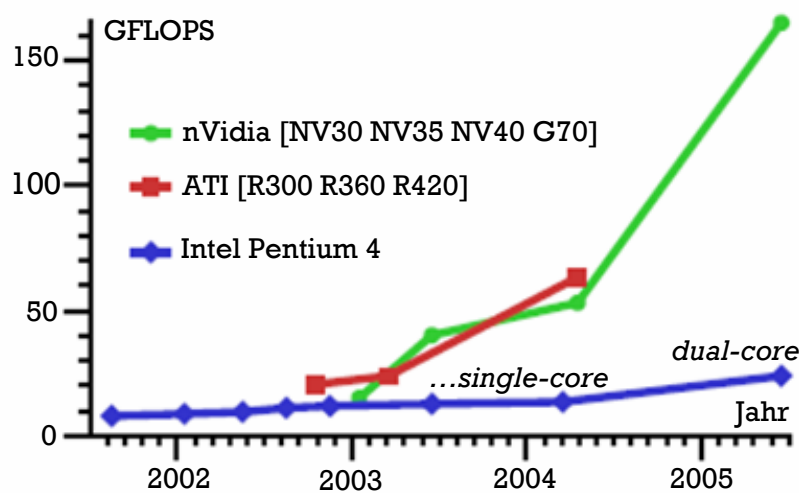


# Universeller Einsatz moderner Grafikkarten am Beispiel der Optimierung eines Spracherkennungssystems

In den letzten Jahren wurde Grafikkarte aufgrund der stark wachsenden Computerspielebranche immer leistungsfähiger, wobei die Kosten für eine Grafikkarte dank der hohen Verkaufszahlen in etwa gleich geblieben sind. Die sehr spezielle Hardware führt eine Vielzahl von Berechnungen parallel aus und kann mittlerweile höhere Rechenkapazitäten und Speicherbandbreiten aufweisen als aktuelle CPU's.



GPU's „programmable floating-point performance“ (gemessen aufgrund von multiply-add Instruktionen) - Ian Buck, Stanford University -

## GPGPU – Ein Framework

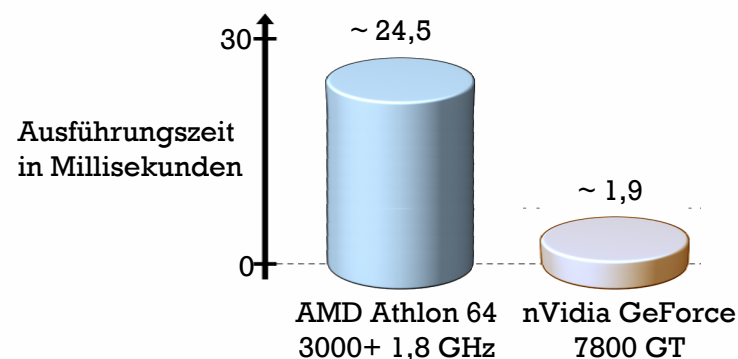
Diese enormen Rechenkapazitäten auch für Anwendungen fernab der Grafik zu nutzen ist aktuelles Forschungsgebiet der General Purpose Computation On Graphics Processor Units (kurz GPGPU) und ebenfalls Gegenstand dieser Arbeit.

Da bestehende Algorithmen nicht problemlos auf eine Grafikkarte verlagert werden können, sondern in Termen der Grafikprogrammierung neu überdacht und reimplementiert werden müssen, wurde zunächst ein entsprechendes Framework auf Basis von OpenGL und DirectX erstellt. Mit Hilfe dessen ist es möglich geworden, Berechnungen auf Grafikkarte auszuführen, ohne sich um die grafiklastige und sehr aufwändige, sowie fehleranfällige Kommunikation zwischen CPU und GPU beschäftigen zu müssen.

## Einsatz in einem Spracherkennung

Die Qualität eines Spracherkennungssystems steigt mit der Anzahl der zugrunde liegenden Sprachmodelle. Allerdings werden für diese Qualitätssteigerung auch erhöhte Rechenkapazitäten benötigt um den ebenfalls steigenden Aufwand der Bewertung von Merkmalsvektoren in der verfügbaren Zeit bewältigen zu können.

Unter Einsatz des entstandenen Frameworks konnten die benötigten Berechnungen zur Merkmalsbewertung erheblich optimiert werden. Durch Ausnutzung der verschiedenen Möglichkeiten, die heutige Grafikkarte bietet, wurde schließlich bei einer beispielhaften Datenmenge von 48.000 Bewertungen, ein Speedup um den Faktor 12 gegenüber einer gleichwertigen CPU-Implementierung erzielt.



## Ausblick und weitere Entwicklung

Das Framework hat sich im praktischen Einsatz bewährt und kann aufgrund der ausführlichen Dokumentation, sowie den erstellten Beispielprogrammen zur Einarbeitung in die Thematik, sowie als Grundlage für weitere GPU-Implementierungen verwendet werden, insbesondere da im GPGPU-Umfeld Dokumentationen bisher noch recht spärlich sind.

Auch die Optimierung der Merkmalsbewertung hat sich aufgrund der durchgeführten Tests als sinnvoll und äußerst performant erwiesen. Für die tatsächliche Integration in ein reales Spracherkennungssystem wurden bereits die notwendigen Schritte innerhalb der schriftlichen Ausarbeitung aufgezeigt und resultieren in einem hybriden System (CPU/GPU) zur Bewertung der Merkmalsvektoren.