

Optimierung von FE-Rechenkernen für die niederfrequente Schallfeldsimulation

Guido Bartsch*

FGAN – Forschungsinstitut für Hochfrequenzphysik und Radartechnik
Neuenahrer Straße 20, 53343 Wachtberg-Werthhoven

19. März 2003

Der Optimierung von Rechenkernen für die wellentheoretisch basierte Schallfeldsimulation mittels der Methode der Finiten Elemente wurde bislang wenig Beachtung geschenkt. Grund für diese Vernachlässigung war die bislang bestehende Meinung, daß durch Taktratensteigerung moderner CPUs, der größte Teil an Geschwindigkeitssteigerung von FE-Codes erzielt werden könne. In [1] konnte aber nachgewiesen werden, daß nicht die Taktrate der CPU, sondern die Übertragungsgeschwindigkeit zwischen Speicher und CPU die dominierende Größe hinsichtlich der Performance von FE-Codes ist. Eine Optimierungstechnik, die diesem Umstand Rechnung trägt, wird hier vorgestellt.

1 Einleitung

Angesichts der rasanten Entwicklung der Rechnerhardware schieben sich noch vor einigen Jahren Rechenzeiten in der Zukunft kaum noch ein Problem darzustellen. Die 'Binaurale Echtzeit-Auralisation', unter Berücksichtigung von Drehung und Bewegung des Zuhörers oder der Schallquelle im Raum, ist keine Zukunftsmusik mehr, sondern bereits Bestandteil aktueller Entwicklungen von gekoppelten akustisch, visuell und taktil virtuellen Umgebungen. Gerade dazu wurden in den letzten Jahren verschiedene Projekte in und außerhalb Deutschlands bearbeitet [3, 4, 5].

Vergleicht man aber die Ergebnisse aus Computersimulationen untereinander und mit Meßergebnissen, so beobachtet man folgende Phänomene [6]:

- Manche Programme liefern hinreichend genaue Ergebnisse; die größere Anzahl liefert allerdings Ergebnisse, die deutlich stärker vom Sollwert abweichen, als z.B. durch die Standardabweichung bei wiederholten Messungen angegeben ist. Die Abweichungen sind zudem größer als die subjektiv wahrnehmbaren Unterschiedsschwellen bezüglich der einzelnen raumakustischen Parameter.
- Die Genauigkeit der Vorhersage der raumakustischen Kenngrößen hängt weniger von der Detailtreue des Computer-Modells ab, als vielmehr vom Geschick des Programmierers, Wesentliches von Unwesentlichem zu unterscheiden.
- Die Materialdaten, die den Raumflächen zugeordnet werden, sind teilweise ungenau oder nicht vollständig vorhanden. Der Absorptionsgrad wird typischerweise zu klein geschätzt.
- Systematische Abweichungen und Tendenzen in den Ergebnissen der Computersimulationen sind allen Programmen gemeinsam und lassen sich eindeutig auf eine unzureichende Modellbildung der 'physikalischen Wirklichkeit' zurückführen. Die in [6] genannten Abweichungen sind für das untersuchte Raumbeispiel und für den betrachteten Frequenzbereich um 1 kHz nur bei drei von siebzehn Programmen akzeptabel.

Und vor allem:

- Keines der heute verfügbaren Programme erlaubt eine effektive Simulation im unteren Frequenzbereich.

In [1] werden neue Ansätze und Implementierungen von raumakustischen Simulationsalgorithmen für die niederfrequente Schallfeldsimulation vorgestellt. Ein Teilaspekt, die Optimierung von FE-Rechenkernen, soll hier näher vorgestellt werden.

2 Problemidentifikation

Bei Anwendung der Finiten-Elemente-Methode zur Schallfeldsimulation wird der zu betrachtende Raum Ω zunächst in eine Menge von Teilkörpern, den sogenannten Finiten Elementen, unterteilt. Dies sind o.B.d.A. einfache Körper, wie z.B. Tetraeder oder Hexaeder, von denen gefordert wird, daß die Materialeigenschaften innerhalb eines jeden Elements konstant sind.

Der erste Schritt zur näherungsweise Lösung des Randwertproblems

$$\Delta p + k^2 p = 0 \quad \text{in } \Omega, \quad (1)$$

$$\frac{\partial p}{\partial n} + i k \varrho_0 c_0 v_n = 0 \quad \text{auf } \Gamma_1, \quad (2)$$

$$\frac{\partial p}{\partial n} + i k \varrho_0 c_0 A_n p = 0 \quad \text{auf } \Gamma_2, \quad (3)$$

mit den Nebenbedingungen

$$\int_{\Omega} p \, d\Omega = 0, \quad (4)$$

ist dessen Überführung in eine entsprechende Variationsformulierung bzw. die Überführung der strengen Formulierung in eine sogenannte schwache Formulierung [2].

Multipliziert man (1) mit einer geeigneten Testfunktion q , z.B. aus dem Modalraum $\Phi(\Omega)$, d.h. $q \in \Phi(\Omega)$, und integriert über das Feldgebiet Ω , so erhält man als Ergebnis eine Darstellung des Feldproblems, die die Randbedingungen nunmehr implizit beinhaltet:

$$\begin{aligned} \int_{\Omega} \nabla p \cdot \nabla q \, d\Omega + i k \varrho_0 c_0 \int_{\Gamma_2} A_n p q \, d\Gamma_2 - k^2 \int_{\Omega} p q \, d\Omega \\ = -i k \varrho_0 c_0 \int_{\Gamma_1} v_n q \, d\Gamma_1. \end{aligned} \quad (5)$$

Unter Ausnutzung der konstanten Materialeigenschaften innerhalb eines Finiten Elements kann (5) schließlich in Matrixform umgeschrieben werden:

$$(S - i k C - k^2 M) p = -i k f. \quad (6)$$

In (6) bezeichnet $S \in \mathbb{R}^{m \times m}$ die sogenannte Steifigkeitsmatrix, $C \in \mathbb{C}^{m \times m}$ die Dämpfungsmatrix, $M \in \mathbb{R}^{m \times m}$ die Massenmatrix, $f \in \mathbb{C}^m$ den Kraftvektor und $p^T \in \mathbb{C}^m$ den für eine vorgegebene Anregungskreisfrequenz $\omega = k c_0$ gesuchten Koeffizientenvektor.

*Bartsch@FGAN.de

3 Methoden

Unabhängig davon, ob eine modale oder nicht modale Methode zur Lösung von (1) bis (4) herangezogen wird, identifiziert man das Lösen von linearen Gleichungssystemen als den rechenaufwendigsten Teil eines FE-Lösers. Als Kernroutine stellt hier die Matrix-Vektor-Multiplikation $y \leftarrow A \cdot x$ den Hot-Spot innerhalb des Lösers dar.

```

Initialisiere Zähler  $i$  für betrachtete Matrixzeile:  $i \leftarrow 0$ 
Für jeden  $i$ -ten Zeilenvektor  $A_i$  der Matrix  $A_{ij}$  {
  Führe Skalarmultiplikation  $y_i \leftarrow A_i^T \cdot x$  durch,
   $i \leftarrow i + 1$ 
}
  
```

Abbildung 1: Nicht optimierte Matrix-Vektor-Multiplikation

Definiert man als spezifische algorithmische Performance SAP eines Rechenkerns die Anzahl notwendig¹ ausgeführter 'floating point'-Operationen $O_{FP,min}$ pro Prozessortakt

$$SAP := \frac{O_{FP,min}}{T_{CPU}}, \quad (7)$$

so läßt sich hiermit die Effizienz eines Codes auf einer beliebigen Zeilarchitektur quantitativ beurteilen.

Wie in [1] gezeigt wurde, ist eine effektive Simulation mit FEM selbst für große Schallfeldprobleme möglich, wenn effiziente Algorithmen verwendet werden, die möglichst umfassend von der Sparsizität der Systemmatrizen Gebrauch machen. So werden z.B. sparse Matrizen in sogenannten 'compressed row storage'-Format (CRS) abgespeichert, einer Repräsentation, bei der nur die von Null verschiedenen Elemente abgelegt werden. Die Effizienzsteigerung der Matrix-Vektor-Multiplikation ergibt sich dann zum einen aus der mitunter erheblichen Reduktion der zu verarbeitenden Datenmenge, zum anderen aus der Möglichkeit, das Problem noch 'in core' rechnen zu können.

```

Initialisiere Zähler  $k$  für betrachteten 4-er Matrixzeilenblock:  $k = 0$ 
Beginnend mit der ersten Zeile  $i = 4k$  des  $k$ -ten Matrixzeilenblocks {
  Hole Daten für die  $(i + 1)$ -te Skalarmultiplikation und führe parallel dazu die  $i$ -te Skalarmultiplikation  $y_i \leftarrow A_i^T \cdot x$  durch.
  Hole Daten für die  $(i + 2)$ -te Skalarmultiplikation und führe parallel dazu die  $(i + 1)$ -te Skalarmultiplikation  $y_{i+1} \leftarrow A_{i+1}^T \cdot x$  durch.
  Hole Daten für die  $(i + 3)$ -te Skalarmultiplikation und führe parallel dazu die  $(i + 2)$ -te Skalarmultiplikation  $y_{i+2} \leftarrow A_{i+2}^T \cdot x$  durch.
  Hole Daten für die  $(i + 4)$ -te Skalarmultiplikation und führe parallel dazu die  $(i + 3)$ -te Skalarmultiplikation  $y_{i+3} \leftarrow A_{i+3}^T \cdot x$  durch.
   $k \leftarrow k + 1$ 
}
  
```

Abbildung 2: Optimierte Matrix-Vektor-Multiplikation

Sparsizität und nahezu statistische Besetzung der FE-Systemmatrizen mit Nicht-Null-Elementen bringen aber das Problem mit sich, daß das Speicherzugriffsverhalten der Algorithmen, zumindest aus Prozessorsicht, als statistisch charakterisiert werden muß. Bei größeren Problemen hat dies zur Folge, daß der sogenannte 'hardware prefetch' einer modernen CPU nicht mehr in der Lage ist, die benötigten Daten rechtzeitig aus dem Hauptspeicher zu holen und dem Rechenkern der CPU zur Verfügung zu stellen. Die sogenannten 'floating point pipelines' laufen leer; der die Performance charakterisierende SAP-Wert sinkt auf ein Bruchteil seines möglichen Spitzenwertes, der bei einer CPUs mit nur einer arithmetischen Pipeline bei $SAP_{max} = 1$ liegt.

Das Versagen des 'hardware prefetch' bei CRS-basierten FE-Algorithmen kann aber durch den Einsatz von 'software data

prefetch'-Mechanismen, kompensiert werden. Unter 'software data prefetch' versteht man ein vom Algorithmus initiiertes Anfordern von Daten aus dem Speicher, ohne diese direkt nutzen zu wollen.

4 Ergebnisse

Wie Abbildung 3 illustriert, läßt sich die Performance einer CRS-basierten Matrix-Vektor-Multiplikation durch Optimierung (vgl. Abb. 2) auf mehr als das Doppelte gegenüber der nicht optimierten Version (vgl. Abb. 1) steigern.

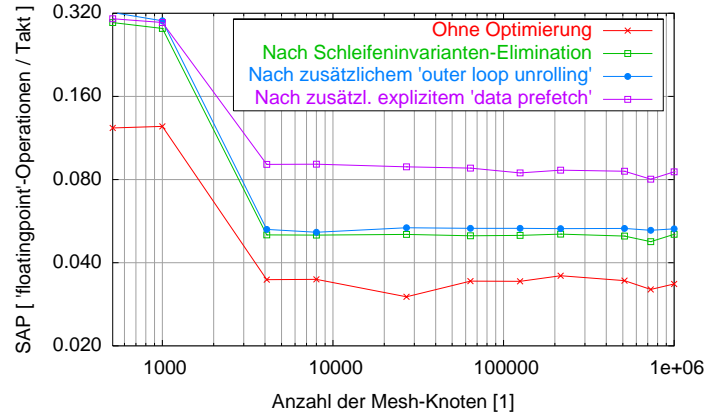


Abbildung 3: Steigerung der SAP durch Optimierung

Dazu ist es aber notwendig, daß neben klassischen Optimierungstechniken wie Schleifeninvarianten-Elimination und dem in Abbildung 2 dargestellten 4-fachen 'loop unrolling', auch der CPU-unterstützten Parallelität von Rechnung und Datenfluß Augenmerk geschenkt wird. Diese läßt sich bei CRS-basierten FE-Codes insbesondere durch den Einsatz von 'software data prefetching' wesentlich verbessern.

Details zur Optimierung und Implementierung effizienter Löser für die niederfrequente Schallfeldsimulation findet man in [1].

Literatur

- [1] BARTSCH, G.: *Effiziente Methoden für die niederfrequente Schallfeldsimulation*. Driesen Edition Wissenschaft. Verlag Dr. H. H. Driesen, Taunusstein, 1. Auflage, 2003.
- [2] HACKBUSCH, W.: *Theorie und Numerik elliptischer Differentialgleichungen*. Teubner, Stuttgart, 1996.
- [3] JANTZEN, M.: *Aufbau eines Signalprozessorsystems zur Echtzeitverarbeitung von Raumimpulsantworten*. Diplomarbeit, Institut für Technische Akustik, RWTH Aachen, Deutschland, 1993.
- [4] LEHNERT, H.: *Vereinfachung von binauralen Impulsantworten zur Auralisierung von Rückwürfen*. In: *Fortschritte der Akustik, DAGA*, Seiten 303–306, 1995.
- [5] NASSHAN, K., G. SCHUPP und U. STEPHENSON: *Echzeitauralisation als Element optischer und akustischer virtueller Realität*. In: *Fortschritte der Akustik, DAGA*, Seiten 307–310, 1995.
- [6] VORLÄNDER, M.: *International Round Robin on Room Acoustical Computer Simulations*. In: *Proc. 15th International Congress on Acoustics*, Seiten 689–692, Trondheim, Norwegen, 1995.

¹Notwendig sei eine Operation dann, wenn sie zum Ergebnis numerisch beiträgt.