

Eine Einführung in die im weiteren verwendeten Schreibweisen und Methoden findet sich beispielsweise in [Hackbusch]. Wir führen die Aufgabenstellung, Lösung eines elliptischen Randwertproblems, und die dazugehörigen iterativen Lösungsverfahren ein, die wir in den weiteren Kapiteln dann parallelisieren werden. Zunächst notieren wir, was wir unter dem Modellproblem verstehen wollen.

2.1. Modellproblem und Notation

Gegeben sei folgendes Modellproblem: Es wird eine Näherung u für die Lösung eines elliptischen Randwertproblems zweiter Ordnung auf einem polygonal berandeten Gebiet Ω aus $\Re^2$ gesucht:

$$L\,u = f \qquad \text{in } \Omega$$
$$u = 0 \qquad \text{auf } \partial\Omega$$

mit dem selbstadjungierten linearen skalaren elliptischen Differentialoperator zweiter Ordnung L. Dieser läßt sich schreiben als

$$L\,u = -\sum_{i,l=1}^{2} \frac{\partial}{\partial x_i}\, a_{il}\, \frac{\partial u}{\partial x_l} + au$$

mit uniform positiv definiter symmetrischer Matrix $\{a_{il}(x)\}$ und nicht negativem a(x) für a, $a_{il} \in L^\infty(\Omega)$.

In der variationellen Formulierung des Problems wird bei gegebenem $f \in L^2(\Omega)$ eine Funktion $u \in H_0^1(\Omega)$ gesucht, so daß gilt

(1) $\quad a(u,v) = (f,v) \qquad \forall\, v \in H_0^1(\Omega)$

mit dem L^2-Skalarprodukt (.,.) und dem verallgemeinerten Dirichlet-Integral

$$a(u,v) = \sum_{i,l=1}^{2} \int_\Omega a_{il}\, \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_l}\, dx + \int_\Omega auv\, dx.$$

Für die Diskretisierung mit finiten Elementen konstruieren wir eine Folge von Triangulierungen $\tau_0 \subset \tau_1 \subset \ldots \subset \tau_j$ des Gebietes Ω, für die die Ausgangstriangulierung τ_0 ganz Ω mit Dreiecken überdeckt, und jedes Dreieck mit einem anderen eine Kante, einen Punkt oder nichts gemeinsam hat. Jede weitere Triangulierung entsteht aus der Vorhergehenden durch Zerlegen einzelner Dreiecke in 4 kongruente Teildreiecke oder durch Seitenhalbierung in zwei Dreiecke, die beide im weiteren aber nicht mehr verfeinert werden. Die Mengen der so entstehenden Eckpunkte der Dreiecke im Inneren von Ω bezeichnen wir mit Ω_k .

Wir konstruieren geschachtelte lineare Finite-Elemente-Räume V_k über den Triangulierungen τ_k mit $0 \le k \le j$. In diesen Räumen diskretisiert, schreibt sich (1) über einer Knotenbasis $\{\Phi_i^k\}$ mit gesuchtem $u_k = (u_{k,i})_i$, $1 \le i \le \dim V_k$

(2) $\quad \sum_i a(\Phi_i^k, \Phi_l^k)\, u_{k,i} = (f, \Phi_l^k) \quad 1 \le l \le \dim V_k.$

Mit Hilfe der Steifigkeitsmatrizen

(3) $A_k = (a(\Phi_i^k, \Phi_l^k))_{i,l}$

und der rechten Seiten

$f_k = (f, \Phi_l^k)$

kann man daraus eine Folge von linearen Gleichungssystemen konstruieren. Im folgenden geht es um die effiziente Lösung des Gleichungssystems

(4) $A_j u_j = f_j$.

2.2. Vorkonditionierer

Wir werden nun einige iterative Verfahren zur Lösung von (4) konstruieren, im wesentlichen Vorkonditionierer für CG-Verfahren.
Wir betrachten den diskreten Galerkin-Operator A_k auf V_k, die diskrete L^2-Orthogonalprojektion Q_k von V_j nach V_k und die a-Orthogonalprojektion P_k von V_j nach V_k. Ein Mehrgitter V-Zyklus mit einem Vorglättungsschritt durch den symmetrischen Glätter $S_k = Id - R_k A_k$ läßt sich ausdrücken durch

(5) $M_j^{MG} = \prod_{k=0}^{j} (Id - T_k)$

mit $T_0 := Id - P_0$,

$T_k := (Id - S_k) P_k = R_k A_k P_k = R_k Q_k A_j, \qquad k > 0.$

Unter Vernachlässigung aller Terme höherer Ordnung in T entsteht daraus ein additives Iterationsverfahren

$M_j = Id - \sum_{k=0}^{j} T_k$ [Bramble, Pasciak, Wang & Xu],

das auf der Zerlegung

$u_j = Q_0 u_j + \sum_{k=1}^{j} (Q_k - Q_{k-1}) u_j$

und der Vernachlässigung der Terme Q_{k-1} darin beruht.
Mit $R_0 := A_0^{-1}$ und $R_k := r_k Id$ für $k > 0$ und r_k^{-1} näherungsweise dem Spektralradius ρ von A_k kann man daraus einen Vorkonditionierer für ein CG-Verfahren konstruieren,

(6) $B_j = \sum_{k=0}^{j} R_k Q_k$.

Unter der Vorraussetzung, daß es ein C>0 gibt mit

$\| (Id - Q_{k-1}) v \|^2 \le C \rho(A_k)^{-1} a(v,v) \qquad \forall v \in V_j,$

gilt für zu A_k^{-1} spektral äquivalente R_k

(7) $\kappa(B_j A_j) \le O(j^2)$ [Bramble, Pasciak & Xu].

Unter der stärkeren Regularitätsannahme, daß es ein $C>0$ und ein $\alpha \in (0,1]$ gibt mit

$$a((Id - P_{k-1}) v,v) \le (C \rho(A_k)^{-1} (A_k v)^2)^{\alpha} \; a(v,v)^{1-\alpha} \qquad \forall v \in V_k$$

folgt

(8) $\kappa(B_j A_j) \le O(j^{1/\alpha})$ [Bramble, Pasciak & Xu].

Für quasiuniforme Hierarchien von Triangulierungen gilt sogar

(9) $\kappa(B_j A_j) \le O(1)$ [Oswald].

Mit einer L^2-Orthonormalbasis $\{\Psi_i^k\}$ von V_k ist $Q_k u = \sum_i (u, \Psi_i^k) \Psi_i^k$. Für den praktischen Vorkonditionierer genügt die jeweilige Knotenbasis $\{\Phi_i^k\}$ von V_k. Damit ist

(10) $$B_j^{BPX} u = \sum_{k=0}^{j} \sum_i (u,\Phi_i^k) \Phi_i^k,$$

Die Terme $\{ (u,\Phi_i^k) \}_i$ lassen sich aus $\{ (u,\Phi_i^{k+1}) \}_i$ durch Linearkombinationen gewinnen, genauso wie die Knotenbasisdarstellungen in V_j der Terme $\{ \Phi_i^k \}_i$ aus den Darstellungen der $\{ \Phi_i^{k+1} \}_i$ durch Linearkombinationen folgen. Damit läßt sich die Auswertung von $B_j^{BPX} u$ in $O(\sum_{k=0}^{j} \dim V_k)$ Operationen berechnen.

Im folgenden werden wir teilweise auch noch den Hierarchischen-Basis Vorkonditionierer

(11) $$B_j^{hier} = \sum_{k=0}^{j} \sum_{i \text{ für die } \Phi_i^k \in V_k \setminus V_{k-1}} (u,\Phi_i^k) \Phi_i^k$$

betrachten, für den in zwei Dimensionen

(12) $\kappa(B_j^{hier} A_j) \le O(j^2)$ [Yserentant]

in einer scharfen Abschätzung gilt. Eine Auswertung des hierarchischen Vorkonditionierers kostet $O(\dim V_j)$ Operationen. Für Ω aus $\Re^3$ degeneriert die Abschätzung (12) zu $O(2^j)$. Im Gegensatz dazu bleiben die Abschätzungen (7)-(9) für den BPX-Vorkonditionierer in Räumen höherer Dimension erhalten.

3. Parallelisierung von Multilevel-Methoden

Wir werden im folgenden die Aufteilung der Rechenoperationen zur Auswertung von $B_j^{BPX} u$ und die dabei entstehenden Datenabhängigkeiten und damit die notwendigen Kommunikationsoperationen untersuchen. Das tun

wir zunächst anhand eines theoretischen massiv-parallelen Rechners und dann durch Zusammenfassen einzelner Operationen für reale Rechnerarchitekturen. Wir untersuchen verschiedene Strategien der Parallelisierung und entwickeln dann ein Kostenfunktional, das später zur Optimierung der Aufteilung verwendet wird.
Wir gehen zunächst von einem geometrischen Wachstum der $\{\dim V_k\}$ aus,

(13) $\dim V_k \geq C \dim V_{k-1}$, $C > 1$, für $k>0$.

3.1. Theoretische Parallelrechner

Sei ein theoretischer paralleler Rechner mit genügend vielen Prozessoren gegeben zur Berechnung von $B_j^{BPX}u$. Alle Terme (u,Φ_i^k) können unabhängig voneinander berechnet werden. (u,Φ_i^k) ist Linearkombination von $O(C^k)$ Werten. Dafür muß also auf $O(C^k)$ Prozessoren in Baumanordnung mindestens $O(k \log C)$ Rechenzeit aufgewendet werden.
Die Summe $\sum_{k=0}^{j} \sum_{l} (.,.) \Phi_i^k$ besteht in jedem Gitterpunkt aus höchstens $O(j C)$ Summanden, benötigt also mit $O(j \dim V_j)$ Prozessoren $O(\log j)$ Rechenzeit, bzw. mit $O(\dim V_j)$ Prozessoren $O(j)$ Zeit.
Insgesamt ist also durch massive Parallelisierung mit dem Einsatz von $O(\dim V_j)$ Prozessoren eine Reduktion der Ordnung von

(14) $O(\dim V_j)$ auf $O(j)$

erreichbar, das geometrische Wachstum der $\{\dim V_k\}$ vorausgesetzt. Die gleiche Ordnung ist auch bei einem Mehrgitter V-Zyklus mit lokalem, vollständig parallelisierbarem Vorglätter, wie z.B. einer Schachbrett-Gauß-Seidel-Iteration, und auch bei Anwendung des hierarchischen Basis-Vorkonditionierers erreichbar.
Liegt schwächeres Wachstum der $\{\dim V_k\}$ vor, dann kann unter Einsatz von $O(j \dim V_j)$ Prozessoren der kleinere Wert

(15) $O(\log (j \dim V_j))$

für den BPX-Vorkonditionierer und den hierarchischen Basis Vorkonditionierer erreicht werden. Für den Mehrgitter V-Zyklus ist eine solche Verbesserung von $O(j)$ auf einen kleineren Wert im allgemeinen nicht zu erwarten, da der Glätter für den Wert an einem Gitterpunkt stets Nachbarwerte

braucht, die aber nach obiger Reihenfolge der Berechnung teilweise erst später berechnet werden.

3.2. Ansätze für Parallelisierungen

Wir betrachtem im folgenden Ansätze, diese Parallelität auf realen Parallelrechnern nutzbar zu machen.

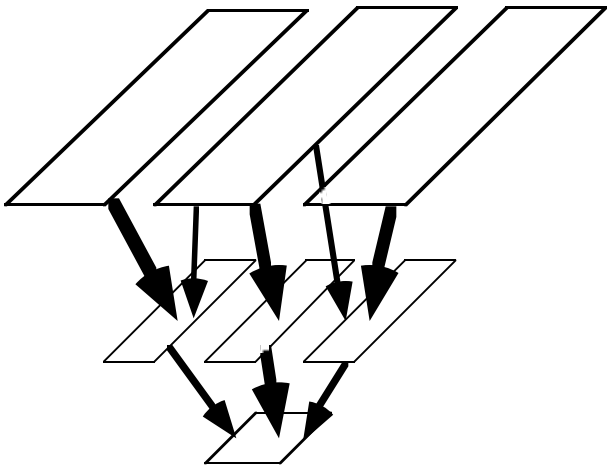

Abb 3.2.1. *Aufteilung von Gittern mit Datenabhängigkeiten*

Die Abbildung 3.2.1 zeigt eine mögliche Aufteilung von drei Verfeinerungsebenen und die möglichen Datenabhängigkeiten.
Bei realen MIMD-Rechnern hat man eine konstante Zahl von Prozessoren zur Verfügung, die im allgemeinen wesentlich kleiner als die Zahl der Unbekannten ist. Man muß also Rechenoperationen, die theoretisch parallel durchgeführt werden könnten, zu Gruppen zusammenfassen und einem Prozessor zuordnen. Damit wird es wichtig, gleiche Operationen nicht mehrfach durchzuführen, was bei genügend Prozessoren kein Problem war.
Wir stellen uns eine Auswertung des Vorkonditionierers in einer sequentiellen mehrgitterartigen Implementierung mit Restriktionen, Lösung auf dem gröbsten Gitter und Prolongationen vor. Regelmäßige Aufteilungen können sowohl elementorientiert, als auch punktorientiert durchgeführt werden. Dabei werden die Rechenoperationen und Datenabhängigkeiten so aufgeteilt, daß entweder Dreiecke oder Gitterpunkte immer auf einem gemeinsamen Prozessor berechnet werden.

In der Abbildung 3.2.2 ist eine elementorientierte Aufteilung eingezeichnet, wobei die Zeilen drei verschiedene Verfeinerungsstufen, von fein nach grob darstellen. Die Punkte sind als Kugeln, die Datenabhängigkeiten sind als Linien gekennzeichnet.

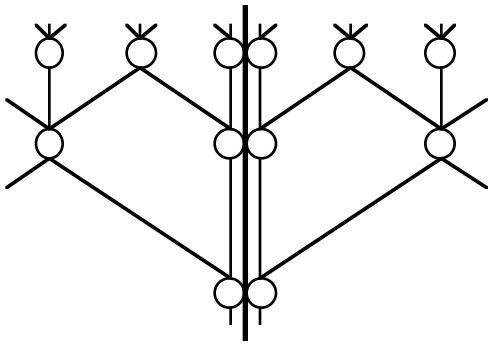

Abb 3.2.2. *Aufteilung von Elementen*

Die Aufteilung erfolgt entlang dem senkrechten Balken. Damit sind alle Punkte entlang der Trennungslinie doppelt vorhanden, wobei sie jeweils nur Teilergebnisse enthalten und nicht, wie die inneren Punkte, konsistent sein müssen. Nach einem vollständigen Restriktionsschritt über alle Ebenen müssen für die Grobgitterlösung die Teilergebnisse an den Kanten zusammengefaßt werden, was einer lokalen Kommunikation entspricht. Damit werden auch diese Werte konsistent und stimmen auf beiden Nachbarprozessoren überein. Die Prolongationen anschließend arbeiten wieder mit inkonsistenten Rändern, die erst in einem abschließenden lokalen Kommunikationsschritt zusammengefaßt werden.
Eine Variante der Aufteilung ist die zeitliche Hintereinanderausführung der Operationen zweier Gebiete mit gemeinsamer Kante [Leinen]. Dabei kann eine Addition je Punkt beim Zusammenfassen durch den lokalen Kommunikationsschritt eingespart werden, wobei allerdings eine Färbung der einzelnen Gebiete und die zeitliche Hintereinanderausführung der verschieden gefärbten Gebiete nötig wird. Weiterhin wird die Aufteilung in mindestens doppelt so viele (zwei Farben) Teile wie Prozessoren nötig, was die Lastverteilung behindern kann.
Eine Auswertung kann mit zwei lokalen Kommunikationsschritten durchgeführt werden. Das ist aber nur so lange richtig, wie die Aufteilung in

Elemente bereits auf der gröbsten Triangulierung stattfindet. Für eine gute Lastverteilung der Prozessoren und damit große Beschleunigung durch die Parallelisierung ist aber im allgemeinen eine Aufteilung erst auf einem feineren Gitter möglich, was die Kommunikationsstruktur komplizierter macht.
Dazu betrachten wir nun die andere regelmäßige Aufteilung, die Verteilung der Gitterpunkte. In der Abbildung 3.2.3 ist die gleiche Aufteilung nur dieses Mal für Punkte gezeichnet.

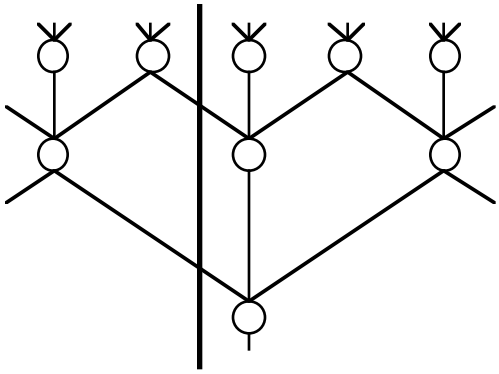

Abb 3.2.3. *Aufteilung von Gitterpunkten*

Die Werte der Gitterpunkte sind in jedem Schritt konsistent, weil sie nur genau einmal vorhanden sind. In jedem Restriktions- und Prolongationsschritt von einer zur nächsten Verfeinerungsebene muß jeweils eine lokale Kommunikation stattfinden, in diesem Fall allerdings nur in einer Richtung. Der senkrechte Balken der Aufteilung auf die Prozessoren kann überall verlaufen, stets ist eine lokale Kommunikation in genau einer Richtung nötig.
Damit ist diese Aufteilung für eine gute Lastverteilung günstiger als die Verteilung der Elemente der gröbsten Ebene, weil die Punkte der feinsten Ebene frei verteilt werden können. Die Menge der zu transferierenden Daten ist bei beiden Methoden der Zerlegung gleich, nur die Zahl der Kommunikationsoperationen unterscheidet sich.
Für eine gute Lastverteilung, bei der eine bestimmte Effizienz der Parallelsierung erreicht werden soll, wird also zunächst eine Aufteilung und Kommunikation wie für Gitterpunkte durchgeführt, um ab einer festgelegten Verfeinerungsebene eine Aufteilung in Elemente zu verwenden und damit Kommunikationsoperationen zu sparen.

Wenn im folgenden von der Aufteilung von Gitterpunkten die Rede sein wird, so ist damit die Aufteilung der auf den gröberen Gittern als Punkte aufzufassenden Elemente und Hierarchien von Elementen auf dieser festgelegten Ebene zu verstehen.
Wie diese Ebene festgelegt werden muß, wird in Kapitel 6.1. besprochen.
Für die Parallelisierung von Mehrgitterverfahren ist eine Aufteilung in Elemente im allgemeinen nicht sinnvoll, da der Glätter konsistente Werte einer Gitterebene benötigt. Damit ist für jede Gitterebene ein lokaler Kommunikationsschritt nötig und eine Parallelisierung mittles Aufteilung der Gitterpunkte effizienter. Hier wird also bis zur feinsten Ebene punktorientiert verteilt.
Der im weiteren häufig verwendete Begriff der lokalen Kommunikation soll andeuten, daß jeder Prozessor nur mit einer beschränkten Zahl von Nachbarn Informationen austauscht. Das ist insbesondere für Abschätzungen für große Prozessorzahlen wichtig. Die Struktur der Prozessoren und ihrer Nachbarn wird auch als Prozessortopologie bezeichnet, wobei diese logische Struktur mit der physikalischen Struktur der Verknüpfungen der Prozessoren zusammenhängen sollte, um effizient kommunizieren zu können. Diese Strukturen sind sehr stark hardwareabhängig. In einem zweidimensionalen Netz von Prozessoren für ein Randwertproblem in zwei Raumdimensionen sollte jeder Prozessor mit mindestens 6 Nachbarn logisch verbunden sein, was aus der Struktur von Restriktion und Prolongation für innere Punkte hervorgeht.

Im folgenden gehen wir von einer mehrgitterartigen Implementierung von $B_j^{BPX}u$ oder $B_j^{hier}u$ sowie einer Aufteilung der Gitterpunkte auf P Prozessoren aus.

3.3. Das Kostenfunktional für Punktverteilungen

Wir betrachten für die Konstruktion des Kostenfunktionals eine Aufteilung der Gitterpunkte, die später durch die Minimierung des Funktionals optimiert werden soll. Jedes Gitter Ω_k wird also auf die P Prozessoren verteilt. Jeder Prozessor p erhält das Teilgebiet D_k^p aus Ω_k. Wir betrachten Aufteilungen mit

(16) $\bigcup_p D_k^p = \Omega_k$, die zunächst disjunkt seien,

(17) $D_k^i \cap D_k^j = \varnothing$ für $i \neq j$, woraus
$\sum_{p=1}^{P} \#D_k^p = \#\Omega_k$ folgt.

Wir stellen ein Kostenfunktional für einen Modell-Parallelrechner auf, das die Zahl der wesentlichen Rechenoperationen und die Menge der Datenübertragungsoperationen bzw. die verlangsamende Benutzung des gemeinsamen Speichers, je nach Rechnerarchitektur, berücksichtigt und eine obere Schranke für die Rechenzeiten liefert. Im wesentlichen vernachlässigt werden dabei Einsparungen durch Überlappung von Rechnung und Kommunikation. Seien

(18) $\mathrm{supp}(Op \mid D) := \bigcup \{ \mathrm{supp}(f) ;$ mit $\mathrm{supp}(Op(f))$ aus D und f aus Definitionsbereich von Op $\}$

die notwendigen Argumente für den auf den Bildbereich D´ eingeschränkten Operator Op. Eine Restriktion mit dem Operator r von Ω_k nach Ω_{k-1} kostet dann

(19) $W_k^R = \max_p \left\{ \#D_{k-1}^p \, T_R + \#(\mathrm{supp}(r \mid D_{k-1}^p) \setminus D_k^p) \, C_R \right\}$

mit einem Maß T_R für die wesentlichen Rechenoperationen und einem Maß C_R für Kommunikationsoperationen oder die Verzögerung bei Zugriffen auf den gemeinsamen Speicher. Die Prolongation mit $p = r^*$ von Ω_{k-1} nach Ω_k kostet entsprechend

(20) $W_k^P = \max_p \left\{ \#D_k^p \, T_P + \#(\mathrm{supp}(r^* \mid D_k^p) \setminus D_{k-1}^p) \, C_P \right\}.$

Die exakte Lösung des Problems in V_0 kostet, mittels Cholesky-Zerlegung oder eines anderen direkten Lösers auf einem Prozessor mit maximalem $\#D_0^p$ durchgeführt,

$$(21) \qquad W_0^E = (\#\Omega_0)^2\, T_E + (\,\#\Omega_0 - \max_p \#D_0^p\,)\, C_E.$$

Eine Anwendung von A_k mit dem diskreten a-Operator auf dem Gitter Ω_k kostet

$$(22) \qquad W_k^A = \max_p \left\{ \#D_k^p\, T_A + \#(\,\mathrm{supp}(a \mid D_k^p) \setminus D_k^p\,)\, C_A \right\}.$$

Die für den Mehrgitter V-Zyklus verwendeten Glätter S_k auf Ω_k kosten, sofern sie lokal und vollständig parallelisierbar mittels Färbung {Col} der Gitterpunkte sind,

$$(23) \qquad W_k^S = \sum_{\text{Col aus } \Omega_k} \max_p \left\{ \#(D_k^p \cap \mathrm{Col})\, T_S + \#(\,\mathrm{supp}(S \mid D_k^p \cap \mathrm{Col}) \setminus D_k^p\,)\, C_S \right\}.$$

Für das Schachbrett-Gauß-Seidel Verfahren ist {Col} beispielsweise gleich { {schwarze Punkte von Ω_k}, {rote Punkte von Ω_k} }. Die Vektoradditionen innerhalb des CG-Verfahrens bzw. des Mehrgitterverfahrens können ohne Kommunikation durchgeführt werden und kosten deshalb

$$(24) \qquad W_k^V = \max_p \{\#D_k^p\}\, T_V.$$

Damit ergibt sich das Kostenfunktional für einen Mehrgitter V-Zyklus zu

$$(25) \qquad W_j^{MG\text{-}V} = \sum_{k=1}^{j} (\, W_k^R + W_k^A + W_k^V + W_k^S + W_k^P\,) + W_0^E$$

Ein Skalarprodukt für das umgebende CG-Verfahren ist mit

$$(26) \qquad W_k^{SC} = \max_p \{\#D_k^p\}\, T_{SC} + (\log P)\, C_{SC}$$

anzusetzen, wenn die Teilergebnisse in einer Baumstruktur zusammengefaßt werden. Dies ist der einzige Teil des Verfahrens, in dem wegen der unterschiedlichen Anordnung der Teilsummen die sequentielle und die parallele Implementierung verschiedene numerische Ergebnisse liefern können.

Wenn die Prozessoren logisch als Hypercube vernetzt sind und jeder gleichzeitig Senden und Empfangen kann, so genügen $\log_2 P$ Kommunikationsschritte um das Ergebnis parallel zu berechnen und zu

verteilen. Wenn Sende- und Empfangsoperationen hintereinander ausgeführt werden müssen, kann man einen binären Baum zum Berechnen des Skalarproduktes und denselben Baum in umgekehrter Richtung zum Verteilen des Ergebnisses verwenden, was zur doppelten Zahl von Kommunikationsschritten führt. Damit kostet insgesamt ein BPX-vorkonditionierter CG-Schritt

$$(27) \qquad W_j^{CG} = \sum_{k=1}^{j} (W_k^R + W_k^P) + W_0^E + W_j^A + 2\, W_j^{SC} + W_j^V .$$

Um damit einen mit der hierarchischen Basis vorkonditionierten CG-Schritt abschätzen zu können, ersetzen wir in den Definitionen von W_k^R und W_k^P alle Terme der Form D_i^p durch $(D_i^p \setminus D_{i-1}^p)$ mit der Konvention $D_{-1}^p = \varnothing$.

3.4. Abgeleitete Kostenfunktionale

Wir werden nun einige Varianten des Kostenfunktionals diskutieren, die das Verhalten bestimmter Parallelrechner besser modellieren.

Für Parallelrechner mit gemeinsamem Speicher spielen die Konstanten C die Rolle von Verlangsamungen durch Benutzung des gemeinsamen Speichers und den daraus entstehenden Speicherzugriffskonflikten. Sie werden auch teilweise mit dem allgemeineren Begriff der Speicherbankkonflikte bezeichnet, die aber auch bei ungünstiger Ausrichtung der Daten im Speicher auftreten. Der bremsende Einfluß ist proportional zur Größe der Ränder der $\{ D_k^p \}$ und teilweise auch proportional zu P.

Für Parallelrechner mit verteiltem Speicher sind die Konstanten C die Zeit, die für Sende- und Empfangsoperationen für die Randdaten benötigt wird. Die reinen Übertragungskosten sind, von Einflüssen der kleinsten Übertragungseinheiten abgesehen, proportional zur Menge der Daten. Meistens sind aber auf realen Parallelrechnern die Initialisierungskosten höher als die auftretenden eigentlichen Übertragungskosten, so daß das Modell durch Ersetzen der Zahl der Randpunkte der Form

$$\#(\mathrm{supp}(Op \,|\, D_i^p) \setminus D_j^p)$$

durch die Zahl der Nachbarprozessoren $\#N_{j,Op}^p$ mit

$$(28) \qquad N_{j,Op}^p := \left\{ p' : \text{mit } (\mathrm{supp}(Op \,|\, D_i^p) \setminus D_j^p) \cap D_j^{p'} \neq \varnothing \right\},$$

die für eine lokale Kommunikationsstruktur beschränkt ist, mit anderen C realistischer wird. Diese Zahl der Nachbarprozessoren wird für ein

zweidimensionales Prozessornetz mindestens 6 sein. Wenn die Zerlegung der Ω_k nicht disjunkt ist, also

(29) $D_k^i \cap D_k^j = \emptyset$ für $i \neq j$

verletzt ist, bleiben die obigen Audrücke weiterhin gültig. Das könnte insbesondere dann von Interesse sein, wenn auf den gröbsten Ω_k einige Werte schneller berechnet als gesendet und empfangen werden können. Dazu müssen natürlich dann auch die dazu nötigen Randdaten schnell verfügbar sein, was im allgemeinen aber nicht der Fall ist.

4. Ein Beispielprogramm

Bevor wir versuchen, die allgemeinen Kostenfunktionale W^{MG} und W^{CG} durch geeignete $\{D_k^p\}$ zu minimieren, zunächst ein einfaches Beispiel für ein Randwertproblem und eine Folge von Gittern. Es wird eine bezüglich einfachen Kriterien optimale Aufteilung der Gitterpunkte und eine Näherung für das Kostenfunktional konstruiert, die in diesem Spezialfall das Laufzeitverhalten gut beschreibt.

4.1. Formulierung

Wir betrachten eine quasiuniforme regulär verfeinerte Triangulierung des Einheitsquadrates. Die Gebiete $\{D_k^p\}$ seien Rechtecke der Kantenlänge des Quadrates. Die Prozessoren sind also in einer Kette angeordnet, jeder Prozessor p mit $1<p<P$ hat die Nachbarn p-1 und p+1. Wir verwenden ein BPX-vorkonditioniertes CG-Verfahren, der diskrete Differentialoperator sei durch einen 3×3 Stern charakterisiert. Jede Triangulierung entsteht durch die Zerlegung jedes Dreiecks der vorhergehenden Triangulierung in jeweils 4 kongruente Teildreiecke. Die Triangulierung auf Ebene 0 entsteht dabei aus der Zerlegung eines in zwei Dreiecke geteilten Quadrates. Damit enthält Ω_0 eine Unbekannte. Alle Ω_k können als quadratische Gitter aufgefaßt werden, für die sich dann Restriktion und Prolongation als 3×3 Sterne ergeben.

4.2. Parallelisierung

Die Aufgabe, die rechteckigen $\{D_k^p\}$ zu finden, reduziert sich, wenn wir die Ränder mitzählen, auf die Aufteilung der Zeilen $\{0 ,\dots ,2^{k+1}\}$ von Ω_k in P Intervalle $\{d_k^p\}$ für jede Ebene k. Um die Zwischengitterkommunikation aus W_k^R und W_k^P gering zu halten, soll zwischen p und p+1 für Restriktion und Prolongation jeweils nur eine Zeile in genau einer Richtung übertragen werden. Damit folgt

(30) $2n \in d_k^p \Rightarrow n \in d_{k-1}^p$ für $k>0$,

und damit wird die Kommunikation von Restriktion und Prolongation symmetrisch. Es bleibt noch die Aufteilung des feinsten Gitters Ω_j in $\{d_j^p\}$ zu finden. Zur Minimierung von W_j^A und damit W^{CG} muß gelten

(31) $$\#d_j^p \in \left(\frac{2^{j+1}+1}{p} - 1 \ , \frac{2^{j+1}+1}{p} + 1 \right).$$

Falls $\frac{2^{j+1}+1}{p}$ keine ganze Zahl ist, dann bleiben noch Freiheitsgrade, die aufgewendet werden sollten für

(32) $\#d_j^1 \geq \#d_j^n \ \forall n$, und $\#d_j^P \geq \#d_j^n \ \forall n>1$,

weil am Rand etwas weniger Arbeit für die Prozessoren sowohl an Kommunikation als auch an Rechenoperationen für die Randpunkte des Gebietes anfallen. Die übrigen überschüssigen Zeilen können gleichmäßig auf die inneren Prozessoren verteilt werden.
Aus (30) folgt, daß die Eigenschaften (31) und (32) für alle $\{d_k^p\}$ mit $k \leq j$ gelten.
Damit wird W^{CG} unter allen rechteckigen $\{D_k^p\}$ der Länge Eins minimiert, sobald $k > k_0$, das von P und C_{eff} / T_{eff} abhängt. Für kleinere k ist es sinnvoll, einige $d_k^p = \varnothing$ zu wählen.

4.3. Ergebnisse

Wir werden nun die entstehenden Parallelisierungen analysieren, sowohl die Aufteilungen der Gitter als auch die Werte für Beschleunigung und Effizienz, die in einem späteren Kapitel mit experimentellen Ergebnissen verglichen werden.
Es folgt ein Beispiel für eine solche Aufteilung der Gebiete. Die Zeilen der Gitterstufen 0 bis 7 werden auf 9 Prozessoren verteilt. Zusätzlich in der Abbildung 4.3.1 eingezeichnet sind alle nötigen Kommunikationsoperationen

für eine Auswertung von $B_7^{BPX}u$. Ein Pfeil zeigt dabei den Transfer einer Randzeile an. Deutlich zu sehen ist die Symmetrie von Restriktion und Prolongation und der Leerlauf einiger Prozessoren auf den beiden gröbsten Gitterebenen.

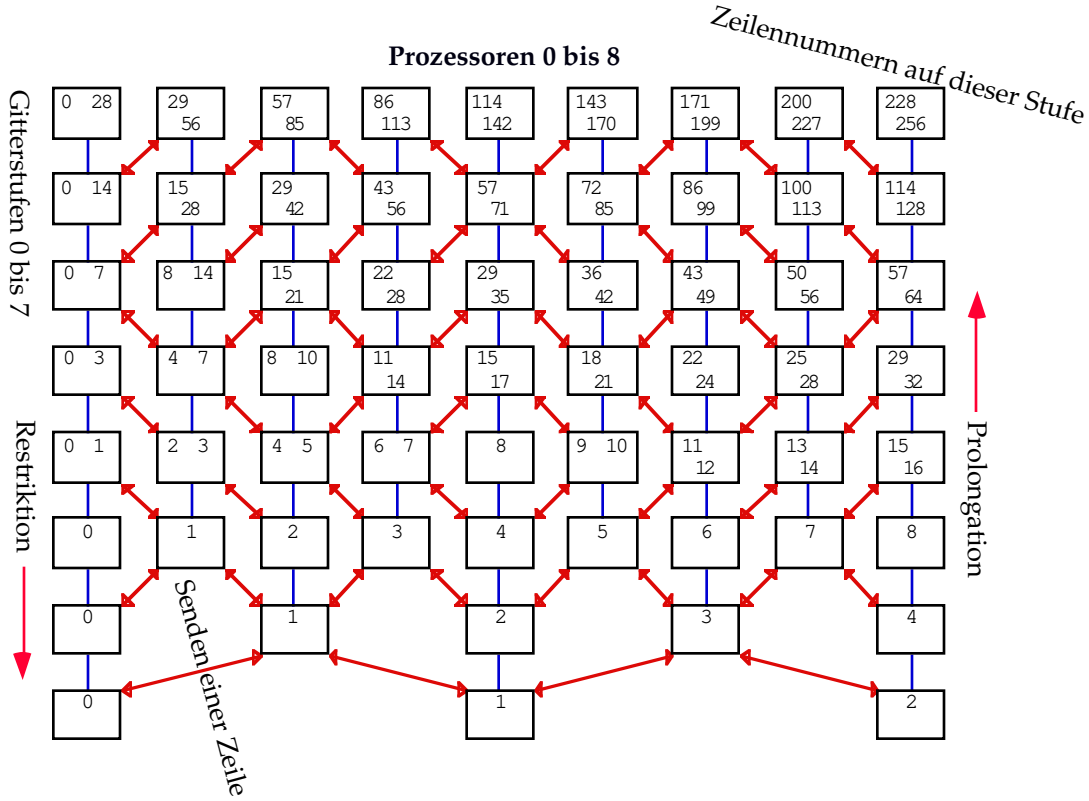

Abb 4.3.1. *Kommunikation für eine Auswertung desVorkonditionierers*

Zum Vergleich dazu ist in Abbildung 4.3.2 die Kommunikationsstruktur für die Auswertung eines Skalarproduktes mit 9 Prozessoren und der Hilfe eines binären Baumes dargestellt. Auch hier ist der Leerlauf einiger Prozessoren während der Operation zu beobachten. Die einzelnen Zeilen deuten die Zeitschritte an, in denen die Kommunikation abläuft. Das Zusammenfassen der einzelnen Teilergebnisse auf einem Prozessor muß dabei nicht immer in der angegebenen Reihenfolge stattfinden und kann bei entsprechender Programmierung zu einem nichtdeterministischen Einfluß von Rundungsfehlern führen.

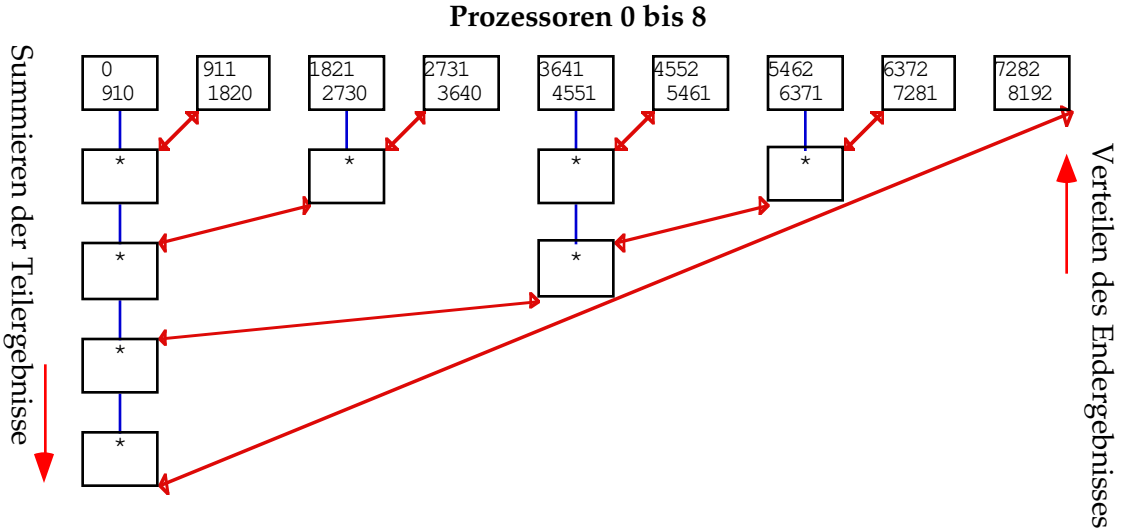

Abb 4.3.2. *Kommunikation für das Skalarprodukt mit binärem Baum*

Für einen CG-Schritt sind auf einem Rechner mit verteiltem Speicher $(2\,j + 2 + 4\log_2 P)$ Sende- und Empfangsoperationen nötig. Es kann asymptotisch eine Geschwindigkeitssteigerung um

$$(33) \qquad Sp^{CG}_{distr} = \frac{P}{1 + P(\,1 + j + 2\log_2 P)\,4^{-j}\,C_{eff} / T_{eff}}$$

und damit eine Effizienz

$$(34) \qquad Eff^{CG}_{distr} = (1 + P(\,1 + j + 2\log_2 P)\,4^{-j}\,C_{eff} / T_{eff})^{-1}$$

erreicht werden. Die höchste Ausführungsgeschwindigkeit ergibt sich damit für eine Rechnung mit

$$(35) \qquad P^{CG}_{opt} = 4^j \ln 2 \;\; T_{eff} / C_{eff}$$

Prozessoren. Für einen Rechner mit gemeinsamem Speicher ergeben sich die Werte zu

$$(36) \qquad Sp^{CG}_{shared} = \frac{P}{1 + P\,2^{-j}\,C_{eff} / T_{eff}} \quad \text{und}$$

$$(37) \qquad Eff^{CG}_{shared} = (1 + P\,2^{-j}\,C_{eff} / T_{eff})^{-1}$$

für kleine P, solange der Einfluß von $(\log_2 P)$ der Skalarprodukte vernachlässigbar bleibt, was bei realen Rechnern mit gemeinsamem Speicher aber stets gewährleistet ist. Der Term $(P\,2^{-j})$ beschreibt dabei den Anteil der Ränder an $\{D^p_k\}$. Diese Werte, genauso wie die Werte der genaueren Modelle für W^{CG}_j konnten in Experimenten mit guter Übereinstimmung gemessen werden, wie auch die Ergebnisse am Ende dieses Textes zeigen.

5. Adaptive parallele Multilevel-Methoden

Wir werden nun die übrigen Komponenten eines Finite-Elemente Programms und deren Parallelisierung, in das der iterative Löser eingebettet ist, analysieren.
Das gesamte Lösungsverfahren besteht aus

- Fehlerschätzer,
- Gittermanipulation und
- iterativem Löser.

Die Verwaltung der Gitter erzeugt die Datenstrukturen, mit denen der Löser arbeitet, und sollte nur einen geringen Einfluß auf den Gesamtaufwand des Verfahrens haben. Daher ist eine gute Parallelsierung dieser Komponenten für Rechner mit großen Prozessorzahlen wichtig.

5.1. Fehlerschätzer

Wir betrachten nun die Parallelisierung von Fehlerschätzern.
Das Finite-Elemente-Verfahren startet mit einer vorgegebenen Anfangstriangulierung τ_0. Die feineren Triangulierungen τ_k werden adaptiv erzeugt. Üblich sind lokale Fehlerschätzer, die einzelne Punkte, Kanten oder Elemente markieren, die dann anschließend verfeinert werden. Diese lokalen Fehlerschätzer arbeiten mit lokalen Schätzungen für Ableitungen, Fehlerentwicklungen [Bank, Weiser] oder Näherungslösungen in einem quadratischen Finite-Elemente-Ansatzraum [Deuflhard, Leinen, Yserentant] oder feineren linearen FE-Räumen. Alle genannten lokalen Fehlerschätzer sind mit ihrer lokalen Kommunikationsstruktur ähnlich einem Schritt des CG-Verfahrens parallelisierbar.
Um aus dem Fehlerschätzer einen Fehlerindikator zu gewinnen, der zu verfeinernde Punkte, Kanten oder Elemente markiert, wird häufig der lokale geschätzte Fehler mit einer Kombination aus globalem Maximum und Mittelwert aller geschätzten Fehler verglichen. Das erfordert eine globale Kommunikation, die die gleiche Komplexität O(log P) besitzt wie ein Skalarprodukt.
Insgesamt gilt damit $O(W^{Fehler}) \leq O(W^{CG})$.

5.2. Gittermanipulation

Die Erzeugung der neuen Gitter ist im allgemeinen nicht so problemlos zu parallelsieren. Wir werden also auch Änderungen an existierenden Finite-Elemente-Programmen in Betracht ziehen müssen.
Im Gittermanipulationsteil muß eine gültige neue Triangulierung τ_j erzeugt werden. Zweckmäßigerweise wird dabei für die statische Aufteilung der Gebiete auch die Verteilung der neuen und die Umverteilung der alten Elemente $\{D_j^p\}$ durchgeführt.
Legt man die Verfeinerungsregeln von [Bank] für die Gitterverfeinerung um eine Stufe zugrunde, so werden Dreiecke mit zwei zu verfeinernden Kanten in 4 (kongruente) Dreiecke, Dreiecke mit nur einer zu verfeinernden Kante in 2 Dreiecke zerlegt (grüne Triangulierung). Auf einem sequentiellen Rechner läßt sich bei geeigneter lokaler Programmierung die neue Triangulierung τ_{j+1} in $O(\#\tau_{j+1} - \#\tau_j)$ erzeugen [Leinen].
Auf einem parallelen Rechner mit verteiltem Speicher kann dagegen folgende Situation eintreten:

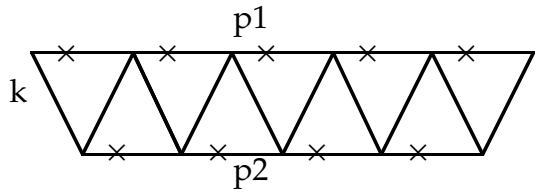

Abb 5.2. *Triangulierung mit markierten Kanten*

Die zu verfeinernden Kanten sind in der Abbildung 5.2 mit „×“ gekennzeichnet. Abhängig von der Verfeinerung der Kante k werden damit alle abgebildeten Dreiecke halbiert oder in 4 kongruente Dreiecke zerlegt. Dazu müssen zwischen den Prozessoren p_1 und p_2 im schlimmsten Fall $\#\{D_j^{p1} \cap D_j^{p2}\}$ Botschaften ausgetauscht werden.
Auf Parallelrechnern mit gemeinsamem Speicher müssen entsprechend viele Synchronisierungen durchgeführt werden. Die gleichen Probleme entstehen auch mit der Strategie, Dreiecke mit markierter Kante als vollständig zu verfeinernd zu markieren und darauf Algorithmen zum regelkonformen

Abschluß der Triangulierung anzuwenden. Die grünen Triangulierungen führen im ungünstigsten Fall zu

$$(38) \qquad W_j^{Gitter} \geq \max_p \left\{ \#D_{j+1}^p \, T_G + \#(\, supp(r \mid D_j^p) \setminus D_{j+1}^p \,) \, C_G^{init} \right\}$$

mit den Initialisierungskosten C_G^{init} für eine Kommunikation, was auf realen Parallelrechnern um Ordnungen größer ist als der Kommunikationsaufwand von W_j^{CG}. Es muß daher nach anderen Verfahren gesucht werden, so daß nicht der Gitterverwaltungsteil die eigentliche Rechenzeit dominiert. Eine Möglichkeit, $W_j^{Gitter} << W_j^{CG}$ zu erhalten, ist, auf die grüne Triangulierung zu verzichten und statt dessen alle Verfeinerungen von Dreiecken regulär durchzuführen. Dabei können gebundene innere Knoten entstehen, deren Werte durch Interpolation bestimmt werden und die keinen Freiheitsgrad darstellen. Damit kann sehr einfach

$$(39) \qquad W_j^{Gitter} = \max_p \left\{ \#D_{j+1}^p \, T_G + \#(\, supp(r \mid D_j^p) \setminus D_{j+1}^p \,) \, C_G^{data} \right\}$$

erreicht werden, mit den Kosten C_G^{data}, um Daten über die Randelemente den Nachbarprozessoren bekannt zu machen. Für Rechner mit verteiltem Speicher kann dies mit einem Satz von lokalen Sendeoperationen durchgeführt werden.

6. Lastverteilung adaptiver paralleler Multilevel-Methoden

Wir kommen nun zu verschiedenen Strategien, für allgemeinere Gebiete $\{\Omega_k\}$ gute Aufteilungen in $\{D_k^p\}$ zu finden, die die Ausführungszeiten des iterativen Lösers minimieren sollen. Dabei werden wir dynamische Scheduling-Algorithmen mit Verfahren für statische Aufteilungen mit den Mitteln der kombinatorischen Optimierung vergleichen. Wir werden eine Klasse von Gitterverfeinerungen konstruieren, die eine effiziente Verteilung auf mehr als eine beschränkte Zahl von Prozessoren unmöglich macht. Daraus leiten wir für große Prozessorzahlen Bedingungen an das Wachstum der Dimensionen der Finite-Elemente-Räume ab. Wenn diese nicht erfüllt werden, können wir bezüglich schwächerer Kriterien mit einigen Erweiterungen der Aufteilungsverfahren immer noch gute Aufteilungen erhalten. Zuletzt werden wir noch ein neues Lastverteilungsverfahren konstruieren, dessen Komplexität im Gegensatz zu den anderen vorgestellten Verfahren nicht größer der des

iterativen Lösers ist, das sich aber an die Strukturen verschiedener Parallelrechner und Löser anpasst.

6.1. Dynamische Ansätze

Lastverteilungsverfahren, die die Aufteilung der Punkte dynamisch während der iterativen Lösung ständig neu erzeugen, beispielsweise durch Scheduling kleiner Gitterteile [Leinen], sind für große Prozessorzahlen P ungeeignet, da der Scheduling-Prozeß die Lokalität der Kommunikationsstruktur durchbricht. Es muß also einen zentralen Prozeß geben, der mit allen anderen kommuniziert. Die Gitterteile müssen für den Mehrgitter V-Zyklus im allgemeinen für jede Ebene k getrennt bearbeitet werden. Dagegen können für die Vorkonditionierer auch ganze Hierarchien von Elementen zu einem Gitterteil zusammengefaßt werden, so daß für eine grobe Lastaufteilung teilweise schon eine Aufteilung in die Elemente der gröbsten Anfangstriangulierung genügt. Feinere Aufteilungen können auf einer der Prozessorzahl P angemessenen feineren Triangulierung durchgeführt werden, wobei auf den gröberen Gittern ein anderes Verteilungsverfahren zu Einsatz kommt.
Um die globale Kommunikation und den damit verbundenen zentralen Scheduler zu umgehen, bieten sich Hierarchien von Schedulern an, wie sie z.B. als Loadbalancer in CDL [Veer] üblich sind. Auf Parallelrechnern mit verteiltem Speicher ist ein solches Verfahren ungünstig, weil die gesamten Daten des Problems über mehrere Zwischenstufen transportiert werden müssen, was den zentralen Scheduler-Prozeß nach wie vor zum Flaschenhals macht. Im übrigen müssen alle Daten im Speicher eines Prozessors Platz finden, was dem Prinzip der Datenparallelität widerspricht und was bei großen Problemen nicht der Fall sein muß.
Für Rechner mit gemeinsamem Speicher, für die P naturgemäß beschränkt ist und kein Datentransfer nötig ist, sondern nur Steuerinformation über die Gitterhierarchien, ist dieses Verfahren aber durchaus möglich und liefert, abhängig von der Granularität der Unterteilung in Gitterteile, verschieden gute Werte. So wird man die Gitterebene k, auf der unterteilt wird, in Abhänigigkeit von P und der zu erzielenden Effizienz wählen, so daß für die Zahl der

Elemente der Triangulierung der Ebene k, oder in einer Näherung die Zahl der Gitterpunkte $\#\Omega_k$ der Ebene k, ungefähr gilt

(40) $\#\tau_k \geq \frac{P}{1 - \text{Effizienz}}$ mit der Effizienz <1.

6.2. Der schlechteste Fall für statische Ansätze

Nachdem für Rechner mit verteiltem Speicher realistischerweise nur noch eine statische Aufteilung der Gitter in Frage kommt, wollen wir untersuchen, ob das unter den bisher gemachten Voraussetzungen überhaupt sinnvoll ist.
Es bleibt also die Alternative einer statischen Aufteilung für eine Folge von Triangulierungen $\{ \tau_k \}_{k=0,\dots,j}$. Wird durch Verfeinerung ein neues τ_{j+1} erzeugt, so muß ein neuer Satz von $\{ \overline{D}^p_k \}_{k=0,\dots,j+1}$ erzeugt werden, der W_{j+1} näherungsweise minimiert. Für Rechner mit verteiltem Speicher und große P sind für gute Abschätzungen des Kostenfunktionals die beschränkten Zahlen der Nachbarprozessoren $\{\#N^P_{k,Op}\}$ mit

(41) $N^p_{j,Op} := \left\{ p' : \text{mit } (\text{supp}(Op \mid D^p_i) \setminus D^p_j) \cap D^{p'}_j \neq \varnothing \right\}$,
$\#N^P_{k,Op} \leq N_{Op} \quad \forall p,k$

wichtig. Es liegt also eine lokale Kommunikationsstruktur vor. Sei diese im folgenden vorausgesetzt.

6.2.1. Das Problem

Wir betrachten folgende Verfeinerung: Alle Dreiecke von $D_j^{\overline{p}}$ werden verfeinert, die übrigen Dreiecke von Ω_j dagegen nicht.

6.2.2. Ansätze

(i) Ist $\overline{D}^p_k = D^p_k$ für alle $k \leq j$, so kann kann Ω_{j+1} höchstens auf N_r+1 Prozessoren verteilt werden. Für jeden dieser Prozessoren gilt $\#\overline{D}^p_{j+1} \approx \frac{3\,\#D_j^{\overline{p}}}{N_r+1} + \#\overline{D}^p_j$. Mit Wiederholung dieser Verfeinerung, erhält man für $j \to \infty$ eine Aufteilung der Last auf höchstens N_r+1 Prozessoren, also eine maximale

Beschleunigung $Sp^{CG} \le N_r+1$ unabhängig von P für BPX-Vorkonditionierer und Mehrgitter V-Zyklus. Für den hierarchischen Basis-Vorkonditionierer ergibt sich $Sp^{CG} \to 1$ für P>3. {dim V_k} wächst dabei höchstens geometrisch mit dem Faktor $(1+\frac{3}{N_r+1})$.

(ii) Für $\overline{D}_k^p \neq D_k^p$ wird bei optimaler Lastaufteilung auf Ω_{j+1}, $j \to \infty$ und globaler Kommunikation erreicht, daß $N_r^{\overline{p}} \to P$. Für lokale Kommunikation kann die Last auf Ω_{j+1} auch noch optimal verteilt werden, die Aufteilungen auf $\{\Omega_k\}_{k=j,\dots,0}$ werden aber exponentiell schlechter, was im ungünstigsten Fall, für $P > O(N_r^2)$ und zusammenhängende Gebiete $\{D_k^p\}$ dazu führen kann, daß dann $W_k^{CG} \approx W_{j+1}^{CG}\ \forall k$ gilt.
Für Ω aus $\Re^3$ ist entsprechend $P > O(N_r^3)$ zu setzen.

Damit entstehen, unabhängig von der verwendeten Lösungsstrategie, denkbar schlechte Beschleunigungen, deren Wert unabhängig von der Zahl der eingesetzten Prozessoren nur von der Kommunikationsstruktur der Prozessoren abhängt.

6.2.3. Zusätzliche Voraussetzungen

Wir werden nun Möglichkeiten diskutieren, die die Problemstellung der statischen Lastverteilung doch noch sinnvoll machen.
Die diskutierte ungünstige Lastverteilung kann grundsätzlich ausgeschlossen werden, wenn für das geometrische Wachstum der {dim V_k} gilt

(42) $$\dim V_{k+1} > (1+\frac{3}{N_r+1}) \dim V_k.$$

Für geringeres Wachstum bleibt noch die Möglichkeit, nicht zusammenhängende Gebiete $\{D_k^p\}$ zu verwenden, die einen größeren Kommunikationsaufwand für die Randdaten erfordern.
Die Topologie der logischen Verbindungen zwischen den einzelnen Prozessoren, die für die geforderte lokale Kommunikation üblicherweise ein regelmäßiges zweidimensionales Netz ist, kann auch so gestaltet werden, daß eine baumartige Struktur eingebettet wird. Damit verläuft der Abfall der

Auslastung der Prozessoren auf den einzelnen Ebenen nicht mehr exponentiell. Für dieses Vorgehen bieten sich quintäre Bäume an, die die Daten schneller übertragen können, als durch Verfeinerung des Gitters je Ebene neue Punkte und damit Daten entstehen. Diese Struktur ist auch für beschränkte N_r, $P \to \infty$ möglich und erfordert je Prozessor höchstens 6 weitere Verbindungen zu anderen Prozessoren.

6.3. Übersicht über statische Ansätze

Unter der Voraussetzung, daß die Problemstellung der statischen Lastverteilung sinnvoll ist, wollen wir nun verschiedene Standardansätze betrachten und feststellen, daß diese für Multilevel-Verfahren nur bedingt geeignet sind.
Im Fall der hierarchischen Vorkonditionierung muß nicht $\{ D_k^p \}$ verteilt werden, sondern $\{ D_k^p \setminus D_{k-1}^p \}$, was durch die Kopplung der einzelnen Gitter Ω_k durch lokale Kommunikation aber zu denselben Ergebnissen führt.
Sei im folgenden ein geometrisches Wachstum der $\{\dim V_k\}$ mit einem Faktor größer $(1 + \frac{3}{N_r+1})$ angenommen. Gesucht ist eine Verteilung

(43) $\{ \overline{D}_k^p \}_{k=0,\dots,j}$, die W_j^{CG} minimiert,

mit der Nebenbedingung, daß N_r beschränkt ist für $j \to \infty$. Diese Verteilung sollte in

(44) $$O(W_j^{CG}) = O(\frac{\dim V_j}{P} + \log P) \text{ Zeit}$$

auf einem Parallelrechner, einschließlich aller nötigen Datentransfers, geschehen können. Es ist zu teuer, die exakte Lösung des Minimierungsproblems zu berechnen, da das Problem im allgemeinen NP-vollständig ist.
Es bieten sich Heuristiken der kombinatorischen Optimierung zur näherungsweisen Minimierung an, die ausgehend von der Verteilung $\{D_k^p\}_{k=0,\dots,j-1}$ in einer Art von Diffusionsprozeß eine gute Näherung für die gesuchte Verteilung liefern, wie z.B. Ansätze mit „simulated annealing“ [Fox & Otto] oder auch Metaheuristiken wie „tabu search“ [Glover] zur geschickten lokalen Minimierung von W_j^{CG}. Diese Ansätze liefern im allgemeinen Verfahren von deutlich höherer Ordnung als gesucht.

Andere Verfahren arbeiten mit rekursiver Bisektion des Gebietes Ω_j, das von zwei entgegengesetzt liegenden Gitterpunkten ausgehend bezüglich von Abstandsmaßen und Maßen für die Zahl von Verbindungen in zwei gleich große Teile geteilt wird, und verteilen dann in einem weiteren Schritt diese Teilgebiete möglichst gut auf die gegebene Prozessortopologie [Bastian], [Berger & Bokhari], [Literatur zu MinCut]. Diese Verfahren sind mit einem Aufwand $\geq O(\log P \dim V_j)$ nicht nur zu teuer, sondern nehmen in dieser Form auch keine Rücksicht auf die Verbindung der verschiedenen Gitter Ω_k untereinander.
Ansätze, die auf einer von den Triangulierungen zunächst unabhängigen, a priori wählbaren Aufteilung der Gebiete beruht, scattered decomposition [Fox & Otto], sind vom Aufwand her billiger. Wenn sich die Aufteilung an der Maschenweite des gröbsten Gitters orientiert, sind die Aufteilungen für große Prozessorzahlen im allgemeinen zu grob, so daß einzelne Prozessoren keine Gitterteile erhalten können. Orientiert sich die Verteilung aber an feineren Gittern, so kann an nicht verfeinerten Stellen des Gitters globale Kommunikation notwendig werden. Im übrigen ist die Aufteilung bezüglich der Gebietsränder normalerweise schlecht, während die Last für feine Maschenweiten gut verteilt werden kann. Für adaptive Multilevel-Methoden ist das Verfahren in dieser Form weniger geeignet, obwohl die Komplexität genügen würde.

6.3. Statistischer Ansatz mit geeigneter Komplexität

Nachdem die besprochenen Standardverfahren zur Parallelisierung für Multilevelverfahren nicht optimal geeignet sind, werden wir versuchen, ein an die Struktur meherer Gitter und an die Kommunikationsstruktur der Prozessoren angepaßtes Lastverteilungsverfahren zu konstruieren, das keine größere Komplexität hat als der iterative Löser selbst und ihn damit vom Aufwand her nicht dominiert.
Wir machen daher im folgenden Aufteilungsverfahren einen statistischen Ansatz. Gegeben sei eine Aufteilung von Ω_{j-1} auf P Prozessoren und eine lokale Kommunikationsstruktur als symmetrische, reflexive Relation R zwischen den Prozessoren. Gesucht sei eine Abbildung G von $\Omega_j \setminus \Omega_{j-1}$ auf die

Menge {1,…,P}. G sei bereits auf Ω_{j-1} definiert. Für ein $g \in \Omega_j \setminus \Omega_{j-1}$, das durch Halbierung einer Kante mit den Endpunkten g_1 und g_2 entstanden ist, gelte

$$(45) \qquad \mathbf{P}(\, G(g) = p \,) = \frac{w(p)}{\sum_{R'(p',G(p_1)) \wedge R'(p',G(p_2))} w(p')}$$

für alle p mit $R'(p,G(p_1))$ und $R'(p,G(p_2))$ und einer reflexiven, nicht mehr notwendig symmetrischen Relation $R' \subset R$ und der Wahrscheinlichkeit $\mathbf{P}$.

G ist für $R' \cup R'^{*} = R$ wohldefiniert. Für eine gute Lastverteilung muß nun das Maximum der Erwartungswerte der $\{\#D_j^p\}_p$ minimiert werden. Diese Gleichverteilung, in einer geeigneten Norm gemessen, führt für gegebenes oder heuristisch gewähltes R´ auf ein nichtlineares Gleichungssystem der Dimension P für die Terme w(p), das iterativ näherungsweise gelöst werden kann.

Das Gesamtverfahren hat damit keine höhere Ordnung als W_j^{BPX} . Aussagen über die Qualität der Lastverteilung können damit allerdings nur noch für große $\dim V_j$ gemacht werden.

Um sowohl die Aufteilung des Beispiels für das uniform verfeinerte Einheitsquadrat zu erreichen, als auch in allgemeineren Fällen gute Näherungen für das Minimum der W_j^{CG} zu erzielen, müssen die Ränder der Gebiete klein sein. Das kann man im allgemeinen dadurch erreichen, daß innere Punkte bereits gesondert vor der Berechnung der w(p) behandelt werden und daß Punkte g mit gleichen Nachbarprozessoren und gleichem G(g) in zusammenhängenden Gruppen angeordnet werden. Für stark adaptiv verfeinerte Gebiete, also nur langsam steigende $\{\dim V_j\}$, werden die entstehenden Ränder naturgemäß anteilig größer als für quasi-regulär verfeinerte Gebiete.

Es kann aber sowohl eine gute Lastverteilung als auch die Einhaltung der gegebenen lokalen Kommunikationsstruktur wie auch eine günstige Ordnung des Verteilungsverfahrens selbst erreicht werden.

7. Experimentelle Ergebnisse

Nachdem wir bisher auf der Basis des Kostenfunktionals optimiert haben, wollen wir überprüfen, in wie weit die Annahmen, die zu dem Funktional

geführt haben, mit der Praxis übereinstimmen. Dazu vergleichen wir Meßwerte verschiedener Parallelrechner mit den Ergebnissen des Kapitel 4.
Die adaptiven Verfahren der Kapitel 5 und 6 wurden bisher noch nicht vollständig auf Parallelrechnern implementiert. Deren Ergebnisse werden aber in einer folgenden Arbeit vorgestellt werden.
Wir betrachten das nicht adaptive Beispiel des quasiuniformen regulär verfeinerten Einheitsquadrates. Der diskrete Differentialoperator wird durch einen 3×3 Stern charakterisiert. Jede Triangulierung entsteht durch die Zerlegung jedes Dreiecks der vorhergehenden Triangulierung in jeweils 4 kongruente Teildreiecke. Das Problem wird mit einem BPX-vorkonditionierten CG-Verfahren gelöst. Die in einer Kette angeordneten Prozessoren bearbeiten jeweils ein Rechteck der Kantenlänge des Quadrates auf jeder Verfeinerungsebene, also die Zeilen mit den Nummern $\{d_k^p\}$. Die Messwerte werden verglichen mit dem allgemeinen Kostenfunktional W^{CG} und den speziell für dieses Beispiel hergeleiteten Näherungen (33)-(37) aus Kapitel 4.

7.1. Ergebnisse auf Intel iPSC/2

Die in den Abbildungen 7.1.1 bis 7.1.4 aufgeführten Messungen stammen von einem Rechner mit verteiltem Speicher (Intel iPSC/2), dessen Prozessoren in Hypercube-Topologie verbunden sind und asynchron kommunizieren. Die Übertragungszeiten sind dabei fast wegunabhängig. Die meisten Kommunikationen finden allerdings zwischen physikalischen Nachbarn statt, so daß sich dieser Effekt kaum auswirkt.
Die Meßwerte für die Beschleunigung der Rechnung durch den Einsatz mehrerer Prozessoren gegenüber einem Prozessor sind in Abbildung 7.1.1 gegenübergestellt dem Kostenfunktional W_j^{CG}, hier W geschrieben, das die Anzahl der Rechenoperationen und die Anzahl der Sende- und Empfangsoperationen berücksichtigt. Zum Vergleich sind noch die Werte der Formel (33) angegeben. Die Werte für die Beschleunigung sind angetragen gegen die feinste verwendete Verfeinerungsstufe. Die einzelnen Kurven stellen verschiedene Zahlen von Prozessoren dar.

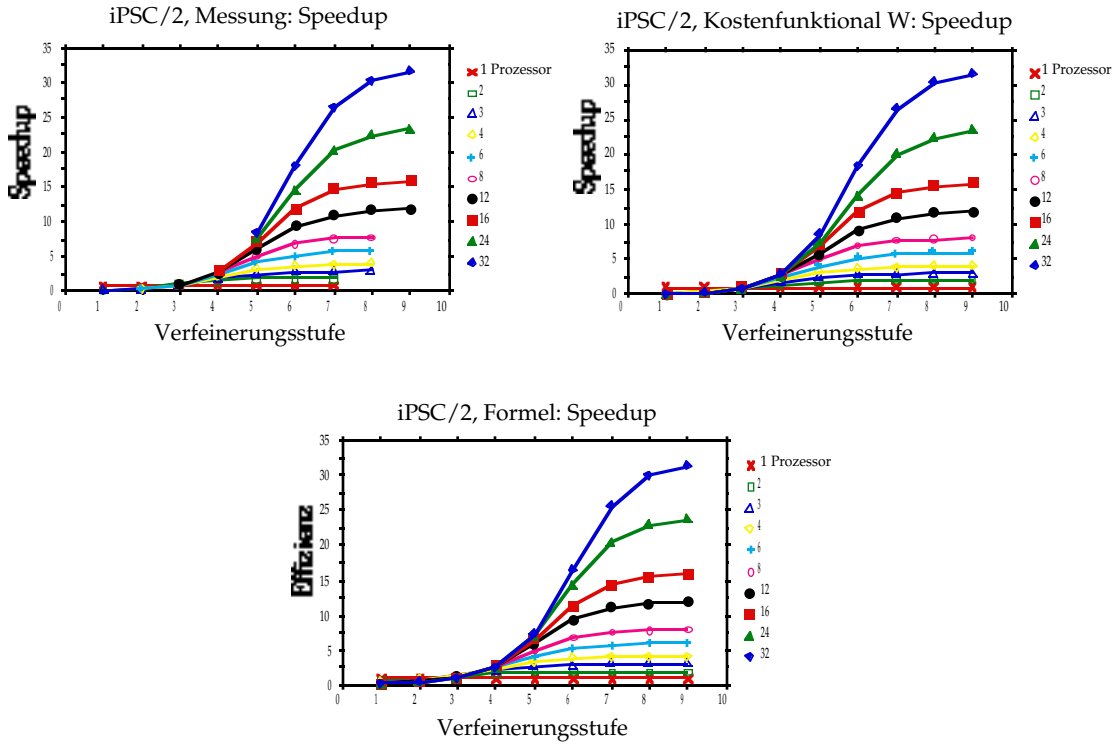

Abb 7.1.1 a-c. *Speedup: Messung, Kostenfunktional und Formel*

Die Meßwerte zeigen, daß für große Verfeinerungsstufen, also große Zahlen von Unbekannten, nahezu jede mögliche Beschleunigung erreicht werden kann. So können auch auf Verfeinerungsstufe 9 alle 32 Prozessoren sinnvoll eingesetzt werden und das Ergebnis in über 31.4-facher Geschwindigkeit berechnen gegenüber einem Prozessor, wenn dieser genügend Speicher hätte. Die Vergleichswerte für einen Prozessor mit den Stufen 8 und 9 mußten extrapoliert werden, was bei der hohen Meßgenauigkeit und Güte der Modellrechnung aber sehr genau möglich war. Im übrigen ist eine gute Übereinstimmung der Meßwerte mit den beiden Modellen zu beobachten, obwohl hier als einziger rechnerabhängige Wert das Verhältnis von Rechengeschwindigkeit zu mittlerer Kommunikationszeit eingeht.

Die Grafiken in Abbildung 7.1.2 geben die Effizienz der Ausnutzung der Prozessoren wieder. Dabei sind links jeweils Kurven für einzelne Prozessorzahlen abgebildet, wobei die erzielte Effizienz gegen die feinste Verfeinerungsstufe angetragen ist. Auf der rechten Seite ist mit Kurven für einzelne Verfeinerungsstufen die Effizienz gegen die Zahl der verwendeten Prozessoren

angetragen. In den folgenden drei Zeilen sind zunächst die gemessenen Werte angegeben. Dann folgen die Werte, die mit dem Kostenfunktional für W_j^{CG} bestimmt wurden, und schließlich werden diese mit der Formel (34) verglichen.

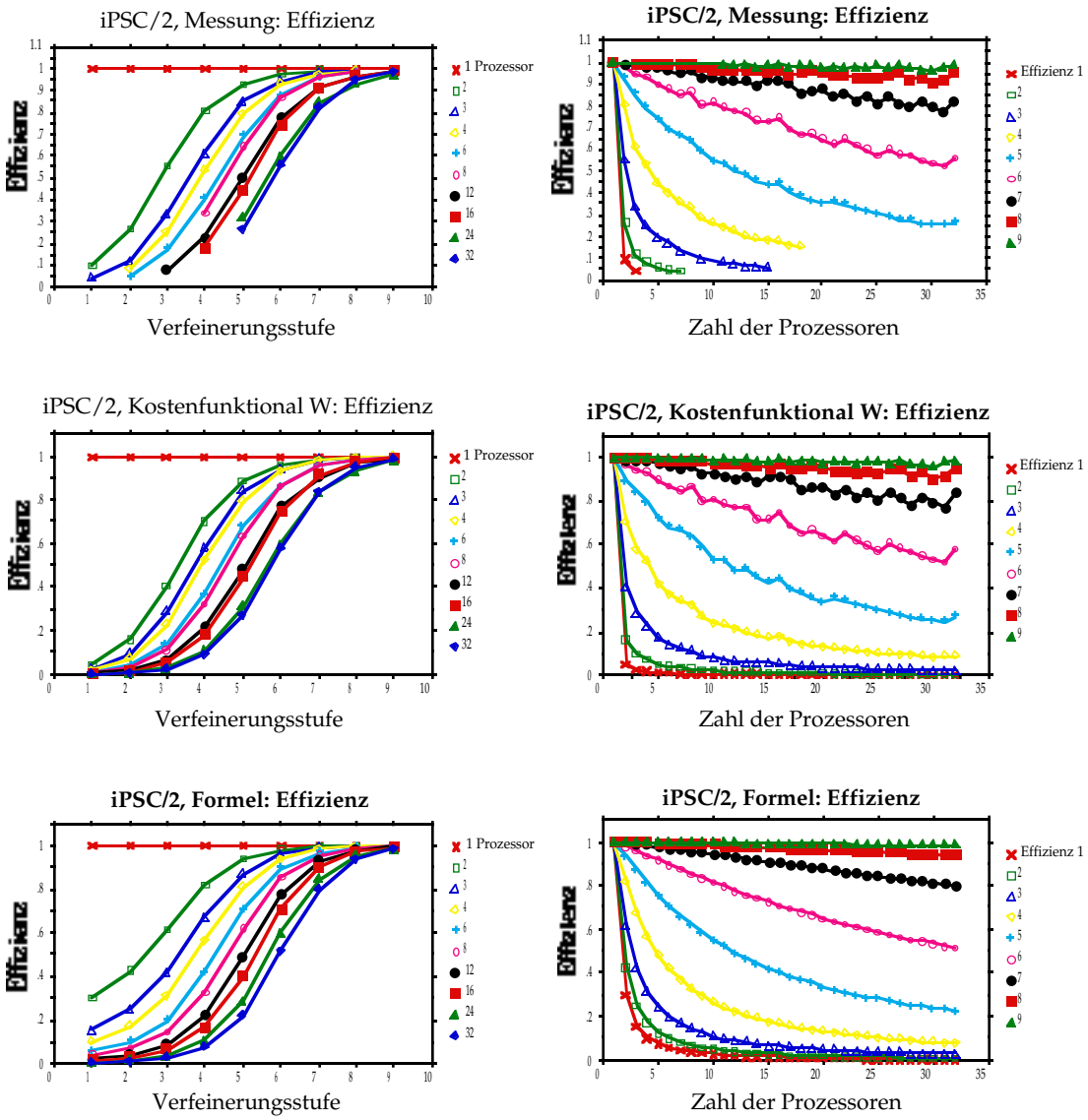

Abb 7.1.2 a-f. *Effizienz: Messung, Kostenfunktional und Formel*

Die Meßwerte zeigen, daß für große Verfeinerungsstufen Werte für die Effizienz von 96%, oft über 98%, bei kleinen Prozessorzahlen auch über 99%, erreicht werden können. Für kleinere Verfeinerungsstufen ist deutlich der Abfall der Effizienzen bei steigenden Prozessorzahlen zu beobachten.

Die beiden Modellrechnungen zeigen wiederum eine hohe Übereinstimmung mit den Meßwerten. Da die Formel (34) stetig differenzierbar in P und j ist, erscheinen in dieser einfachen Näherung keine Sprünge oder Zacken, obwohl die Werte qualitativ richtig sind.
Das Kostenfunktional W_j^{CG} dagegen kann auch ungleiche Lastverteilungen, die durch die Division der Zeilenzahlen durch die Prozessorzahlen entstehen, modellieren und gibt daher selbst Sprünge und Zacken in den Kurven richtig wieder. So ist kein nennenswerter Unterschied zwischen den Meßwerten und dem Kostenfunktional zu beobachten, was die Bedeutung dieses Funktionals und seiner theoretischen Begründung zeigt.

Im folgenden wird die Frage untersucht, mit welcher Zahl von Prozessoren sich die kürzesten Ausführungszeiten ergeben. Diese Zahlen sind, nach Verfeinerungsstufen geordnet, für die aus (33) abgeleitete Formel (35) (gerundete Werte), das Kostenfunktional W_j^{CG} und die eigentlichen Meßwerte in Tabelle 7.1.3 angegeben.

Verfeinerungs-stufe	Formel (35)	W_j^{CG}	Messung
1	0.2	1	1
2	0.8	1	1
3	3.1	1	2
4	12.2	16	16
5	49.0	32 (max)	32 (max)
6	195.8 (>129)	32 (max)	32 (max)
7	783.2 (>257)	32 (max)	32 (max)
8	3132.8 (>513)	32 (max)	32 (max)
9	12531.3 (>1025)	32 (max)	32 (max)

Tab 7.1.3. *optimale Zahl von Prozessoren*

Ab Verfeinerungsstufe 5 reichen die vorhandenen 32 Prozessoren nicht mehr aus, um eine sinnvolle Aussage zu machen. Wir können lediglich sagen, daß alle vorhandenen Prozessoren zu einer Geschwindigkeitssteigerung beitragen und daher sinnvoll eingesetzt werden können.
Die Formel (35) zeigt aber, daß ab Verfeinerungsstufe 6 alle durch die zeilenweise Aufteilung des Gebietes möglichen Prozessorzahlen zu einer

Beschleunigung führen werden. Auch in diesem Fall liefern beide Modellrechnungen Ergebnisse der gleichen Größenordnung wie die Messungen.

Die Abbildung 7.1.4 zeigt absolute Ausführungszeiten (in ms) für die beiden Modellrechnungen (33) und W^{CG}und die Messung, angetragen gegen die Zahl der Prozessoren für die Verfeinerungsstufe 3.

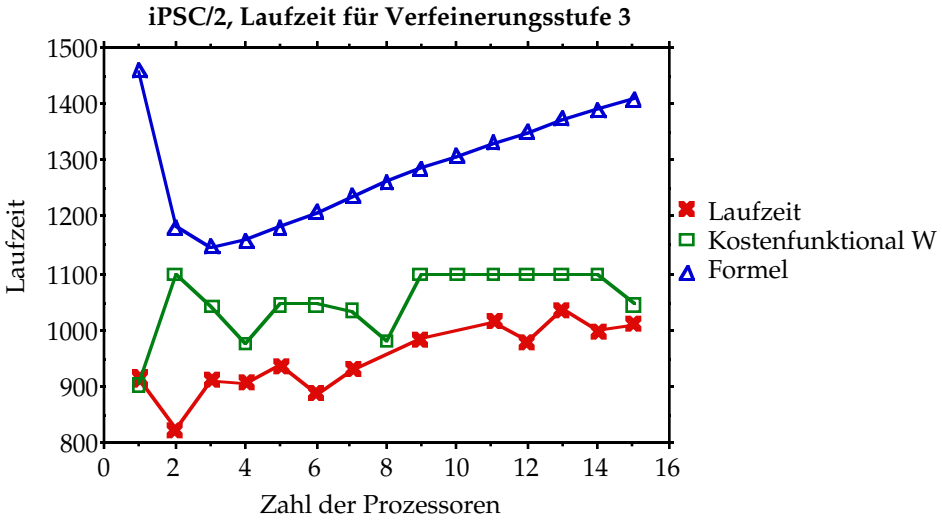

Abb 7.1.4. *Laufzeit: Messung, Kostenfunktional und Formel*

Deutlich ist die Übereinstimmung der Größenordnungen der drei Kurven zu erkennen, wobei für diese relativ geringe Verfeinerungsstufe Einflüsse der Gebietsränder und zeitliche Überlappungseffekte mit den gröberen Gittern die absoluten Werte noch stark beeinflussen. Die Werte stehen in Übereinstimmung mit der Tabelle für die optimalen Prozessorzahlen und illustrieren deren Ergenisse für die Verfeinerungsstufe 3.

7.2. Ergebnisse auf T800

Die Werte der Abbildungen 7.2.1 und 7.2.2 sind Messungen mit einer Kette von Transputern (T800), also einem Rechner mit verteiltem Speicher. Durch das Betriebssystem Helios bedingt, sind die Initialisierungskosten für Sende- und Empfangsoperationen relativ zur etwa 9-fachen Rechengeschwindigkeit deutlich höher, was die Ergebnisse negativ beeinflußt.
Die Prozessoren konnten aus Hardware-Gründen nicht als Hypercube vernetzt werden, weshalb hier die Konfiguration als Kette gewählt wurde. Damit steigt der Aufwand für W^{SC} nicht mit log P sondern mit P/2, was aber für die kleinen Prozessorzahlen bis zu 8 und dem geringen Rechenzeitanteil der Skalarprodukte an der Gesamtrechenzeit vernachlässigbar bleibt.
Ein weiteres Problem entsteht auf den gröbsten Gittern, wenn einige Prozessoren keine Punkte mehr zu berechnen haben. Da die Kommunikationsstrukturen bereits während der Übersetzungsphase des Programmtextes festgelegt werden müssen, können die untätigen Prozessoren nicht einfach mehr übersprungen werden, sondern müssen die Informationen weiterleiten. Damit werden Kommunikationsoperationen auf den gröbsten Gittern teurer als auf feineren Gittern, was aber keinen großen Einfluß auf die Gesamtergebnisse hat.

Es folgen Meßwerte in Abbildung 7.2.1 für die Beschleunigung der Rechnung durch den Einsatz mehrerer Prozessoren gegenüber einem Prozessor. Auf der linken Seite sind die Werte für die Beschleunigung angetragen gegen die feinste verwendete Verfeinerungsstufe. Die einzelnen Kurven stellen verschiedene Zahlen von Prozessoren dar. Auf der rechten Seite sind die Meßwerte gegen die Zahl der Prozessoren angetragen, wobei für jede Verfeinerungsstufe eine Kurve eingezeichnet wurde.

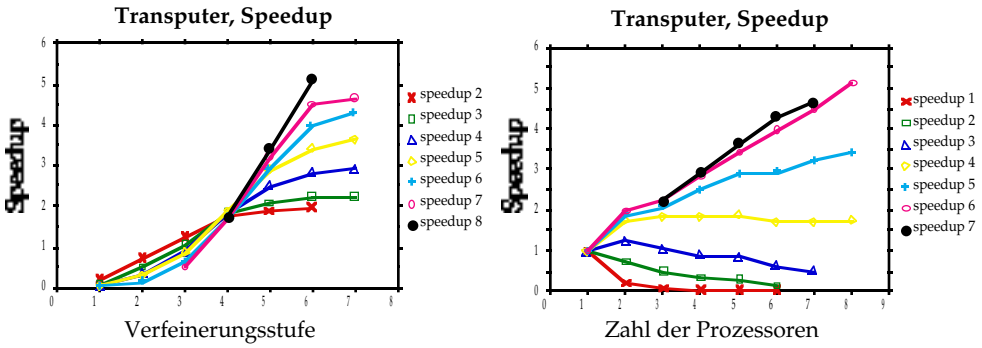

Abb 7.2.1 a,b. *Messung Speedup*

Die Meßwerte zeigen, daß für große Verfeinerungsstufen, also große Zahlen von Unbekannten, nahezu jede mögliche Beschleunigung erreicht werden kann. So können auch auf Verfeinerungsstufe 7 alle vorhandenen 8 Prozessoren sinnvoll eingesetzt werden und das Ergebnis in über 5.1-facher Geschwindigkeit berechnen gegenüber einem Prozessor. Im übrigen ist eine gute qualitative Übereinstimmung der Meßwerte mit den Werten des iPSC/2 zu beobachten, obwohl hier eine andere Prozessortopologie verwendet wurde.

Die Grafiken in Abbildung 7.2.2 geben die Effizienz der Ausnutzung der Prozessoren wieder. Dabei sind links jeweils Kurven für einzelne Prozessorzahlen abgebildet, wobei die erzielte Effizienz gegen die feinste Verfeinerungsstufe angetragen ist. Auf der rechten Seite ist mit den Kurven für einzelne Verfeinerungsstufen die Effizienz gegen die Zahl der verwendeten Prozessoren angetragen.

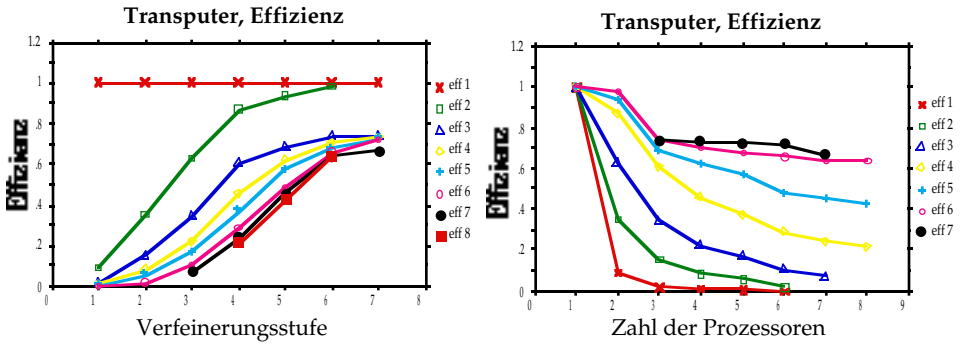

Abb 7.2.2 a,b. *MessungEffizienz*

Die Meßwerte zeigen, daß für große Verfeinerungsstufen Werte für die Effizienz von 64% bis 73%, bei zwei Prozessoren auch bis zu 98%, erreicht werden können. Für kleinere Verfeinerungsstufen ist deutlich der Abfall der Effizienzen bei steigenden Prozessorzahlen zu beobachten. Sobald mehr als zwei Prozessoren eingesetzt werden, wirkt die Kommunikation der inneren Prozessoren mit ihren jeweiligen Nachbarn sehr bremsend. Der Datentransfer kann nur jeweils in eine bestimmte Richtung durchgeführt werden und wirkt damit synchronisierend auf die Prozesse, was zu den zu beobachtenden Leistungseinbußen führt.
Die Messungen zeigen wiederum eine qualitative Übereinstimmung mit den Meßwerten des iPSC/2.

7.3. Ergebnisse auf Alliant FX/2800

Die folgenden Messungen wurden mit einem Parallelrechner mit gemeinsamem Speicher und 8 Prozessoren durchgeführt (Alliant FX/2800). Die Messungen wurden mit dem vorhandenen Multitasking-Multiuser-Betriebssystem durchgeführt. Jeder Hierarchie von Teilgebieten $\{D_k^p\}_k$ wurde dabei ein Systemprozess zugeordnet. Die Abbildung der Prozesse auf Prozessoren wurde durch das Betriebssystem zur Laufzeit durchgeführt. Die Kommunikation der Prozesse erfolgt über die jeweiligen Randdaten der Gebiete, die im gemeinsamen Speicher angeordnet sind. Die inneren Daten sind aus Geschwindigkeitsgründen in privaten Speicherbereichen abgelegt. Die Rolle der Kommunikationsoperationen, also der Sende- und Empfangsoperationen, übernehmen hier Synchronisationsoperationen mit Semaphoren, die die Kohärenz der Daten sicherstellen. Dabei entstehen die gleichen Lastausgleichsprobleme wie für Rechner mit verteiltem Speicher.
Die Meßwerte hängen deutlich von der übrigen Belastung des Rechners ab und schwanken daher. Es wurde jeweils der beste Wert einer Reihe von Messungen bei möglichst unbelasteter Maschine verwendet.

Es folgen in Abbildung 7.3.1 Meßwerte für die Beschleunigung der Rechnung durch den Einsatz mehrerer Prozessoren gegenüber einem Prozessor. Dabei bezieht sich die Beschleunigung hier auf die Gesamtlaufzeit der Rechnung,

unabhängig davon, wieviel Rechenzeit den einzelnen Prozessen von dieser Echtzeit vom Betriebssystem zugeteilt werden. Die Meßwerte sind gegen die Zahl der Prozesse angetragen, wobei für jede Verfeinerungsstufe eine Kurve eingezeichnet wurde.

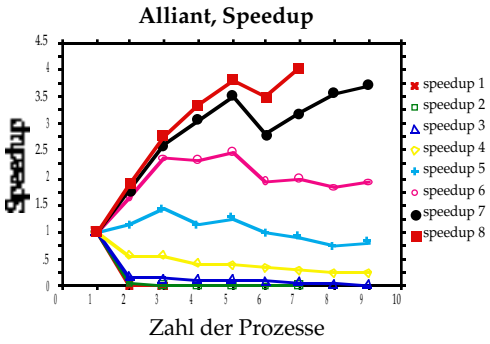

Abb 7.3.1. *Messung Speedup*

Der Meßwert für 9 Prozesse zeigt eine Laufzeitverbesserung gegenüber kleineren Werten durch eine verbesserte Granularität des Gesamtprogramms. Die Meßwerte zeigen, daß für große Verfeinerungsstufen, also große Zahlen von Unbekannten, gute Beschleunigungen erreicht werden können. So können auch auf Verfeinerungsstufe 9 alle vorhandenen 8 Prozessoren sinnvoll eingesetzt werden und das Ergebniß in 4-facher Geschwindigkeit berechnen gegenüber einem Prozessor. Im übrigen ist eine gute qualitative Übereinstimmung der Meßwerte mit den theoretischen Werten zu beobachten. Der Abfall der Leistung beim Übergang von 5 auf 6 und 7 Prozesse kann mit den Hardware-Eigenschaften des Rechners erklärt werden. Die Datenpfade der Prozessoren zu den Speichermodulen sind so ausgelegt, daß die Hälfte der Prozessoren nahezu wechselwirkungsfrei auf privatem Speicher arbeiten kann. Arbeiten aber mehr Prozessoren mit ähnlichen Speicherzugriffen, so treten starke bremsende Bus- und Cache-Konflikte auf.

Die Grafiken in Abbildung 7.3.2 geben die Effizienz der Ausnutzung der Prozessoren wieder. Die Effizienz bezieht sich hier im Gegensatz zur Beschleunigung auf die maximale, von einem Prozeß verbrauchte Rechenzeit. Dabei sind links jeweils Kurven für einzelne Prozeßzahlen abgebildet, wobei die erzielte Effizienz gegen die feinste Verfeinerungsstufe angetragen ist. Auf

der rechten Seite ist mit den Kurven für einzelne Verfeinerungsstufen die Effizienz gegen die Zahl der verwendeten Prozesse angetragen.

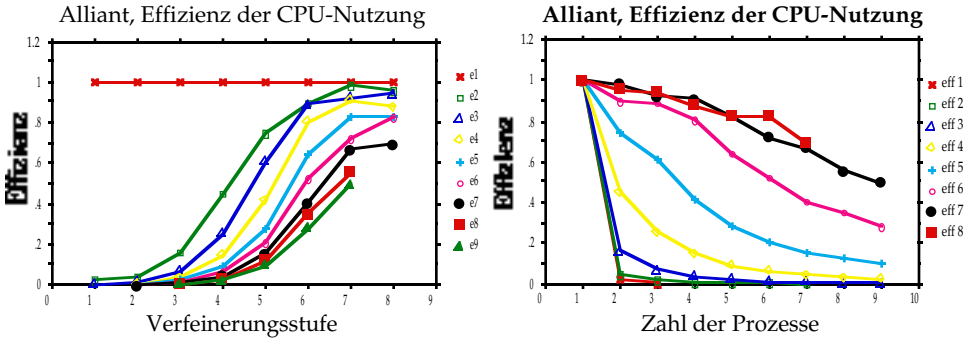

Abb 7.3.2 a,b. *Messung Effizienz*

Die Meßwerte zeigen, daß für große Verfeinerungsstufen Werte für die Effizienz von 70-85%, bei zwei Prozessen auch 98%, erreicht werden können. Für kleinere Verfeinerungsstufen ist deutlich der Abfall der Effizienzen bei steigenden Prozessorzahlen zu beobachten. Dieser Abfall ist aber für alle Verfeinerungsstufen zu erkennen. Das bedeutet, daß hier nicht nur Lastausgleichsprobleme und Zugriffe auf gemeinsame Randdaten bremsend wirken, sondern auch der unabhängige Ablauf des Programms auf mehreren verschiedenen Prozessoren mit unabhängigen Daten. Dieser Effekt kann auch während des normalen Multiuser-Betrieb des Rechners beobachtet werden. Eine bestimmte Grundrechenlast auf einigen Prozessoren wirkt sich fatal auf die übrigen aus.
Die Messungen zeigen wiederum eine gute qualitative Übereinstimmung mit der Theorie (36) und (37), bis auf die Leistungsabfälle für innere Gitterpunkte bei großen Prozessorzahlen.

Es folgen analog zu den abgebildeten Effizienzen aufgebaute Grafiken in Abbildung 7.3.3, die die reziproken Werte der Effizienzen, hier als Kosten der Parallelisierung bezeichnet, zeigen.

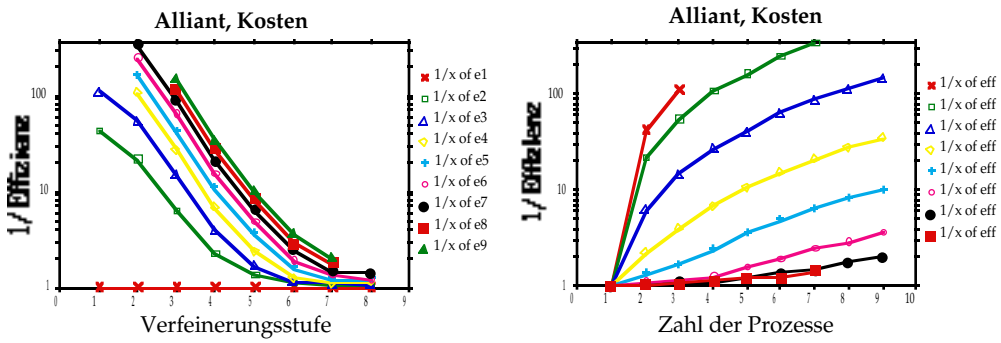

Abb 7.3.3 a,b. *Messung Kosten der Parallelisierung*

An den Kosten der Parallelisierung ist deutlich zu sehen, daß sich aufgrund der starken Beeinflussung der Prozessoren untereinander eine Parallelisierung einer derart auf Durchsatz und weniger auf Spitzenleistung hin optimierten Maschine, wie es die Werte nahelegen, nur für große Probleme und Verfeinerungsstufen lohnt. Wenn die Zahl der Unbekannten beispielsweise kleiner als Tausend ist, kostet die Parallelisierung mindestens einen Faktor 14 an Rechenzeit, bei zwei Prozessen aber zumindest den Faktor 6 mehr als die sequentielle Rechnung.

Bei diesen Vergleichen muß man aber weiterhin berücksichtigen, daß ein Prozessor der Alliant eine Größenordnung schneller arbeitet als ein Prozessor des iPSC/2, und damit immer noch Faktoren schneller als ein T800 Prozessor. Bei diesen Zuwächsen an Rechengeschwindigkeit wird es zusehend schwieriger, die Kommunikations und Datentransferleistung der Rechner in ähnlichen Größenordnungen zu steigern.

8. Schlußbemerkungen

Wir haben die Parallelisierung der einzelnen Komponenten eines von Bramble, Pasciak und Xu vorgeschlagenen vorkonditionierten Verfahrens der konjugierten Gradienten, eingebettet in ein adaptives Finite-Elemente-Programm diskutiert, die zusätzliche Forderungen an Triangulierungen, Finite-Elemente-Räume und Gittermanipulationsalgorithmen stellen. Nachdem Standardverfahren der Lastverteilung teils eine größere Komplexität als das Lösungsverfahren selbst besitzen und zum anderen Teil nicht genügend an Multilevelverfahren angepaßt sind, haben wir einen Ansatz für ein neues Aufteilungsverfahren gemacht, das auf einem statistischen Ansatz beruht. Mit

einer gemischten Strategie der Aufteilung von Gitterpunkten und Dreiecken konnte ein Gesamtverfahren von optimaler Ordung erreicht werden.
Die experimentellen Ergebnisse auf einigen Parallelrechnern zeigten eine hohe Übereinstimmung mit denen durch das Kostenfunktional vorhergesagten Werten, das die Grundlage unserer Analysen bildete. Für statische Aufteilungen konnte eine ideale Parallelisierbarkeit des Verfahrens gezeigt werden. Die Unterschiede zu Parallelisierungen anderer bekannter Multilevelverfahren waren gering, konnten aber im Modell erfaßt werden.
Die nächste Aufgabe besteht in einer praktischen Umsetzung der Verfahren im adaptiven Fall. Entsprechende Ergebnisse werden in einer folgenden Arbeit vorgestellt werden.
Ziele weiterer Untersuchungen könnten Experimente mit vollständig parallelisierten Finite-Elemente-Programmen oder auch genauere Untersuchungen des vorgeschlagenen Lastverteilungsverfahrens sein. Die Abstimmung des Verfahrens wird aber mit Sicherheit stark vom verwendeten Parallelrechner, vom umgebenden iterativen Löser und den Ansprüchen an die Qualität der Verteilung abhängen. Eine andere Frage ist, ob trotz der höheren Komplexität in der Praxis nicht doch ein teureres Lastverteilungsverfahren zu günstigeren Meßwerten führt.

9. Danksagungen

An dieser Stelle möchte ich dem Lehrstuhl Prof. Dr. Ritter für die Erlaubnis zur Benutzung des verwendeten Transputersystems danken.
Für die kritische Durchsicht des Textes danke ich insbesondere Prof. Dr. R. Hoppe und Dr. M. Griebel, wobei die Verantwortung für alle Mängel des Textes natürlich beim Autor bleiben.

Literatur

R.E. Bank: *PLTMG. A Software Package for Solving Elliptic Partial Differential Equations*., Philadelphia, SIAM 1990

R.E. Bank, T. Dupont, H. Yserentant: *The Hierarchical Basis Multigrid Method*, Numer. Math. 52 (1988), 427-458

R.E. Bank, A. Weiser: *Some a posteriori Error Estimators for Elliptic Partial Differential Equations*., Math. Comp. 44 (1985), 283-301

P. Bastian: *Kommunikation in Transputernetzen und Lastverteilung unstrukturierter Gitter*, Vortrag, GAMM Seminar 31.05.91 Univ. Heidelberg

M.J. Berger, S. Bokhari: *A Partitioning Strategy for Nonuniform Problems on Multiprocessors,* IEEE Trans. Comput., vol. C-36, no. 5 (1987), 570-580

P.E. Bjorstad, O.B. Widlund: *To overlap or not to overlap: A note on a domain decomposition method for elliptic problems*, SIAM J. Sci. Stat. Comput. 10 (1989), 1053-1061

J.H. Bramble, J.E. Pasciak, A.H. Schatz: *The construction of preconditioners for elliptic problems by substructuring I*, Math. Comp. 47 (1986), 103-134

J.H. Bramble, J.E. Pasciak, J. Wang, J. Xu: *Convergence estimates for product iterative methods with applications to domain decomposition and multigrid*., Math. Sciences Inst. Cornell Univ., Techn Report 90-39 (1990)

J.H. Bramble, J.E. Pasciak, J. Xu: *Parallel Multilevel Preconditioners*., Math. Comp. 55 (1990), 1-22

B. Briggs, L. Hart, S. McCormick, D. Quinlan: *Multigrid Methods on a Hypercube*, in S.F. McCormick (Ed.), Proc. Third Copper Montain Conf. Multigrid Methods, New York, Marcel Dekker (1988), 63-83

P. Deuflhard, P. Leinen, H. Yserentant: *Concepts of an Adaptive Hierarchical Finite Element Code*., IMPACT of Computing in Science and Engineering 1 (1989), 3-35

G.C. Fox, S.W. Otto: *Concurrent Computation and the Theory of Complex Systems,* in M.T. Heath (Ed.), Hypercube Multiprocessors 1986, Philadelphia, SIAM, 244-268

F. Glover: *Tabu Search,* CAAI Report 88-3, University of Colorado, Boulder (1988)

W. Hackbusch: *Multi-grid methods and applications*, Springer, Berlin (1985)

P. Leinen: *Ein schneller adaptiver Löser für elliptische Randwertprobleme auf Seriell- und Parallelrechnern*. Dissertation, Dortmund (1990)

H. Mierendorff: *Parallelization of Multigrid Methods with Local Refinements for a Class of Nonshared Memory Systems*, in S.F. McCormick (Ed.), Proc. Third Copper Montain Conf. Multigrid Methods, New York, Marcel Dekker (1988), 449-465

P. Oswald: *On discrete norm estimates related to multilevel preconditioners in the finite element method*., Erscheint in: Proc. Int. Conf. Constr. Theory of Functions Varna´91

H. Yserentant: *Two Preconditioners Based on Multi-Level Splitting of Finite Element Spaces*. Preprint SC 89-9, Konrad Zuse Zentrum für Informationstechnik Berlin, 1990

B. Veer:*The CDL Guide*, Distributed Software Ltd., 1990

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

33

34

35

36

37

38

39

40

41

42

43

44

45