

Wie gut ist die Arbeit verteilt?

CFK: direkte Load-Balancing-Kontrolle | 18.09.2026

18. September 2026

Inhaltsverzeichnis

1	Zusammenfassung	2
1.1	Versuchsmatrix und Kontrollen	2
1.2	Messdefinition	2
2	Lastverteilung der drei Phasen	3
3	Bewegung und Messaufwand	4
4	Einordnung und Folgerungen	5
4.1	Was das Scheduling leisten kann	5
4.2	Nächste technische Schritte	5
4.3	Grenzen und Reproduktion	5

Die erste Aufbau-Serie mass Phasenzeiten, aber keine Lastverteilung. Diese Ergaenzung erfasst Aufgaben und Arbeitszeiten pro Thread, ohne die feste Praefixfront auf Tiefe 6 zu optimieren. Das Ziel ist eine Diagnose, nicht der Nachweis einer bereits besseren Lastverteilung.

Dreihemden: complete, 102 Prozesse, 528 Messzustaende, 6.6 Minuten einschliesslich exakter Referenzkontrollen.

Schmalzberg: complete, 102 Prozesse, 528 Messzustaende, 14.2 Minuten einschliesslich exakter Referenzkontrollen.

1 Zusammenfassung

1.1 Versuchsmatrix und Kontrollen

100 000 und 300 000 Teilchen; vier Verteilungen; drei Seeds; Profilierung bei 1, p und 2p Threads (p = physische Kerne) sowie ohne Profilierung bei p Threads. Ein Aufwaermzustand, danach drei Messzustaende je Prozess. Reihenfolge je Datenblock gemischt. Zusaetzlich 40 bewegte Hotspot-Zustaende mit drei Seeds und gepaarter Profilierung an/aus. Jeder Zustand wird ausserhalb der Zeitmessung vollstaendig mit dem seriellen Referenzgitter verglichen und auf Konformitaet geprueft.

Hotspot: 90 Prozent der Teilchen bewegen sich in einem Quadrat der Kantenlaenge 0.015 um $(0.6 + 0.18 \sin(t), 0.6 + 0.18 \cos(t))$; 10 Prozent bilden einen Hintergrund. Es handelt sich um vorgegebene Bewegung, nicht um Gravitation oder eine physikalische Simulation.

1.2 Messdefinition

Fuer Thread j sei B_j die Summe seiner gemessenen Task-Wandzeiten. Der Balanceindex ist $L = \sum(B_j) / (p \max(B_j))$. Auch Threads ohne Aufgaben werden mit $B_j = 0$ einbezogen. $L = 1$ bedeutet gleiche gemessene Arbeitszeiten; L nahe $1/p$ bedeutet Dominanz eines Threads. Die Besitzersuche misst die Schleifendauer pro Thread ohne abschliessende Team-Barriere.

Diese Groesse ist keine CPU-Auslastung und kein Speedup: Allokation, Speicherwartzeit und Praemption beeinflussen die Wandzeiten. Barrierewartzeiten werden nicht direkt gemessen. Erfasst sind ausserdem Eingangs- und Ausgangsmengen je Aufgabe sowie Teilchenzahl und durchlaufene Tiefen je Besitzer-Thread. Das ermoeeglicht die Trennung von Arbeitsmenge und Dauer.

2 Lastverteilung der drei Phasen

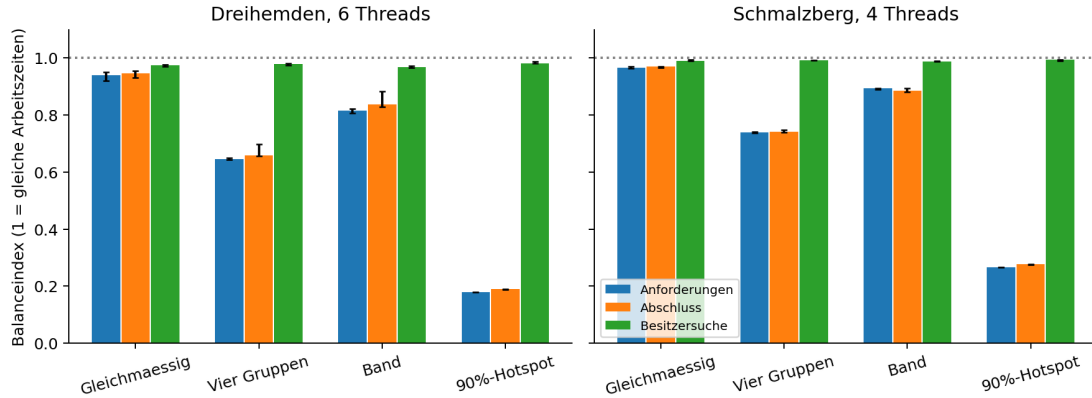


Abbildung 1: Balanceindex der drei parallelen Phasen bei 300000 Teilchen.

300 000 Teilchen, physische Kerne. Zuerst wird ueber die drei Frames je Lauf gemittelt, danach Median und Minimum/Maximum der drei Seeds gebildet. Keine Konfidenzintervalle; aufeinanderfolgende Frames sind keine unabhaengigen Replikate.

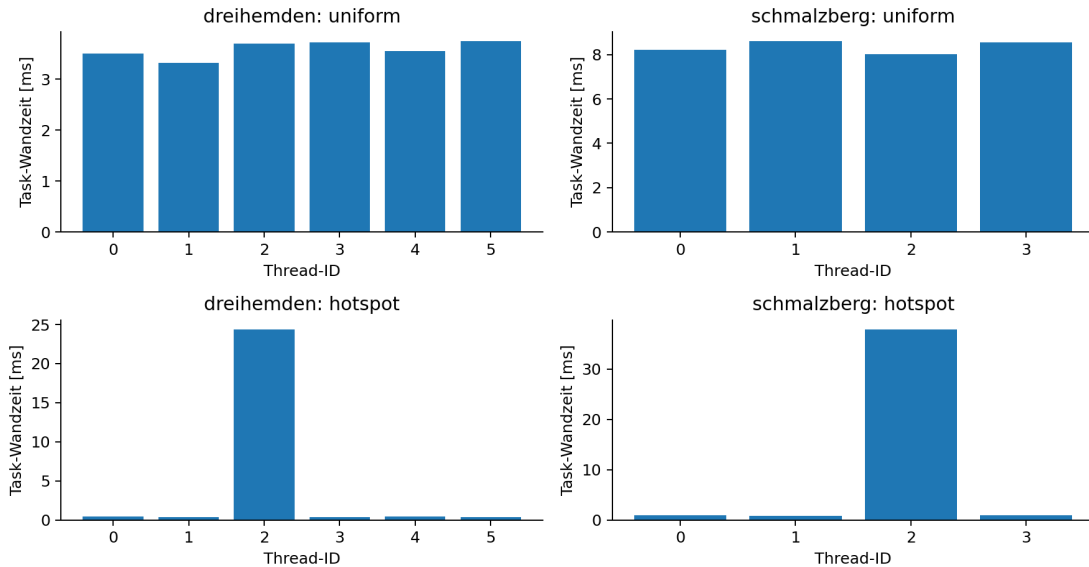


Abbildung 2: Arbeitszeiten der Praefixaufgaben je Thread: uniforme Kontrolle und Hotspot.

Konkrete Praefix-Arbeitszeiten je Thread: Seed 807, erster Messzustand. Die Thread-ID der groessten Aufgabe kann durch das dynamische Scheduling wechseln. Entscheidend sind die ungleichen Balken, nicht die Identitaet eines bestimmten Kerns.

3 Bewegung und Messaufwand

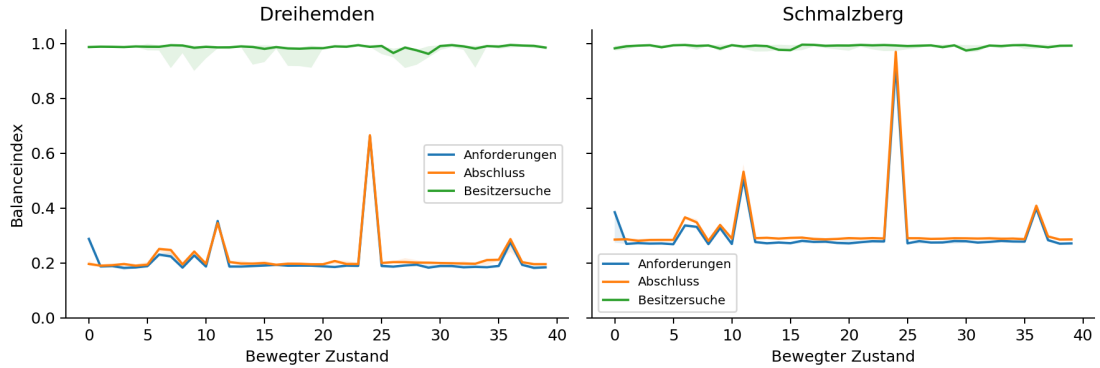


Abbildung 3: Balanceindex entlang der bewegten Hotspot-Folge.

Bewegter 90%-Hotspot mit 100 000 Teilchen. Linien: Median dreier Seeds pro Zustand; Baender: deren volle Spanne. Aenderungen zeigen, wie die feste grobe Front eine bewegte Konzentration auf eine oder mehrere Aufgaben verteilt.

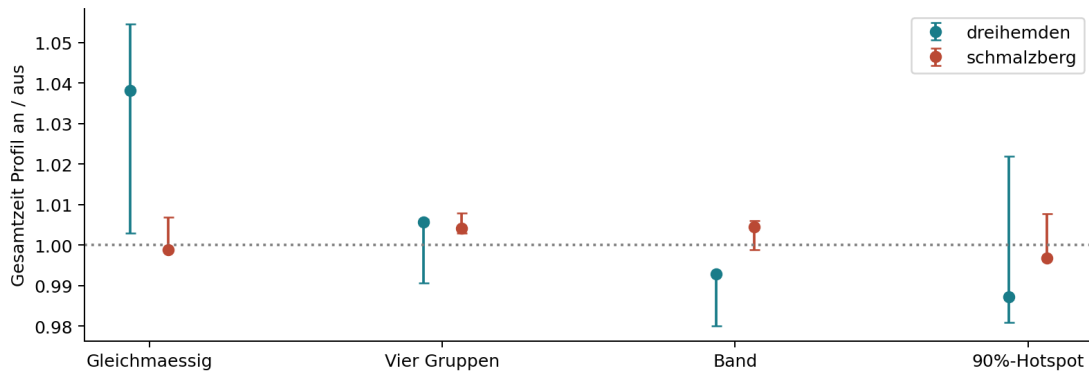


Abbildung 4: Gepaarter Vergleich der Gesamtzeit mit und ohne Instrumentierung.

Messaufwand bei 300 000 Teilchen und p Threads: gepaarter Quotient der gesamten Aufbauzeit mit/ohne Profilierung. Eins bedeutet keine gemessene Aenderung. Werte unter eins koennen aus Laufstreuung entstehen und sind keine Beschleunigung durch die Uhr. Es wird kein Overhead nachtraeglich von den Ergebnissen abgezogen. Auch das Zaehlen der Baumtiefen in der Besitzersuche gehoert zum Instrumentierungsaufwand.

4 Einordnung und Folgerungen

Dreihemden, Hotspot mit 300 000 Teilchen: medianer Balanceindex Anforderungen 0.179, Abschluss 0.191, Besitzersuche 0.982.

Die groesste Praefixaufgabe umfasst dabei 90.3 Prozent der Teilchen. Dies ist ein mengenbasierter Hinweis auf die Ursache, unabhaengig von den gemessenen Zeiten.

Schmalzberg, Hotspot mit 300 000 Teilchen: medianer Balanceindex Anforderungen 0.266, Abschluss 0.276, Besitzersuche 0.994.

Die groesste Praefixaufgabe umfasst dabei 90.3 Prozent der Teilchen. Dies ist ein mengenbasierter Hinweis auf die Ursache, unabhaengig von den gemessenen Zeiten.

4.1 Was das Scheduling leisten kann

dynamic,¹ verteilt fertige Aufgaben an freie Threads. Eine bereits laufende grosse Aufgabe wird dadurch nicht weiter geteilt. Ein schnellerer Scheduler allein beseitigt daher keine Dominanz eines einzelnen Praefixgebiets. Die feste Tiefe 6 ist die zu pruefende Granularitaet, nicht ein mathematisch notwendiger Bestandteil des konformen Abschlusses.

4.2 Naechste technische Schritte

Eine datenabhaengige Praefixfront sollte stark besetzte Gebiete weiter aufteilen, bis genuegend handhabbare Aufgaben vorhanden sind. Teilchenzahl ist ein erster Indikator; Trennungstiefe und Abschlussgroesse beeinflussen die wirkliche Arbeit ebenfalls. Das Merge bleibt eine separate Kostestelle aus der ersten Serie. Bessere Balance garantiert deshalb noch keinen schnelleren Gesamtaufbau. Beide Optimierungen muessen mit denselben Kontrollfaellen geprueft werden.

4.3 Grenzen und Reproduktion

Keine physikalische Anwendungsmessung, keine direkte Messung von CPU-Zeit oder Barrierestillstand, keine kontinuierliche Temperaturkontrolle. GCC 13.3 auf Dreihemden und GCC 11.4 auf Schmalzberg; absolute Rechnerunterschiede sind daher keine reine Hardwaremessung. Rohdaten enthalten auch Ein-Thread- und SMT-Kontrollen; die Hauptgrafiken konzentrieren sich auf physische Kerne.

Lokal bestanden: alle regulaeren Tests, Gittertests ohne OpenMP sowie ASan/UBSan ohne Leckpruefung. Zusaetzlich wird geprueft, dass Profilierung keine Blattzuordnung veraendert und dass die aufgezeichneten Mengen mit den globalen Zaehlern uebereinstimmen. Kein Race-Detector-Nachweis.

Quellen: separat archivierter C++-Stand, ausfuehrbare Dateien, Compilerflags, Rohmessungen und Versuchsplan. Auswertung: report_balance_study.py. Die Tests wurden von Pascal Bauer beauftragt; Instrumentierung und Auswertung wurden von ChatGPT umgesetzt. Arbeit 0.3 blieb unveraendert. Der erste Bericht wurde lediglich auf die bereitgestellte LaTeX-Vorlage umgestellt; seine Messdaten und Auswertung blieben unveraendert.