



Zentren für
Kommunikation und
Informationsverarbeitung e.V.

c/o
Hochschule Heilbronn
Max-Planck Straße 39
D-74081 Heilbronn

Weiterentwicklung des Hochleistungsrechnens in Deutschland

Positionspapier

Stand: 02.12.2013

Redaktion:

Sprecherteam des ZKI-Arbeitskreises Supercomputing:
Prof. Dr.-Ing. Stephan Olbrich, Universität Hamburg
Dr. Harald Ziegler, Friedrich-Schiller-Universität Jena

Inhalt

1	Motivation und Ziele	3
2	Ausgangslage	4
3	Probleme und Lösungsansätze	5
4	Quellen	8

1 Motivation und Ziele

Hochleistungsrechnen ist für weite Teile der Wissenschaften zum unverzichtbaren Werkzeug geworden. Ein weiteres Wachstum dieses Werkzeuges in Breite und Tiefe ist zu erwarten und für konkurrenzfähige Forschung und Entwicklung unverzichtbar. Verschiedene Förderprogramme und Initiativen auf nationaler wie europäischer Ebene adressieren diese Entwicklung und haben, insbesondere auf dem Gebiet des Höchstleistungsrechnens, gute Erfolge gezeigt.

Aus Sicht des ZKI gibt es dennoch Mängel und Verbesserungspotenziale, die in diesem Papier aufgezeigt werden sollen. Dies betrifft insbesondere:

1. Die klassische Geräteförderung des Hochleistungsrechnens nach Art. 91b GG behindert den Zugang zu geeigneten Hochleistungsrechenkapazitäten auf den Ebenen 2 und 3, da zur länderübergreifenden Nutzung hohe Hürden zu überwinden sind. Insbesondere in kleinen Ländern ist so der Zugang zu HPC-Kapazitäten problemangemessener Architektur behindert. Der Zugang zu Kapazitäten der Ebene 2 fehlt bisweilen ganz.
2. Der Anteil der Betriebskosten (Energie, Kühlung) an den HPC-Gesamtkosten wächst. Während die Investitionskosten förderfähig sind (Bund, Länder), sind die Betriebskosten regelmäßig von den tragenden Institutionen (meist Universitäten) zu übernehmen. Dies behindert den Aufbau angemessener Kapazitäten sowie die institutions- und länderübergreifende Nutzung.
3. Software-Entwicklung, -Anpassung und -Optimierung ist, verglichen mit dem Einsatz von Hardware, eine deutlich nachhaltigere Investition. Eingesetzte Personalkosten in diesem Bereich können Hardware- und Betriebskosten deutlich überkompensieren. Organisationsformen und Förderkonzepte im Bereich der Pflege wissenschaftlicher Software sind aber nur schwach ausgebildet.

2 Ausgangslage

Die bereits hohe und stark zunehmende Bedeutung von High-Performance Computing (HPC) in den daten- und rechenintensiven wissenschaftlichen Disziplinen geht mit einem entsprechenden Zuwachs der Erwartungen der Nutzer an die Betriebseinheiten der Universitäten und Forschungseinrichtungen einher sowie dem Bedarf an fortgeschrittener Nutzerunterstützung, bis hin zu kooperativer Begleitforschung und -lehre in den Bereichen der Methodenwissenschaften. In den Empfehlungen und Positionspapieren der Deutschen Forschungsgemeinschaft (DFG) [1], des Bundesministeriums für Bildung und Forschung (BMBF) [2], des Wissenschaftsrats (WR) [3], des Gauss Centre for Supercomputing (GCS) [4] spiegeln sich diese Gesichtspunkte bereits wider. Dabei liegt das Modell einer HPC-Leistungs- und -Organisationspyramide zugrunde, das in 4 Ebenen („Tiers“) gegliedert ist:

Tier 0: Europäische Ebene – PRACE (Partnership for Advanced Computing in Europe)

Tier 1: Nationale Ebene – Gauss Centre for Supercomputing (GCS),
Verbund dreier Trägereinrichtungen

Tier 2: Regionale oder fachspezifische Ebene – Gauß-Allianz (GA),
Vereinsorganisation mit 9 ordentlichen und 8 assoziierten Mitgliedern

Tier 3: Hochschulen und öffentliche Forschungsinstitutionen,
Vereinsorganisation (ZKI) mit 223 Mitgliedern (ca. 80 % der deutschen Hochschul-IT)

Auf der Basis jeweils landesweit abgestimmter Arbeitsteilung werden kooperative Konzepte entwickelt und umgesetzt, die die nationale und europäische HPC-Leistungspyramide und die damit verbundene Supportkonzepte ergänzen (z. B. in Baden-Württemberg [5]).

Zu den Stärken der derzeitigen HPC-Versorgung – vor allem in Tier 0 und 1 – zählen:

- International konkurrenzfähige Tier-0- und Tier-1-Versorgung (Europa und GCS) mit Rechenleistung für exzellente Forschung mit Anforderungen im „Capability Computing“
- Verfügbarkeit untereinander abgestimmter Rechnerarchitekturen auf Tier 1
- Aufbau forschungsnaher Beratungs- und Unterstützungsleistungen im GCS (Tier 1)
- Bereitstellung von Fördermitteln für Projekte und Kompetenzzentren, z. B. „Software für Exascale-Computing“ (DFG-Schwerpunktprogramm) sowie „HPC-Software für skalierbare Parallelrechner“ und „Management und Analyse großer Datenmengen (Big Data)“ (BMBF)

Aus Sicht des ZKI sind jedoch auf den folgenden Gebieten Problemlagen im Rahmen der künftigen Förderungspolitik und der Zukunftskonzepte – vor allem in Tier 2 und 3 – zu beobachten:

- Förderung und Nutzbarkeit von Forschungs Großgeräten bzw. Forschungsbauten (Tier 2)
- Finanzierung der Betriebs- und Personalkosten
- Beschaffungszyklen und Planungssicherheit
- Rolle der Rechenzentren bei Betrieb, Beratung und Software Stewardship
- Koordinierung von Förderungen
- Organisation von Kompetenzzentren

Auf die konkreten Probleme und Lösungsansätze sowie auf den Koordinierungs- und Förderbedarf wird im folgenden Kapitel näher eingegangen.

3 Probleme und Lösungsansätze

Im Zuge der strategischen Planung der Weiterentwicklung des HPC sind aus Sicht des ZKI die folgenden, heute schon zu beobachtenden Problemlagen im Rahmen der künftigen Förderungspolitik und der Zukunftskonzepte – vor allem in Tier 2 und 3 – mit zu berücksichtigen. Zu dem jeweiligen Themenbereich werden hier auch Empfehlungen für Lösungsansätze gegeben.

(a) Förderung und Nutzbarkeit von Forschungs Großgeräten bzw. Forschungsbauten (Tier 2)

Die Förderung von Tier-2-Zentren – typischerweise als themenbezogene Forschungsbauten – ist insbesondere für kleine Länder nicht mehr adäquat. Da in kleinen Ländern in der Regel nur ein Zentrum für die Landesversorgung bestehen wird – oder eine länderübergreifende Nutzung eines Zentrums –, haben solche Zentren noch stärker als in großen Ländern themenübergreifenden Charakter. Die thematisch bezogene Förderung nach Art. 91b GG wird dafür als nicht adäquates Instrument angesehen (de facto wird hiervon bereits in großzügiger Auslegung der Gesetzeslage abgewichen, siehe z.B. HLRN). Außerdem ist die Verfügbarkeit von Tier-2-Systemen für die Nutzer je nach Bundesland sehr unterschiedlich ausgeprägt und zum Teil gar nicht gegeben (z.B. Sachsen-Anhalt, Thüringen).

Die Zugänglichkeit der Kapazitäten von Ebene 2 für alle Anwender (unabhängig von Ländergrenzen) ist wünschenswert, da ein Ausweichen auf Tier 1 oder 3 in vielen Fällen auch keine adäquate Lösung darstellt und gerade auf Ebene 3 eine hinreichende apparative Ausstattung nicht überall gegeben ist, insbesondere wegen der Konkurrenzsituation in Bezug auf die begrenzten, häufig nicht mit Zweckbestimmung versehenen Investitionsmittel, baulicher Unzulänglichkeiten (z. B. Serverräume) und der Betriebskosten (z. B. Energiebedarf). Die dadurch mögliche Auswahl aus mehreren Rechnerarchitekturen stellt einen Zusatznutzen dar.

Lösungsansätze:

Die Kriterien bzw. Richtlinien im Rahmen der Förderung von Forschungs Großgeräten (Art. 91b GG) sollten den Besonderheiten des HPC entsprechend angepasst werden.

Aufgrund der anteiligen Finanzierung aus Bundesmitteln sollte überlegt werden, ob diesem Anteil entsprechend generell eine länderübergreifende Nutzung ermöglicht werden kann.

Eine Förderung aller anfallenden Kosten (TCO) (siehe Abschnitt b) beseitigt Hürden für die länderübergreifende Nutzung.

(b) Finanzierung der Betriebskosten und Personalmaßnahmen – Rolle der Rechenzentren

Die Betriebskosten – insbesondere elektrische Energie und Kühlung – haben einen bereits erheblichen Anteil an den Gesamtkosten (über 5 Jahre bereits ähnlich hoch wie die Investitionskosten) und werden bei steigenden Strompreisen und im Rahmen der absehbaren Entwicklung der Rechnerarchitekturen noch deutlich zunehmen.

Die effiziente Nutzung massivparalleler und zum Teil auf Spezialprozessoren basierender Rechnerarchitekturen erfordert verstärkt Beratungsdienstleistungen, bis hin zu „Cross-Sectional Labs“, die den nutzenden Fachwissenschaftlern die nötigen Methodenkompetenzen zur Verfügung stellen. Nutzerabstimmung, Bedarfsplanung, Konzeption, Beschaffung, Betrieb und Nutzungsunterstützung sind kontinuierliche, personalintensive Begleitprozesse mit hohem Bedarf an Spezialkompetenzen. Hochschulrechenzentren – und hier speziell diejenigen mit HPC-Betriebs- und -Supportleistungen – entwickeln sich zunehmend zu wissenschaftsnahen Dienstleistungszentren: Betreiber von Infrastrukturen (Werkzeugen für die Forschung), Ermöglicher von Forschung, Forschungsberater und Weiterbilder. Die HPC-Rechenzentren müssen an der Wissenschaft teilhaben, um adäquate Dienstleistungen für die Forschung erbringen zu können, z. B. durch Kooperationen im „Computational and Data-Intensive Science and Engineering“. Dies alles erfordert adäquate Maßnahmen zur Personalentwicklung sowie zur Förderung von Personal.

Ein bislang in der Förderlandschaft kaum adressiertes Problem ist die Pflege und Weiterentwicklung von Software über Forschergenerationen und insbesondere Hardwarearchitekturwechsel hinweg. Hierbei ist Software (wie Hardware) als Infrastrukturkomponente für die Forschung anzusehen. Die Pflege dieser langlebigen und kostenintensiven Softwareinfrastrukturen (software stewardship) ist bisher weder finanziell, noch organisatorisch gesichert.

Lösungsansätze:

Es sollten künftig generell Gesamtpakete – nicht nur Investitionen (Hardware), sondern auch Betriebskosten und personelle Dienstleistungen/Projekte (Brainware [6]) – gefördert werden.

Die Beschaffung und insbesondere auch Pflege von Software infrastrukturellen Charakters muss ebenfalls Bestandteil solcher Gesamtpakete sein.

Das Finanzierungsmodell der Ebene 1 (Total Cost of Ownership) sollte auch auf den Ebenen 2 und 3 angewendet werden.

Die Etablierung von „Computational Science and Engineering“ als Fachcommunity, auch in der Form eines eigenen Fachkollegiums innerhalb der DFG, geht aus Sicht des ZKI in die richtige Richtung.

Es sollten, in Analogie zur Forschungsförderung an wissenschaftlichen Bibliotheken, von den Fachwissenschaften der Nutzergruppen unabhängige Förderlinien zur Unterstützung wissenschaftsnaher HPC-Betriebs- und -Kompetenzzentren etabliert werden.

(c) *Koordinierung von Förderungen und Kompetenzen*

Förderungen auf den unteren Versorgungsebenen – Tier 2 und 3 – erfolgen derzeit weitgehend unkoordiniert (Ausnahmen z.B. Baden-Württemberg, Hessen). Unausgeschöpfte Synergiepotenziale und unzulängliche Fairness beim Zugang der Wissenschaftler zu HPC-Kapazitäten sind die Folge. Auch die Komplementarität der Rechnerarchitekturen und Kompetenzverbände sind suboptimal ausgeprägt. Ziel einer deutschen HPC-Strategie muss es sein, der Wissenschaft, unabhängig vom Standort adäquaten Zugang zu den HPC Ressourcen zu gewährleisten, die zur Durchführung ihrer Forschung notwendig sind. Dies umfasst sowohl Hardware-Ressourcen als auch Schulungs- und Beratungskompetenz.

Lösungsansätze:

Im Rahmen der Begutachtung sollte gefordert werden, dass sich die HPC-Versorgung – zumindest auf Tier 2 – in den Ländern jeweils an einem abgestimmten Konzept orientiert.

Die Koordinierung der Förderung von HPC-Systemen sowie der Bereitstellung von Kompetenzen innerhalb Ebene 2 – insbesondere auch länderübergreifend – sollte verstärkt werden.

Länderübergreifende Nutzung sollte durch gesetzliche Regelungen erleichtert werden.

Eine bessere Zusammenarbeit zwischen Ebene 2 und 3 sollte dazu beitragen, standortunabhängig Servicequalitäten zu homogenisieren.

Die Erfahrungen und Anforderungsszenarien der Mitglieder des ZKI – und hier insbesondere im Rahmen der Aktivitäten des ZKI-Arbeitskreises Supercomputing – im Bereich des HPC-Betriebs- und -Anwendungsumfelds sollten in die bedarfsgerechte Intensivierung der Koordinierung der HPC-Versorgung adäquat einfließen.

(d) *Beschaffungszyklen und Planungssicherheit*

Beschaffungszyklen von vier bis sechs Jahren sind im Bereich des Hochleistungsrechnens inzwischen nicht mehr angemessen. Zeitlich gestreckte Beschaffungen und über mehrere Zentren umlaufende Beschaffungen sind Hilfskonstrukte, um diesen Mangel abzumildern. Bei einem kürzeren Beschaffungszyklus könnte flexibler auf den sich wandelnden (qualitativen) Bedarf eingegangen werden (z. B. Auslegung der Rechenknoten, Ergänzung von Spezialprozessoren). Planungssicherheit und damit auch Planungsmöglichkeit erfordert auch, dass mehrere Beschaffungszyklen gesichert sind.

Lösungsansätze:

Die Investitionszyklen sollten generell verkürzt werden, z. B. durch kurzfristige Architekturauswahl innerhalb einer längerfristigen strategischen Planung.

Es sollte eine Halbierung der Beschaffungszykluszeit, bei ungefähr gleich bleibender Standzeit, ermöglicht werden.

Es sollte eine größere Planungssicherheit erreicht werden, indem mehrere Beschaffungszyklen in Folge zugesichert werden.

4 Quellen

1. Deutsche Forschungsgemeinschaft:
Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme.
Empfehlungen der Kommission für IT-Infrastruktur für 2011-2015, Bonn 2010
2. Bundesministerium für Bildung und Forschung:
Die Säulen des Supercomputing.
Höchstleistungsrechnen gibt Antworten auf schwierigste Fragen unserer Zeit.
Berlin, Bonn, 2011
3. Wissenschaftsrat:
Strategische Weiterentwicklung des Hoch- und Höchstleistungsrechnens in Deutschland.
Positionspapier, 2012
4. Gauss Centre for Supercomputing:
High Performance Computing in Deutschland.
Gedanken zur Fortschreibung eines nationalen HPC-Versorgungs- und Nutzungskonzeptes.
Berlin, 15.09.2011
5. Hannes Hartenstein, Thomas Walter und Peter Castellaz:
Aktuelle Umsetzungskonzepte der Universitäten des Landes Baden-Württemberg für Hochleistungsrechnen und datenintensive Dienste.
Praxis der Informationsverarbeitung und Kommunikation, Band 36, Heft 2, Mai 2013, De Gruyter
6. Christian Bischof, Dieter an Mey, Christian Iwainsky:
Brainware for green HPC.
Computer Science – Research and Development, Volume 27, Issue 4, November 2012, Springer