

2012

Strategische
Weiterentwicklung des
Hoch- und
Höchstleistungsrechnens
in Deutschland
Positionspapier

INHALT

Vorbemerkung	5
Kurzfassung	7
A. Anlass und Darstellung	9
A.I Institutionelle Entwicklungen	9
A.II Trends in den Anwendungen	10
A.III Technologieentwicklung	13
A.IV Förderpolitische Entwicklungen	14
B. Analyse und Empfehlungen	18
B.I CSE und HPC in der Wissenschaft	18
I.1 Institutionelle Struktur	19
I.2 Förderinstrumente	21
I.3 Aus- und Weiterbildung	21
B.II HPC-Infrastruktur	22
II.1 HPC-Leistungspyramide	24
II.2 Finanzierung	27
II.3 Koordination	30
II.4 Betrieb	32
Anhang	35
Abkürzungsverzeichnis	37
Literaturverzeichnis	38

Vorbemerkung

Bereits in seinen „Empfehlungen zur zukünftigen Nutzung von Höchstleistungsrechnern“ aus dem Jahr 2000 stellte der Wissenschaftsrat fest, dass die Verfügbarkeit von Infrastrukturen für das Höchstleistungsrechnen „ein entscheidender Standortfaktor im internationalen wissenschaftlichen und industriellen Wettbewerb“ sei. |¹ Höchstleistungsrechnen, kurz HPC (*High Performance Computing*) |², ist der entscheidende Schlüssel für den Fortschritt im Bereich *Computational Science and Engineering* (CSE) |³. Zahlreiche Disziplinen sind an der Methodenentwicklung für die Nutzung von HPC-Infrastrukturen beteiligt. Darüber hinaus hat sich CSE als Querschnittsdisziplin etabliert. Dabei ist CSE in einem umfassenden Sinn als Simulationswissenschaft zu verstehen, die die Entwicklung von Modellen, Algorithmen und Software gleichermaßen beinhaltet und eng miteinander verzahnt. In diesem Sinne ist CSE in einer steigenden Anzahl von Anwendungsfeldern als unverzichtbares Fundament wissenschaftlichen Erkenntnisgewinns neben Theorie und Experiment etabliert. Hierfür wird vielfach das Bild einer „dritten Säule“ der Wissenschaft verwendet.

HPC spielt innerhalb von CSE-Anwendungen eine doppelte Rolle: Zum einen ist HPC eine Infrastruktur, derer sich CSE-Anwender bedienen. Zum anderen stellt die Methodenforschung für die Nutzung von HPC einen Teilbereich von CSE dar, der angesichts der Geschwindigkeit technologischen Fortschritts stetiger Weiterentwicklung bedarf. Bereits 2000 hat der Wissenschaftsrat von der „gro-

|¹ Wissenschaftsrat (2000), S. 233.

|² In Deutschland hat sich – aufgrund von Unterschieden insbesondere organisatorischer und finanzieller Art – die Differenzierung von Hoch- und Höchstleistungsrechnern etabliert und wird daher auch im vorliegenden Positionspapier verwendet. *High Performance Computing* wird hingegen als Bezeichnung für beide Leistungsklassen von Rechensystemen gebraucht.

|³ Das Fachgebiet wird noch nicht einheitlich bezeichnet. Es werden u. a. auch die weitgehend synonym gebrauchten Begriffe *Computational Science*, *Computational Engineering*, *Computational Engineering Science* oder *Simulation Science* verwendet. Im deutschen Sprachraum sind zudem die Bezeichnungen Wissenschaftliches Rechnen, Simulationstechnik oder Simulationswissenschaften gebräuchlich.

ßen Entwicklungsdynamik“ |⁴ gesprochen, der der gesamte Bereich des HPC unterworfen sei. Mehr als eine Dekade später hat dieser Befund nach wie vor Bestand. Die technologische Entwicklung im Bereich des HPC ist mit unverminderter Geschwindigkeit vorangeschritten.

Angesichts der steigenden Bedeutung von CSE für die Wissenschaften und des rasanten technologischen Fortschritts im Bereich HPC ebenso wie aktueller politischer, technologischer und struktureller Entwicklungen (vgl. Kap. A) sieht der Wissenschaftsrat Bedarf für eine erneute Befassung mit den Strukturen der Versorgung von Forschenden mit Kapazitäten des HPC sowie mit der Förderung der für diese Ressourcen erforderlichen Methodenkompetenz. Zu diesem Zweck hat der Wissenschaftsrat im Juli 2011 eine Arbeitsgruppe eingesetzt, die das Positionspapier in zwei Sitzungen vorbereitet hat. In ihr wirkten auch Sachverständige mit, die nicht Mitglieder des Wissenschaftsrates sind. Ihnen weiß sich der Wissenschaftsrat zu besonderem Dank verpflichtet. Zudem dankt er den Experten, die im Rahmen von Befragungen bereitwillig für Auskünfte und Einschätzungen zur Verfügung standen.

Der Wissenschaftsrat hat dieses Positionspapier am 27. Januar 2012 verabschiedet.

|⁴ Wissenschaftsrat (2000), S. 233.

Kurzfassung

- _ Die Bedeutung von *Computational Science and Engineering* (CSE) für den wissenschaftlichen Erkenntnisgewinn und der dafür notwendigen Infrastrukturen für *High Performance Computing* (HPC) wächst weiterhin. CSE-Methodik findet in einer steigenden Zahl von Wissenschaftsbereichen Anwendung. Deutschland hat in der Methodenkompetenz für diesen Bereich eine starke Position im internationalen wissenschaftlichen Wettbewerb. Diese Position sollte durch entsprechende Förderprogramme und verstärkte Aktivitäten in der Aus- und Weiterbildung nachhaltig gefestigt werden.
- _ Die Verfügbarkeit von HPC-Infrastrukturen ist für die Wettbewerbsfähigkeit Deutschlands in zahlreichen Wissenschaftsfeldern und für die Weiterentwicklung der CSE-Methodik unverzichtbar. Der Bedarf wird auf absehbare Zeit steigen. Die erforderlichen Ressourcen sollten durch HPC-Kompetenzzentren bereitgestellt werden, die sich durch eine enge Integration von leistungsfähigen Rechnerinfrastrukturen, Methodenkompetenz, Anwenderberatung sowie Forschung und Lehre auszeichnen. Auf den verschiedenen Rechenleistungsstufen sollte es jeweils mehrere, sich in Rechnerarchitektur und Methodenkompetenz gegenseitig ergänzende Zentren geben. Diese sollten durch koordinierte und zeitlich gestaffelte Beschaffungen für eine optimale Versorgung der HPC-Anwender Sorge tragen.
- _ Die aus verschiedenen Rechenleistungsstufen bestehende Leistungspyramide mit wenigen Zentren an der Spitze und einem in den unteren Leistungsklassen breiter werdenden Angebot an Rechensystemen hat sich bewährt und sollte weiter beibehalten werden. Hierfür ist ein nachhaltiges finanzielles Engagement von Bund und Ländern erforderlich.
- _ Die derzeitigen Finanzierungsmodalitäten für Beschaffungen von Rechensystemen der beiden obersten Leistungsklassen haben für die nachhaltige Versorgung der Wissenschaft und Wirtschaft mit leistungsfähigen und ausdifferenzierten Rechenressourcen signifikante Nachteile. Bund und Länder sollten alternative Modelle prüfen, die nachhaltige Finanzierungssicherheit für (Re)Investitionen und Betriebskosten bieten, den Aufbau von sich ergänzenden Strukturen fördern sowie für alle deutschen Wissenschaftlerinnen und

Wissenschaftler Zugang zu den Systemen der obersten Leistungsklassen sicherstellen.

- _ Die Selbstkoordination der Betreiber von Rechenzentren im Gauss Centre for Supercomputing (GCS) sowie in der Gauß-Allianz (GA) hat sich prinzipiell bewährt. Auf Ebene der GA müssen allerdings derzeit noch bestehende Defizite in der Koordination, der Governance und der Transparenz von Prozessen behoben werden, damit sie für die Erfüllung ihrer Aufgaben adäquat gerüstet ist.
- _ Auf europäischer Ebene sollte eine engere Abstimmung im Vorfeld von Beschaffungen der höchsten Leistungsklasse erfolgen, um auch hier eine aus mehreren sich ergänzenden Zentren bestehende Infrastruktur zu schaffen.

A. Anlass und Darstellung

A.1 INSTITUTIONELLE ENTWICKLUNGEN

Der Wissenschaftsrat hat sich in der Vergangenheit wiederholt und in verschiedenen Kontexten mit der Thematik einer angemessenen Versorgung mit Infrastrukturen im Bereich *High Performance Computing* (HPC) auf nationaler und europäischer Ebene befasst. Dabei hat er sowohl übergeordnete Empfehlungen zu Struktur und Bedingungen einer solchen Versorgung abgegeben |⁵ als auch zu konkreten Beschaffungsvorhaben von Rechensystemen der Länder Stellung genommen. |⁶ Zentrales Gremium hierfür war seit seiner Einrichtung auf Grundlage der „Empfehlung zur künftigen Nutzung von Höchstleistungsrechnern“ aus dem Jahr 2000 der Nationale Koordinierungsausschuss zur Beschaffung und Nutzung von Höchstleistungsrechnern. Mit der Einführung der gemeinsamen Förderung von Forschungsbauten an Hochschulen einschließlich Großgeräten durch Bund und Länder im Jahr 2007 (vgl. A.IV) ging die Aufgabe, Förderempfehlungen für konkrete Beschaffungen an Hochschulen auszusprechen, auf den Ausschuss für Forschungsbauten über.

Etwa zeitgleich etablierten sich in Deutschland zwei wichtige neue Akteure im Bereich des HPC. Im April 2007 wurde das „Gauss Centre for Supercomputing“ (GCS) als Zusammenschluss der drei nationalen Höchstleistungsrechenzentren in Stuttgart, Jülich und Garching gegründet. Im Dezember 2008 erfolgte dann die Gründung der Gauß-Allianz (GA), eines Verbundes auf Ebene der Hochleistungsrechner. Beide Initiativen gingen auf die Empfehlung einer durch das BMBF eingesetzten Expertenkommission zurück, eine strategische HPC-Allianz

|⁵ Vgl. Wissenschaftsrat (1995), Wissenschaftsrat (2000), Wissenschaftsrat (2004).

|⁶ Vgl. bspw.: Wissenschaftsrat (2003), Wissenschaftsrat (2006).

ins Leben zu rufen. |⁷ Ein wesentliches Ziel dieser Gründungen, insbesondere des GCS, war es, eine einheitliche und koordinierte Positionierung Deutschlands in internationalen Abstimmungs- und Wettbewerbsprozessen zu ermöglichen. |⁸ Der Bedarf hierfür war nicht zuletzt durch die Empfehlungen des Wissenschaftsrates zur Einrichtung eines europäischen Höchstleistungsrechners aus dem Jahr 2004 in das Bewusstsein der deutschen HPC-Community sowie der politischen Entscheidungsträger gerufen worden.

Als Reaktion auf die Gründungsinitiativen empfahl der Nationale Koordinierungsausschuss in einem vorläufigen Bericht über seine eigenen Tätigkeiten im November 2007 |⁹ dem Wissenschaftsrat, über die zukünftigen Aufgaben des Ausschusses erst nach Abschluss der Gründungsphase der HPC-Allianz zu entscheiden, wenn sich erwiesen habe, welche Aufgaben diese wahrnehmen könne. |¹⁰ Daher bat der Wissenschaftsrat Bund und Länder, ihm zu gegebener Zeit über die Entwicklung der Allianz zu berichten, damit auf dieser Grundlage eine Entscheidung über die Zukunft des Ausschusses getroffen werden könne.

Zu diesem Zweck fand im Februar 2011 eine Anhörung des Wissenschaftsrates statt, bei der der Bund, die drei Sitzländer der Höchstleistungsrechner sowie Vertreter des GCS und der GA durch eine eigens eingerichtete Ad-hoc-Gruppe bestehend aus Anwendungs- und Methodenwissenschaftlern befragt wurden. Dabei wurde von allen Beteiligten ein Bedarf an einer umfassenderen und raschen Befassung mit der Thematik des HPC durch den Wissenschaftsrat identifiziert. Im Ergebnis schlug die Ad-hoc-Gruppe daher dem Wissenschaftsrat die Einrichtung einer Arbeitsgruppe vor, die sich mit der strategischen Weiterentwicklung des *High Performance Computing* (HPC) in Deutschland befassen solle. Der Wissenschaftsrat folgte dieser Empfehlung im Juli 2011.

A.II TRENDS IN DEN ANWENDUNGEN

Computational Science and Engineering (CSE): Viele Disziplinen verwenden fachspezifische Modelle, die auf Simulationsmethodik angewiesen sind und die

|⁷ Vgl.: European Media Laboratory GmbH (2006).

|⁸ Neben dem Engagement des GCS in PRACE (vgl. Kap. A.IV) ist in diesem Kontext die Mitwirkung der GA in der „European Grid Infrastructure“ (EGI) zu nennen. EGI wurde im Februar 2010 mit dem Ziel gegründet, europaweit Zugänge zu Rechenressourcen der Tier2-Ebene über die Nutzung von Grid-Technologien zu schaffen. EGI befindet sich noch im Aufbau.

|⁹ Mit diesem Bericht folgte der Nationale Koordinierungsausschuss der Empfehlung aus dem Jahr 2000, nach einer ausreichenden Erprobungsphase erneut über die Aufgaben und Funktionen des Ausschusses zu beraten.

|¹⁰ Vgl.: Wissenschaftsrat (2008), S. 287.

se entsprechend disziplinspezifisch weiterentwickeln. Darüber hinaus hat sich CSE als Querschnittsdisziplin etabliert. Die Anwendungsfelder, in denen CSE-Methodik Verwendung findet, sind vielfältig, ihre Zahl wächst beständig. Sie reichen von naturwissenschaftlicher Grundlagenforschung (z. B. im Bereich Magnetismus, Licht und Materie) über geowissenschaftliche Simulationen (z. B. der Tsunami-Ausbreitung und -Überflutung) bis hin zu Anwendungen in der Lebenswissenschaften (z. B. Modellierung der Lunge zur Verbesserung der künstlichen Beatmung von Patienten) und den Ingenieurwissenschaften (z. B. Untersuchungen zur Lärmentwicklung in Triebwerksdüsen). |¹¹

In den meisten Anwendungsfeldern basiert das Simulationsmodell auf einer zugrunde liegenden Theorie und verwendet oft Verfahren der numerischen Mathematik, um Lösungen und Vorhersagen zu berechnen, die über keine geschlossene mathematische Form verfügen. Der Einsatz eines Rechners und die mit ihm gewonnenen Daten sind somit durch spezifische Hypothesen und Fragestellungen motiviert, zu denen sie Antworten liefern sollen. Daneben finden sich mit der Astronomie oder der Systembiologie (z. B. in Verbindung mit *High-Throughput Genome Sequencing*) Wissenschaftsbereiche, in denen aufgrund einer stark datengetriebenen Forschung Datenanalyse und Theoriebildung enger miteinander verschränkt sind und nicht zwingend in zeitlicher Aufeinanderfolge stehen.

HPC-Leistung in den Anwendungen: Grundsätzlich sind im Bereich der Anwendungen drei Arten der Nutzung von HPC-Ressourcen zu unterscheiden: Als *Capability Computing* wird das Rechnen einzelner komplexer Probleme oder Modelle bezeichnet, die sich idealerweise durch gute Parallelisierbarkeit auszeichnen. Anwendungen des *Capability Computing* sind häufig zeitkritisch im Sinne eines Wettbewerbs um Durchbrüche in der Grundlagenforschung (z. B. *Grand Challenges*). Aufgrund dieses Erfordernisses eines schnellen Outputs sowie der Komplexität der Fragestellungen werden für Anwendungen des *Capability Computing* Rechensysteme der obersten Leistungsklasse benötigt. Unter *Capacity Computing* wird dagegen das gleichzeitige Rechnen vieler Probleme bzw. Modelle oder Modellparameter verstanden, die weniger als einzelne sondern vielmehr in der Summe Anforderungen an die Leistungsfähigkeit der Rechensysteme stellen. Anwendungen des *Capacity Computing* (z. B. Monte-Carlo-Simulationen) zielen primär auf einen hohen Durchsatz ab, erfordern aber eine ausreichende Rechenleistung, um unverhältnismäßige Wartezeiten der Anwender zu ver-

|¹¹ Darstellungen von Anwendungsfeldern finden sich bspw. in der Broschüre „Supercomputing at the Leading Edge“ (Gauss Centre for Supercomputing [2010]), in der Publikation „Forschung im HLRN-Verbund 2011“ (Konrad-Zuse-Zentrum für Informationstechnik Berlin; Regionales Rechenzentrum für Niedersachsen, Hannover [2011]) sowie im zweimal jährlich durch das GCS herausgegebenen Magazin „inSiDE – Innovatives Supercomputing in Deutschland“ (Gauss Centre for Supercomputing [2011]).

meiden. Zunehmende Bedeutung erlangt daneben das *Real-Time-Computing* (z. B. für Visualisierungen), das sich von den beiden anderen Formen von Anwendungen insbesondere durch Interaktivität abhebt.

HPC findet insbesondere in einer Reihe von natur- und ingenieurwissenschaftlichen Disziplinen Anwendung. In Feldern wie der Strömungsdynamik oder der Klimaforschung dienen Simulationen mittels HPC als unverzichtbare Methode des Erkenntnisgewinns. Der Bedarf an Rechenkapazität ist dabei im Prinzip unbegrenzt. Denn höhere Rechenleistungen ermöglichen eine feinere Diskretisierung, höherdimensionale Daten und komplexere Parameterräume und versprechen somit immer weitere Fortschritte und neue Ergebnisse. Doch die ansteigende Datenmenge über immer detailliertere Modelle sorgt nicht allein für einen exponentiell wachsenden Rechenbedarf, hinzu kommt die zunehmende Komplexität der Anwendungen. Diese entwickeln sich von der reinen Berechnung spezifischer Probleme zur umfassenden Simulation komplexer Phänomene bis hin zu Multiskalensystemen, die unterschiedliche Modelle auf verschiedenen Skalen miteinander koppeln (z. B. in der Materialforschung oder der Chemie). Aufbauend auf quantitativ hinreichend genauen Modellen können z. B. auch Optimierungsaufgaben in komplexen Netzwerkstrukturen oder inverse Probleme (approximativ) gelöst werden.

Neben diesen etablierten Anwendungsfeldern erlangt HPC auch in anderen Wissenschaftsfeldern zunehmende Bedeutung. Zu nennen sind hier bspw. die Verwendung von Agenten-basierten Modellen zur Simulation komplexer sozialer Phänomene in den Sozialwissenschaften (z. B. Katastrophenforschung), komplexe Simulationen im Bereich der Logistik (z. B. Verkehrsflüsse), die Digitalisierung und Erschließung kulturwissenschaftlicher Güter im Rahmen der *Digital Humanities* sowie Probleme, die durch Graphen beschrieben werden, etwa in der Bioinformatik oder in der Analyse sehr großer Datenmengen. Zudem ist durch die Weiterentwicklung von interaktiv-explorativen und Echtzeit-Anwendungen mit einer Vielzahl neuer Anwendungsbereiche, bspw. in der Medizin, zu rechnen.

Industrie: Nicht nur die Wissenschaft selbst, auch die Industrie setzt in stark zunehmenden Maßen CSE-Methoden ein. Komplexe Simulationsmodelle und auch Echtzeitsimulationen spielen in einer steigenden Anzahl von Wirtschaftsbereichen wie bspw. in der Fertigung und Logistik oder im Finanzsektor eine tragende Rolle bei der Entwicklung neuer Produkte und Dienstleistungen. Insbesondere im Kontext präkompetitiver, explorativer Untersuchungen nutzen Unternehmen hierfür auch öffentlich zugängliche Rechensysteme in Kooperation mit universitären Partnern. Immer größere Bedeutung erlangt daneben der Bereich *Business Intelligence (Big Data Analytics)*, also die Unterstützung von operativen und strategischen Entscheidungen durch systematische Analysen elektronisch aufbereiteter Daten. Die Mehrheit dieser Anwendungen ist derzeit nicht

auf höchste Rechenleistungen angewiesen, was sich auch in einer verstärkten Nutzung von *Cloud Computing* zeigt. Doch auch in diesem Bereich sind entsprechende Weiterentwicklungen denkbar, die zu einem erhöhten Bedarf an HPC-Ressourcen führen könnten. Gegenläufig gibt es Forschungstrends, die für industrielle Anwendungen (z. B. in Produktion und Logistik) nach Lösungen mit dezentralen, selbststeuernden Systemen suchen. Trotz des Trends zur Dezentralisierung steigt aber auch in diesen Anwendungsfeldern der Bedarf an HPC, da die Komplettsimulation der im Modell abzubildenden selbststeuernden Objekte, die bspw. für die Evaluierung der Netzwerke erforderlich ist, hochkomplex ist.

Datenqualität: Durch die steigende Komplexität der Anwendungen in Wissenschaft und Industrie wächst nicht nur der Bedarf an Rechenkapazität. Die Anwendungen zeichnen sich auch durch ein stark wachsendes Datenaufkommen und damit einhergehend einen großen Bedarf an Speicherplatz aus. Fragen nach der Qualitätssicherung und dem richtigen Management für solche Datenmengen erhalten damit immer größere Bedeutung. Der Aspekt der Datensicherheit spielt daher eine zunehmend wichtigere Rolle, auch in finanzieller Hinsicht.

A.III TECHNOLOGIEENTWICKLUNG

Technologie: Das 1965 von Gordon Moore formulierte empirische „Moore’s Law“ besagt, dass sich die Transistordichte auf einem Computerchip etwa alle 18 Monate verdoppelt. Dieser Langzeittrend der technologischen Entwicklung führte in den letzten Jahrzehnten zu immer höher getakteten und damit schnelleren Prozessoren. Eine weitere Steigerung der Transistordichte stößt jedoch mittlerweile an physikalische Grenzen wie bspw. Wärmeeffekte auf dem Chip. Stattdessen wird nunmehr die Zahl der Prozessoren, sog. *Cores*, auf einem Chip beständig gesteigert, was zu zunehmend massiver Parallelität der Rechner führt. |¹² Die maximal erzielbare aggregierte Rechenleistung steigt daher auch weiterhin exponentiell an und es ist davon auszugehen, dass sich dieses Wachstum bis mindestens 2020 fortsetzen wird. Für den bevorstehenden Leistungssprung von Rechnern mit einer Leistung im Bereich von Petaflops (10^{15} *floating-point operations per second*), die von heutigen Spitzensystemen erreicht werden, zu Exaflops (10^{18} Flops) wird bspw. eine Zeitspanne von fünf bis acht Jahren angesetzt. Zugleich wächst in der Langfristperspektive die Notwendigkeit, durch verstärkte Grundlagenforschung im Bereich HPC Lösungen für die Zeit „post-

|¹² Ein Parallelrechner der Spitzenklasse umfasst heute um die 100 Tsd. *Cores*. Hochrechnungen gehen davon aus, dass bereits in fünf bis acht Jahren über eine Million *Cores* pro Rechner verbaut werden.

Moore“ zu entwickeln, wenn sich die Transistordichte auf dem Chip der Ebene einzelner Atome annähern wird.

Systemkomplexität: Die steigende Anzahl von Prozessoren wird begleitet von einem starken Wachstum der Speichergrößen. Dabei kommt zunehmend eine verteilte und somit komplexe Speicherarchitektur zum Einsatz. Verstärkt werden außerdem neue Prozessortypen wie z. B. Grafikprozessoren verwendet, die sowohl als spezialisierte Co-Prozessoren als auch als General-Purpose-Prozessoren eingesetzt werden. Es lässt sich sowohl im Rückblick wie auch im Ausblick auf sich bereits abzeichnende Entwicklungen feststellen, dass die Architekturen moderner Hoch- und Höchstleistungsrechner, in der Folge aber auch Rechner auf Instituts- und Arbeitsgruppenebene, immer komplexer werden. Zudem steigt der Energiebedarf für Strom und Klimatisierung und damit die Betriebskosten. |¹³ Damit nimmt die Gefahr einer ineffizienten Nutzung dieser kostenaufwändigen Infrastruktur durch nicht ausreichend geschulte Anwender stetig zu.

Alternative Trends: Große Aufmerksamkeit innerhalb und außerhalb der HPC-Community erfährt derzeit das *Cloud Computing*. Hierbei werden virtualisierte Rechnerressourcen (Prozessoren, Speicher, Plattenplatz) und standardisierte Services durch externe, momentan i. d. R. kommerzielle Anbieter wie Amazon, IBM oder Microsoft über ein verbrauchsabhängiges Abrechnungsmodell zur Verfügung gestellt. Die flexible Bereitstellung der Ressourcen und die möglichen Einsparungen durch *pay-per-use* bieten Vorteile auch für wissenschaftliche Nutzerinnen und Nutzer. Derzeit richten sich die Angebote in dem noch stark bezuschussten Einsteigermarkt aber schwerpunktmäßig an Firmenkunden im Bereich *Business Intelligence*. Die Möglichkeit, *Cloud Computing* verstärkt auch für HPC-Anwendungen im Bereich des *Capacity Computing* zu nutzen, sollte weiter ausgelotet und aktiv verfolgt werden.

A.IV FÖRDERPOLITISCHE ENTWICKLUNGEN

Die Gründungen der Gauß-Allianz (GA) und insbesondere des Gauss Centres for Supercomputing (GCS, vgl. A.I) wurden von Seiten der Politik intensiv begleitet und auch gefördert.

|¹³ Aufgrund der immer höheren Rechenleistung nimmt der Gesamtenergiebedarf eines Höchst- bzw. Hochleistungsrechners zu. Die Energieeffizienz der Rechner, gemessen in Watt pro Flop, wird dagegen ständig verbessert. Die stetig steigenden Energiekosten können damit aber nicht vollständig kompensiert werden, so dass es zum Anstieg der Betriebskosten kommt.

Gauss Centre for Supercomputing (GCS, Zentren der sog. Tier 1): Die Finanzierung des GCS beruht auf einem im Jahr 2008 geschlossenen Verwaltungsabkommen des Bundes (vertreten durch das BMBF) sowie der drei Sitzländer Baden-Württemberg, Bayern und Nordrhein-Westfalen. |¹⁴ In der Umsetzung des Abkommens bauen diese die drei Standorte des GCS mit Projektfördermitteln in Höhe von bis zu 400 Mio. Euro für Entwicklung, Investitionen und Betrieb aus, wovon bis zu 50 % vom Bund getragen werden. Die Sitzländer tragen jeweils die Hälfte der Kosten für den jeweils eigenen Standort. Diese Finanzierung endet vier Jahre nach der dritten Ausbaustufe (nach derzeitigem Planungsstand Ende 2015). Eine über diesen Zeitpunkt hinausgehende Finanzierung des GCS wurde nicht vereinbart.

Gauß-Allianz (GA, Zentren der sog. Tier 2): Auch auf der Ebene der in der Gauß-Allianz |¹⁵ zusammengeschlossenen Rechner ist der Fortbestand der derzeit geltenden Finanzierungsmodalitäten noch unklar. Dies gilt angesichts zunehmender Finanzengpässe in den Ländern allgemein für diese wesentlich durch die Länder finanzierte Infrastruktur. Darüber hinaus gilt es aber in besonderer Weise für die aktuell einzige Möglichkeit einer Finanzierungsbeteiligung des Bundes bei Hochleistungsrechnern an Hochschulen. Diese besteht im Rahmen des auf Artikel 91b GG basierenden Forschungsbautenprogramms, in dem Bund und Länder jeweils hälftig die Investitionskosten großer Forschungsinfrastrukturen oder Bauten für spezifische Forschungsvorhaben an Hochschulen finanzieren. Für Anträge auf Beschaffung eines Hochleistungsrechners wurde eine eigene „programmatisch-strukturelle Linie“ (PSL) geschaffen, die sich durch spezifisch auf diese Anträge zugeschnittene Kriterien auszeichnet. Insbesondere wurde auf das für die sonstige Förderung in diesem Programm geltende Kriterium einer kohärenten Forschungsprogrammatisierung, die der Nutzung des Forschungsbaus bzw. der Infrastruktur zugrunde liegen soll, verzichtet. Diese PSL war bei ihrer Einführung sowohl zeitlich als auch finanziell begrenzt: Im Zeitraum von 2009 bis 2014 sollten insgesamt 100 Mio. Euro aus dem allgemeinen Programmfördertopf für Bewilligungen von Rechneranträgen eingesetzt werden können. Bereits mit den im Juli 2011 ausgesprochenen Empfehlungen für die Förderung von Forschungsbauten beginnend im Jahr 2012 und damit zwei Jahre vor Auslaufen der Befristung der PSL sind diese Mittel nahezu restlos ver-

| ¹⁴ Die drei Rechenzentren sind in Stuttgart, Garching und Jülich angesiedelt.

| ¹⁵ Zur GA gehören aktuell die folgenden Hochleistungsrechenzentren: RWTH Aachen, norddeutscher HLRN-Verbund (vertreten durch die Betreiberzentren Konrad-Zuse-Zentrum Berlin und Regionales Rechenzentrum für Niedersachsen), TU Dresden, Steinbuch-Zentrum am KIT, TU Darmstadt, RZG der MPG, Deutscher Wetterdienst und Deutsches Klimarechenzentrum. Die letzteren drei sind außeruniversitär und werden vollständig von ihren jeweiligen Mutterorganisationen finanziert. Zusätzlich sind assoziierte Mitglieder der GA: Uni Frankfurt, Uni Paderborn, Uni Erlangen, Uni Köln, DESY und DFN.

braucht (vgl. B.II.2.a). |¹⁶ Die Gemeinsame Wissenschaftskonferenz (GWK) hat daher auf ihrer Sitzung im November 2011 den Beschluss gefasst, die Förderung von Vorhaben nach der PSL „Hochleistungsrechner“ zunächst ohne zeitliche und finanzielle Begrenzung als Teil des Forschungsbautenprogramms fortzuführen. Dazu werden Hochleistungsrechner-Vorhaben nach ihren bisherigen Förderkriterien bewertet und durch den Wissenschaftsrat im Rahmen einer Gesamtreihung mit Anträgen aus der thematisch offenen Förderung zusammen priorisiert. Dieser Beschluss soll gelten, bis auf Grundlage des vorliegenden Positionspapiers die Beratungen der GWK zu einer langfristigen Finanzlösung fortgesetzt werden.

PRACE: Auch auf europäischer Ebene sind weitere Abstimmungsprozesse für die zukünftige Versorgung europäischer Nutzer mit HPC-Kapazitäten erforderlich. Diese Ressourcen werden derzeit auf europäischer Ebene durch den Verein „Partnership for Advanced Computing in Europe“ (PRACE) zur Verfügung gestellt. PRACE – eine verteilte Forschungsinfrastruktur, die aus den Empfehlungen des „European Strategy Forum on Research Infrastructures“ (ESFRI) hervorgegangen ist – ist als internationaler Verein nach belgischem Recht mit derzeit 21 Einrichtungen als Mitgliedern |¹⁷ organisiert. Innerhalb von PRACE stellen Höchstleistungsrechenzentren verschiedener Nationen, die als „hosting members“ bezeichnet werden, jeweils Rechenzeit der höchsten Leistungsklasse im Gegenwert von 100 Mio. Euro für einen Zeitraum von fünf Jahren für europäische Nutzer bereit. Das GCS fungiert als eines von derzeit vier solcher „hosting members“ und übernimmt zudem eine führende Rolle in der Projektadministration von PRACE. Die Europäische Union (EU) fördert die Vorbereitung und Implementierung von PRACE seit 2008 mit rund 48 Mio. Euro, beteiligt sich aber nicht an den Investitions- und Betriebskosten der Rechner, diese werden durch die jeweiligen Sitznationen bestritten. Die EU-Förderung erfolgt im Rahmen des 7. Forschungsrahmenprogramms, dessen Laufzeit bis zum Jahr 2013 angesetzt ist. Die Konsultationen für die Gestaltung des nachfolgenden 8. Forschungsrahmenprogramms („Horizon 2020“) laufen derzeit. Die Bundesregierung hat im Rahmen dieses Prozesses in ihrer Stellungnahme zum Grünbuch der Kommission auf die besondere Bedeutung von Schlüsseltechnologien und Forschungsinfrastrukturen – und damit auch von e-Infrastrukturen wie PRACE – für die Stärkung der Wissenschaftsbasis Europas und die Entwicklung des Eu-

| ¹⁶ Es wurden mit Stand Juli 2011 rund 97,2 Mio. Euro in der PSL bewilligt.

| ¹⁷ Die Mitgliedseinrichtungen von PRACE vertreten folgende Nationen: Bulgarien, Deutschland, England, Finnland, Frankreich, Griechenland, Irland, Italien, Niederlande, Norwegen, Österreich, Polen, Portugal, Serbien, Spanien, Schweden, Schweiz, Tschechien, Türkei, Ungarn und Zypern (Stand Juli 2011). Pro Nation kann nur eine Einrichtung Mitglied in PRACE werden.

ropäischen Forschungsraums (EFR) hingewiesen |¹⁸ und zudem in ihrem Leitlinienpapier für das kommende Rahmenprogramm die weitere Unterstützung des Netzwerkes PRACE als ein vorrangiges Forschungsfeld benannt. |¹⁹

|¹⁸ Bundesregierung (2011a).

|¹⁹ Bundesregierung (2011b), S. 15.

B. Analyse und Empfehlungen

B.I CSE UND HPC IN DER WISSENSCHAFT

Computational Science and Engineering (CSE), verstanden in einem umfassenden Sinn als integrierendes *Co-Design* von Modellen, Algorithmen und Software, ist nicht nur ein Werkzeug für Forschung in den verschiedenen Anwendungswissenschaften, sondern verfügt selbst über Grundlagenforschungscharakter. Derzeit ist Deutschland in diesem Bereich, der auch als Methodenkompetenz bezeichnet wird und als solche auch den Bereich der HPC-Methodik umfasst, sehr gut aufgestellt. Deutsche Wissenschaftlerinnen und Wissenschaftler erzielen mit Hoch- und Höchstleistungsrechnern Ergebnisse, die international Beachtung finden. Deutschland ist hier führend in Europa und übernimmt in einzelnen Gebieten wie bspw. komplexen Multi-Physik-Problemen in ingenieurwissenschaftlichen Anwendungen und Multi-Skalen-Modellen in Physik, Materialwissenschaften und Biochemie sogar weltweit eine Vorreiterrolle. Dies trägt maßgeblich zu Europas Wettbewerbsfähigkeit im Bereich HPC bei. Angesichts der teilweise erheblichen Vorsprünge anderer Nationen im Hinblick auf die verfügbare Rechenleistung (vgl. B.II) ist es insbesondere die Stärke in der Methodenkompetenz, die es Europa ermöglicht, international anschlussfähig zu bleiben.

Wichtig ist, dass Rechenkapazität und Methodenkompetenz gleichwertig stark entwickelt und – bei wachsendem Bedarf – ausgebaut werden. Am Beispiel Chinas etwa zeigt sich, dass sehr hohe Investitionen allein in die Hardware (vgl. B.II) nicht notwendigerweise zu beachtenswerten wissenschaftlichen Resultaten führen, wenn diese nicht durch die entsprechende Förderung von Methodenkompetenz unterfüttert werden. Umgekehrt kann führende Methodenkompetenz langfristig nicht ohne Rechner an der weltweiten Leistungsspitze und ohne ausreichende Rechenkapazität für das gesamte Anwendungsspektrum aufrechterhalten werden.

In Anbetracht dieser Ausgangsposition und der enorm wachsenden Bedeutung von CSE in den Anwendungen sieht der Wissenschaftsrat Chancen, die Vorreiterposition Deutschlands und Europas auf dem Gebiet der Methodenkompetenz durch gezielte Investitionen und Maßnahmen nachhaltig zu festigen. Angesichts der Vielschichtigkeit der Anwendungsfelder, die problemabhängig Rechenleistung auf verschiedenen Niveaus erfordern, müssen die grundlegenden Methoden des CSE wie auch deren Nutzung in den Anwendungswissenschaften weiterentwickelt werden. Hochkomplexe und bisher nur wenig bearbeitete Fragestellungen wie bspw. Optimierung, inverse Probleme oder Multiskalenmodellierung stellen gänzlich neue Herausforderungen an die Entwickler und Anwender von CSE-Methoden.

Durch die zunehmende Bedeutung von CSE in einem breiter werdenden Anwendungsspektrum (siehe A.II) entsteht ein rasch wachsender Bedarf an Rechenkapazität für *Capacity Computing* auf HPC-Leistungsniveau. Der Bedarf für *Capability Computing* an der Leistungsspitze wächst angesichts des weltweiten Wettbewerbs um wissenschaftliche Durchbrüche ebenfalls mit stark zunehmender Rate. So kann man mit heutigen Spitzenrechnern die Dynamik von Molekülen mit 10^5 Atomen für einige Mikrosekunden simulieren. Um anschlussfähig an die internationale Spitzenforschung in diesem für die Materialforschung oder für die biomedizinische Forschung essentiellen Grundlagenbereich zu bleiben, wären aber Simulationen von komplexen Molekülen mit mehr als 10^7 Atomen im Millisekundenbereich erforderlich. Der hieraus resultierende Rechenbedarf übersteigt die prognostizierten Leistungssteigerungen der Hardwaretechnologie um mehrere Größenordnungen. Die Situation in Anwendungen wie Klimaforschung oder Geophysik ist ähnlich. Dabei ist der Rechereinsatz in diesen Bereichen unverzichtbar: Im Bereich der Funktionsanalyse von Proteinen werden durch Moleküldynamiksimulationen, z. T. in Kombination mit quantenchemischen Rechnungen, wesentlich teurere Laborexperimente (z.B. in der Pharmaforschung) vermieden; bei Umweltthemen gibt es gar keine Alternative zu rechnergestützten Simulationen.

1.1 Institutionelle Struktur

CSE als Fachgebiet: CSE etabliert sich immer stärker als eigenständiges Forschungs- und Fachgebiet. Diese Entwicklung sollte sich auch in entsprechenden institutionellen Strukturen niederschlagen. Bei der Deutschen Forschungsgemeinschaft (DFG) ist CSE derzeit kein eigenständiges Fachgebiet mit entsprechendem Fachkollegium, was dazu führt, dass Förderanträge im Normalverfahren von nicht spezifischen Fachkollegien behandelt werden (z. B. Informatik oder die entsprechende Fachwissenschaft). Der Wissenschaftsrat empfiehlt daher der DFG, die Möglichkeit der Einrichtung eines solchen Fachkollegiums zu prüfen.

HPC-Kompetenzzentren: Eine wesentliche Stärke der weltweit führenden amerikanischen HPC-Forschung liegt in der frühen und nachhaltigen Einrichtung interdisziplinärer *Laboratories* wie bspw. in Oak Ridge oder Los Alamos begründet. Der Wissenschaftsrat hat bereits 1995 die Einrichtung von interdisziplinären „Kompetenznetzwerken“ auch für Deutschland empfohlen. |²⁰ Er begrüßt daher entsprechende Initiativen, wie bspw. die Gründung des GCS, die enge Zusammenarbeit des Jülicher Rechenzentrums mit der RWTH Aachen in der „Jülich Aachen Research Alliance – Section High Performance Computing“ (JARA-HPC), die enge Vernetzung des LRZ in Garching mit den bayerischen Universitäten im „Kompetenznetzwerk für Hoch- und Höchstleistungsrechnen in Bayern“ (KONWIHR), die Einrichtung von themenspezifischen *Solution Centers* in Kooperation mit der ortsansässigen Industrie am HLRS in Stuttgart, die Gründung des „Kompetenzzentrums für Höchstleistungsrechnen Bremen“ (BremHLR) sowie universitäre Zentren zur Bündelung von entsprechenden Kompetenzen wie bspw. Simtech an der Universität Stuttgart oder IWR an der Universität Heidelberg. Die genannten Beispiele zeichnen sich in besonderer Weise dadurch aus, dass in diesen Initiativen Rechnerexpertinnen und -experten, CSE-Methodenwissenschaftlerinnen bzw. -wissenschaftler und Vertreterinnen bzw. Vertreter der Anwendungsdisziplinen eng miteinander zusammenarbeiten.

Nach Auffassung des Wissenschaftsrates sollten derartige Initiativen in Anlehnung an das amerikanische Modell zum Ausgangspunkt für die Einrichtung und den Ausbau von nachhaltig finanzierten HPC-Kompetenzzentren in Deutschland gemacht werden, die 1) leistungsfähige Rechnerinfrastrukturen und Methodenkompetenz am selben Ort bündeln („*Racks and Brains*“), 2) intensiv mit Anwendungswissenschaftlerinnen und -wissenschaftlern in geographischer Nähe zusammenarbeiten (gleichzeitig aber auch für nichtlokale Projekte offen sind), und 3) eng an benachbarte Universitäten oder außeruniversitäre Forschungseinrichtungen angebunden sind und damit auch einen wesentlichen Beitrag zur Aus- und Weiterbildung im Bereich CSE leisten. Um die genannten Aufgaben erfüllen zu können, sollten die Kompetenzzentren auch vereinfachte Zugänge zu den Rechensystemen für Nachwuchsgruppen eröffnen.

HPC-Kompetenzzentren sollten auf verschiedenen Leistungsstufen und damit sowohl im Bereich des *Capability Computing* als auch im *Capacity Computing* angesiedelt sein. Dabei sollte es auf jeder Leistungsstufe mehrere, hinsichtlich Rechnerarchitekturen und Methodenkompetenz sich ergänzend ausgerichtete Zentren geben. Auf diese Weise können die unterschiedlichen Anforderungsprofile verschiedener Wissenschaften bzw. Nutzergruppen optimal berücksichtigt

|²⁰ Wissenschaftsrat (1995), S. 63.

werden. Außerdem ermöglicht eine Mehrzahl von Zentren eine Abstimmung der (auf ca. fünf Jahre ausgerichteten) Beschaffungszyklen in den Zentren, so dass zu jedem Zeitpunkt jede der großen wissenschaftlichen Communities Zugang zu einem aktuellen (d. h. vor höchstens zwei Jahren installierten) Spitzenrechner hat.

HPC-Kompetenzzentren dieser Art sollten nachhaltig etabliert und regelmäßig hinsichtlich der wissenschaftlichen Qualität der Methodenforschung, der Zahl und Qualität der Anträge auf Nutzung, der Kompetenz in Anwenderberatung und Rechnerbetrieb sowie ihres Nutzen/Kosten-Verhältnisses evaluiert werden.

1.2 Förderinstrumente

Für die nachhaltige Stärkung und den weiteren Ausbau der spezifischen deutschen und europäischen Stärken im Bereich der CSE/HPC-Methodik ist die Förderung mit öffentlichen Mitteln ein unerlässlicher Baustein. Das Bundesministerium für Bildung und Forschung (BMBF) fördert bereits schwerpunktmäßig HPC-Projekte zur Entwicklung und Anwendung neuer Algorithmen und Softwarearchitekturen. |²¹ Ein Beispiel dieser Förderung ist die Fördermaßnahme „HPC-Software für skalierbare Parallelrechner“ innerhalb des Programms „IKT 2020“, die auf die Entwicklung innovativer Methoden zur Nutzung neuer Rechnerarchitekturen zielt. Die DFG hat aktuell ein Schwerpunktprogramm für die Förderung von Softwaremethoden für Exascale-Computing ins Leben gerufen.

Der Wissenschaftsrat begrüßt diese Initiativen und empfiehlt, die Methodentwicklung für HPC auch langfristig zu fördern. Vor dem Auslaufen bestehender Programme ist daher – unter Berücksichtigung des bereits existierenden Projektportfolios – rechtzeitig Vorsorge zu treffen. Dabei sollte vermieden werden, die Programme thematisch zu eng auszurichten. Offene *Calls*, die den Grundlagencharakter der CSE/HPC-Methodik betonen, die wissenschaftliche Qualität zum obersten Förderkriterium machen und zu einer Verbreiterung der Anwender-Community beitragen, sind aus Sicht des Wissenschaftsrates ein geeignetes Instrument für eine nachhaltige Stärkung des Bereichs.

1.3 Aus- und Weiterbildung

Computational Science und Engineering (CSE) ermöglicht neue Erkenntnisse in einer breiten Palette von Wissenschaftsbereichen. In der Ausbildung des wissenschaftlichen Nachwuchses spielt CSE hingegen in vielen dieser Bereiche, insbesondere auch in den Ingenieur- und Naturwissenschaften trotz des unbestritten hohen Anwendungspotentials, bislang nur eine untergeordnete Rolle. Dies wird

|²¹ Vgl.: Bundesministerium für Bildung und Forschung (2011), S. 29.

der heutigen Bedeutung von CSE für die Forschung in diesen Gebieten nicht gerecht. Zum einen drohen hohe Kosten und Zeitverluste, wenn eine technologisch hochkomplexe Infrastruktur durch nicht hinreichend dafür ausgebildete Forschende genutzt wird. Zum anderen sind für die Forschung nicht ausgeschöpfte Erkenntnisreserven zu befürchten, wenn Nachwuchswissenschaftlerinnen und -wissenschaftler aufgrund fehlender CSE-Kompetenzen vor der Nutzung von HPC-Ressourcen zurückschrecken oder diese aufgrund eines fehlenden Bewusstseins für deren Einsatzmöglichkeiten gar nicht erst in Betracht ziehen.

Der Wissenschaftsrat bekräftigt daher nachdrücklich seine bereits im Jahr 2000 ausgesprochene Empfehlung, „Methoden der Nutzung von Höchstleistungsrechnern stärker in die bestehenden Studiengänge der Natur- und Ingenieurwissenschaften sowie der Informatik und Mathematik zu integrieren“. |²² Dies umfasst Themen der Modellierung ebenso wie die Bereiche Simulation, Algorithmen, Programmierung und Software. Zudem hält der Wissenschaftsrat es im Interesse eines weiteren Ausbaus und einer nachhaltigen Sicherung der in Europa vorhandenen Stärken im Bereich der CSE/HPC-Methoden- und Anwendungskompetenz weiterhin für erforderlich, „spezialisierte Studiengänge mit interdisziplinärer Ausrichtung“ im Bezugsfeld von CSE/HPC-Methodenkompetenz und Anwendungswissenschaften auf Master-Ebene zu schaffen. |²³ Die wenigen bereits etablierten Studienangebote dieser Art stellen aus Sicht des Wissenschaftsrates begrüßenswerte Pilotprojekte dar. Einen weiteren Ausbau sieht er als notwendig an. Zu denken ist dabei auch an Studiengänge im Bereich *Digital Humanities*.

Neben der Etablierung von nichtkonsekutiven Master-Studiengängen ist aus Sicht des Wissenschaftsrates außerdem die Schaffung von Weiterbildungsangeboten zur gezielten Schulung von HPC-Nutzern voranzutreiben. Hierbei sollten vor allem die HPC-Kompetenzzentren (vgl. I.1) eine zentrale Rolle spielen.

B.II HPC-INFRASTRUKTUR

Ziel einer nationalen Versorgungsstruktur unter Einbezug auch der europäischen Ebene muss die ausreichende Versorgung aller Anwender, ein effizienter Ressourceneinsatz und die effektive Nutzung der vorhandenen Infrastruktur sein. Nicht nur angesichts des bereits jetzt feststellbaren Rechenbedarfs in etablierten Anwendungsgebieten von HPC (wie z. B. der Moleküldynamik, vgl. B.I),

|²² Wissenschaftsrat (2001), S. 246.

|²³ Ebd.

die für den Anschluss an den wissenschaftlichen Fortschritt auf immer komplexere Simulationen angewiesen sind, sind nach Ansicht des Wissenschaftsrates zur Erreichung dieses Ziels fortgesetzte Investitionen in HPC-Infrastruktur erforderlich (vgl. Kap. II.2.a). Auch die weiter voranschreitende Etablierung von CSE als eigenes Fachgebiet und der damit einhergehende Aus- und Weiterbildungsbedarf werden einen Ausbau der HPC-Kapazitäten erforderlich machen. Nicht zuletzt werden auch neue Anwendungsfelder, bspw. in den Finanzmarktanalysen, zum Anstieg des Rechenbedarfs beitragen.

Hinweise auf die weltweit verfügbaren HPC-Kapazitäten liefert – nach einzelnen Ländern aufgeschlüsselt – die sogenannte TOP-500-Liste, die zweimal jährlich die 500 leistungsfähigsten Rechner der Welt und deren akkumulierte Gesamtleistung listet. Laut der Fassung der TOP-500-Liste vom Juni 2011 |²⁴ stehen 6 % der leistungsfähigsten Rechner der Welt in Deutschland. Damit liegt Deutschland in etwa gleichauf mit Frankreich (5,0 %), Japan (5,2 %) und Großbritannien (5,4 %) und weltweit auf dem dritten Platz. |²⁵ Allerdings verfügt das auf dem zweiten Platz liegende China mit 61 Systemen bereits über einen mehr als doppelt so hohen Anteil (12,2 %). Dominiert wird die Liste der TOP 500 von den USA, die allein 51 % der weltweit leistungsfähigsten Rechner betreiben. Den bislang größten Anteil an der Rechenleistung erreichte Deutschland mit 11 % Ende der 90er Jahre, demgegenüber hat sich der Anteil in der aktuellen Ausgabe der Liste fast halbiert (6 %). |²⁶ Insgesamt zeigt die Entwicklung im Zeitverlauf, dass die technologische Entwicklung so schnell voranschreitet, dass Rechner relativ zur Leistungsspitze schon nach wenigen Jahren stark zurückfallen.

Betrachtet man – unter Berücksichtigung der Einschränkungen in der Vergleichbarkeit nationaler Beschaffungsstrategien |²⁷ – die Anteile an den leistungsfähigsten Rechnern in Relation zu monetären Kennzahlen, so zeigt sich, dass Deutschland in Bezug auf die Vergleichswerte unterdurchschnittlich abschneidet. Das Bruttoinlandsprodukt (BIP) der USA bspw. liegt viermal höher als das deutsche, die USA erreichen aber eine fast neunmal höhere aggregierte Re-

|²⁴ Die TOP-500-Liste wird seit 1993 zweimal jährlich durch die Universitäten Mannheim und Tennessee in Kooperation mit dem National Energy Research Scientific Computing Centre (NERSC) in Berkeley, California, erstellt. Auf der Grundlage des „Linpack Benchmark“, eines Programms zur Messung der Leistungsfähigkeit von Computern, listet sie die 500 schnellsten Rechner der Welt: <http://www.top500.org/lists/-2011/06> (zuletzt abgerufen am 08.11.2011).

|²⁵ Die Prozentangaben beziehen sich auf die Anzahl der Rechensysteme in der Liste.

|²⁶ Dabei ist zu beachten, dass es sich nicht um eine kontinuierliche Abnahme handelt, da es zum Zeitpunkt größerer Beschaffungen (z. B. „Jugene“ am JRC) zu einem kurzzeitigen Anstieg der Anteile kommt. Dennoch lässt sich seit Ende des Jahres 2000 eine abnehmende Tendenz in den Anteilen Deutschlands feststellen. Der Einfluss der geplanten weiteren Beschaffungen für das GCS hierauf bleibt abzuwarten.

|²⁷ Bspw. ist ein Teil der amerikanischen Investitionen in HPC-Infrastruktur militärisch motiviert.

chenleistung als Deutschland. |²⁸ China kommt mit einem nur etwa 1,5fach höheren Bruttoinlandsprodukt auf eine mehr als dreimal so hohe Rechenleistung wie Deutschland. |²⁹ Dabei liegt der Anteil der Aufwendungen für Forschung und Entwicklung (*Gross Domestic Expenditure on Research and Development*, GERD) am BIP mit 1,7 % in China deutlich niedriger als in Deutschland (2,78 %), während die USA mit 2,79 % einen nur unwesentlich höheren Anteil des GERD am BIP zu verzeichnen haben. |³⁰

Die TOP-500-Liste ist aber lediglich als Indikator für die in verschiedenen Ländern verfügbare Rechenleistung zu verstehen. Investitionen in HPC-Infrastrukturen dürfen selbstverständlich nicht durch einen Wettbewerb um Platzierungen in internationalen Rankings oder Listen getrieben sein, sondern sollten vielmehr auf einer nachfrageorientierten Beschaffungsstrategie basieren. Diese darf nicht nur die Spitzenrechner der obersten Leistungsklasse in den Blick nehmen, sondern muss die gesamte Breite der Anwendungen berücksichtigen.

II.1 HPC-Leistungspyramide

Leistungsstufen: Zur Beschreibung der Versorgungsstruktur mit HPC in Deutschland und Europa wird häufig das Bild einer „Leistungspyramide“ verwendet. Damit soll dargestellt werden, dass das auf die Anzahl der Systeme bezogene schmale Angebot an Rechnern der obersten Leistungsklassen auf einer breiteren Basis der Versorgung mit Rechenkapazitäten der unteren Leistungsklassen fußt. Die verschiedenen Leistungsklassen innerhalb der Pyramide werden als „Ebenen“ bzw. *Tiers* voneinander unterschieden. Derzeit wird diese Pyramide in Deutschland durch die Staffelung von Rechenclustern in Instituten und regionalen Rechenzentren an Hochschulen und Forschungseinrichtungen („Tier 3“), Landeshochleistungsrechnern und fachspezifischen Rechnern z. B. für Klimaforschung („Tier 2“, lose organisiert in der GA) sowie die nationalen Höchstleistungsrechenzentren des GCS („Tier 1“) realisiert.

Erneuerungszyklen: Da die relative Leistung von HPC-Rechner-Systemen im Hinblick auf die jeweils aktuell verfügbare Spitzenrechenleistung sehr schnell abnimmt, sind Rechenzentren aller Tiers auf regelmäßige Ersatzbeschaffungen angewiesen. Durch die Beschleunigung dieser Erneuerungszyklen und zeitlich gestaffelte Beschaffungen der Rechenzentren lässt die technologische Unter-

|²⁸ Bezogen auf die in der TOP-500-Liste erfassten Rechner-Systeme. Die Angaben zur aggregierten Rechenleistung basieren auf den Rpeak-Werten (*theoretical peak performance*) der Liste.

|²⁹ Quelle: Eurostat (Vergleich BIP Deutschland und USA, Bezugsjahr 2011 [Prognose]), UNCTADStat (Vergleich BIP Deutschland und China, Bezugsjahr 2010 [im Falle Chinas Schätzwert]).

|³⁰ Quelle: OECD, Main Science and Technology Indicators Database (Bezugsjahr 2009).

scheidbarkeit der Leistungsstufen – zumindest in der jeweils aktuellen Momentaufnahme – nach. Unterschiedliche Nutzungsprofile der verschiedenen Ebenen sowie organisatorische Aspekte sprechen aber für eine Beibehaltung des Pyramidenmodells. Denn die unterschiedlichen Leistungsklassen von Rechen-systemen erfüllen unterschiedlich geartete Versorgungsaufträge, beginnend mit der lokalen Bereitstellung von Rechenleistung für Einzelwissenschaftlerinnen bzw. -wissenschaftler und Forschergruppen über die Funktion als Rechenzentrum für Institute, Universitäten oder ganze Bundesländer bis hin zu den auf nationale bzw. europäische Spitzenprojekte ausgerichteten Systemen. Um diese Versorgungsaufträge konstant auf einer Leistungsstufe erfüllen zu können, ist nach derzeitigem Stand des technologischen Fortschritts ein Erneuerungszyklus von etwa fünf Jahren erforderlich. Dies gilt gleichermaßen für alle Leistungsstufen. Darüber hinaus hat sich gezeigt, dass für international wettbewerbsfähige Spitzenforschung in vielen Wissenschaftsfeldern Zugang zu einem jeweils dem aktuellsten Stand der Technik entsprechenden System benötigt wird, das nicht älter als zwei Jahre sein sollte. Dies betrifft im Wesentlichen die Rechen-systeme der Tiers 1 und 2. Die ständige Verfügbarkeit eines solchen Systems kann durch koordinierte und zeitversetzte Beschaffungen mehrerer Zentren derselben Leistungsstufe erreicht werden, wie sie derzeit bereits innerhalb des GCS praktiziert wird. Diese Vorgehensweise bietet gegenüber der gleich getaketen Ersatzbeschaffung an nur einem Zentrum den wesentlichen Vorteil, dass durch sich ergänzende Rechnerarchitekturen und arbeitsteilige Strukturen der Methodenkompetenz und Anwenderberatung mehrere Nutzer-Communities optimal versorgt werden können.

Anforderungsprofile: Mit den unterschiedlichen Versorgungsaufträgen gehen unterschiedliche Anforderungsprofile einher. Während bspw. für eine Gruppe von Anwendungen die absolute Spitzenleistung aufgrund zeitkritischer Forschung notwendig ist (*Capability Computing*, vgl. A.II), sind für andere ein ausgewogenes Verhältnis von Leistung und Energieverbrauch oder schnelle Rüstzeiten entscheidend (*Capacity Computing*, *Real-time Computing*). Den Anforderungskriterien entsprechen unterschiedliche Rechnerarchitekturen. Massiv parallele Rechner eröffnen z. B. ein großes Potential für hohe Skalierbarkeit, stellen aber gleichzeitig auch hohe Anforderungen an die verwendete Software. Rechner mit niedrigeren Parallelitätsgraden und insbesondere solche mit großem gemeinsamen Hauptspeicher sind einfacher in der Programmierung, aber limitiert in der Skalierbarkeit. Dabei umfassen viele Forschungsprojekte mit HPC-Anwendung innerhalb der Projektlaufzeit mehrere unterschiedliche Anforderungsprofile. So werden häufig während der Entwicklung eines Projektes Testreihen auf niedrigeren als der final angestrebten Leistungsklasse durchgeführt. Zudem weisen Problemstellungen häufig auch einzelne Teilmodelle oder Algorithmen auf, die aufgrund schlechter Skalierbarkeit auf den unteren Leistungsklassen ein besseres Kosten-Leistung-Verhältnis erzielen.

Mehrstufigkeit und Innovation: Die Übertragung innovativer Methoden zwischen den unterschiedlichen Leistungsklassen ist ein wesentlicher Aspekt in der Weiterentwicklung der Methodenkompetenz. Dabei zeigt sich, dass Innovationen zum einen von der Spitze der Pyramide in die Breite der unteren Leistungsklassen ausstrahlen, indem Methoden für eine effiziente Nutzung der Rechner und die Optimierung eingesetzter Software für breitere Anwendungsfelder übernommen werden. Zum anderen aber entstehen besonders innovative Ansätze oft auch durch die Erprobung unkonventioneller Vorgehensweisen auf unteren Leistungsebenen. Diese finden im Anschluss Eingang in die Leistungsspitze, in der wegen der hohen Nutzungskosten i. d. R. weniger riskante und schon erprobte Methoden Anwendung finden. Dies beinhaltet auch die Möglichkeit, dass Rechner der Tiers 1 und 2 an einem Zentrum betrieben werden.

Mehrstufigkeit in der Ausbildung: Die HPC-Leistungspyramide übernimmt auch in der Aus- und Weiterbildung eine wichtige Funktion, da sie es erlaubt, Nachwuchswissenschaftlerinnen und -wissenschaftler sukzessive an das Potential sehr leistungsstarker Rechner heranzuführen. Diese Funktion wird im Zuge der Stärkung von CSE als eigene Disziplin zunehmend wichtiger werden.

Die nationale Versorgungsstruktur als Pyramide mit wenigen sehr leistungsstarken Zentren an der Spitze und einem in den unteren Leistungsklassen breiter werdenden Angebot an Rechensystemen hat sich aufgrund der oben angeführten Funktionen der Leistungsstufen aus Sicht des Wissenschaftsrates für die Bedienung unterschiedlicher Anforderungsprofile bewährt und sollte beibehalten werden.

HPC-Infrastruktur in Europa: Die europäische Ebene der Leistungspyramide („Tier 0“) wird im Rahmen von PRACE durch einen virtualisierten Zusammenschluss von nationalen Tier 1-Zentren gebildet. Internationale Anschlussfähigkeit setzt in den Anwendungswissenschaften wie bei der Weiterentwicklung der Methodenkompetenz, wie oben beschrieben, Zugänge zu den jeweils leistungsfähigsten verfügbaren Rechensystemen voraus. Ziel eines europäischen Höchstleistungsrechnerverbundes sollte es daher sein, europäischen Forschenden zu jedem Zeitpunkt mindestens ein System der internationalen Spitzenklasse zugänglich zu machen. Nach Auffassung des Wissenschaftsrates sind auch hierfür wie auf der nationalen Ebene mehrere Zentren erforderlich, die in enger Abstimmung und zeitlich gestaffelt Rechensysteme mit sich ergänzenden Architekturen beschaffen und entsprechend spezifische Methodenkompetenz aufbauen und weiterentwickeln.

II.2.a Beschaffungen

Die auf die Leistungsspitze bezogene relative Leistung von HPC-Rechensystemen nimmt aufgrund der rasanten technologischen Entwicklung sehr schnell ab. Um eine konstante Versorgung der Anwender mit Rechenkapazitäten einer bestimmten Leistungsklasse sicherzustellen, sind daher regelmäßige Ersatz- und Neubeschaffungen erforderlich. Derzeit hat sich hierfür ein Erneuerungszyklus von etwa fünf Jahren als realistisch einzuplanen herausgestellt (vgl. auch Kap. II.1). Die unterschiedlichen Ebenen der Leistungspyramide unterliegen dabei derzeit unterschiedlichen Finanzierungsmodi. Tier 3-Systeme werden durch die jeweiligen Sitzländer oder aus den Budgets der betreffenden Forschungseinrichtungen finanziert. Zudem gibt es für Beschaffungsvorhaben bis zu 5 Mio. Euro die Möglichkeit einer Förderung über das Programm „Forschungsgroßgeräte nach Art. 91b GG“ über die DFG auf Grundlage einer nichtvergleichenden Einzelbegutachtung. Auch im Fall der Tier 2-Rechenzentren liegt die Finanzierung wesentlich in der Verantwortung der Länder. Für Rechner an Hochschulen kann über das wettbewerbliche Förderprogramm „Forschungsbauten“ auf der Grundlage von Art. 91b GG eine hälftige Finanzierungsbeteiligung des Bundes erreicht werden (vgl. A.IV). Die fachlich dedizierten Rechensysteme (z. B. das Deutsche Klimarechenzentrum) sind Teil außeruniversitärer Forschungseinrichtungen und werden über deren Budgets finanziert. Die Tier 1-Zentren schließlich werden über eine Projektförderung anteilig vom Bund und den jeweiligen Sitzländern getragen (vgl. A.IV).

Aufgrund der hohen Bedeutung, die der Verfügbarkeit von HPC-Ressourcen für die Wissenschaft und die Wettbewerbsfähigkeit Deutschlands zukommt, hält der Wissenschaftsrat ein fortgesetztes finanzielles Engagement von Bund und Ländern für die Sicherung und Weiterentwicklung der Leistungsfähigkeit der HPC-Leistungspyramide auf allen Leistungsstufen für erforderlich. Der Wissenschaftsrat ist der Auffassung, dass das jetzige System der Finanzierungsstrukturen für HPC-Ressourcen nicht optimal geeignet ist, eine leistungsfähige und ausdifferenzierte Versorgung deutscher Forschender mit Rechenressourcen langfristig zu gewährleisten. Auf absehbare Zeit ist mit einem wachsendem Bedarf und nicht mit fallenden Kosten^{|31} in diesem Bereich zu rechnen. Daher bedarf es eines langfristigen Finanzierungskonzeptes, das aus Sicht des Wissenschaftsrates mit der projektförmigen Finanzierung der Tier 1-Zentren des GCS und der wettbewerblichen Förderung der Tier 2-Systeme derzeit nicht gegeben

^{|31} Aufgrund der konstant steigenden Energiepreise gilt dies auch bei verbesserter Energieeffizienz der Rechner.

ist, da – aufgrund der kurzen „Halbwertszeit“ der relativen Leistung eines Rechners – die Zentren auf regelmäßige Ersatzbeschaffungen angewiesen sind. Zudem erfordert der Betrieb von Hoch- und Höchstleistungsrechnern technisches Know-how und Kompetenzen in der Anwenderberatung, die nur an wenigen Standorten vorhanden sind bzw. sein können und deren Auf- und Ausbau langfristige Planungssicherheit erfordert.

Der Wissenschaftsrat bittet daher den Bund und die Länder, mögliche Alternativen zu den jetzigen Finanzierungsmodellen für die Zentren der Leistungsstufen Tier 1 und Tier 2 zu prüfen. Ein neu zu entwickelndes Finanzierungsmodell sollte aus Sicht des Wissenschaftsrates auf allen Ebenen eine Ausstattung auf dem jeweils aktuellen Stand der Technik erlauben und nach Möglichkeit folgende Anforderungen erfüllen:

- _ nachhaltige Sicherstellung der erforderlichen (Re)Investitionen,
- _ langfristige Sicherstellung der Deckung der Betriebskosten,
- _ Förderung des Aufbaus sich ergänzender Rechnerarchitekturen und Methodenkompetenzen, sowie
- _ Sicherung des offenen Zugangs aller Nutzer bundesweit auch zu den Zentren der Leistungsstufe Tier 2 nach den Anforderungen der jeweiligen Anwendungen entsprechend der bereits bestehenden Zugangsmöglichkeiten zu den Zentren der Leistungsstufe 1.

Ein Finanzierungsmodell, das diesen Anforderungen genügt, würde eine bedarfsgerechte Versorgung der Nutzer mit HPC-Ressourcen sicherstellen und zudem den Aufbau interdisziplinärer Kompetenzzentren (vgl. I.1) nachhaltig und langfristig fördern. Der Aufbau spezifischer Methodenkompetenz, eine enge Koordination der Zentren untereinander sowie regelmäßige Evaluationen auch bei den Beschaffungen würden die Kosteneffizienz der geförderten Zentren erheblich verbessern und damit letztlich auch zu einem effizienteren Mitteleinsatz in der Förderung von Hoch- und Höchstleistungsrechnen durch Bund und Länder führen.

Da die ursprünglich bereitgestellten Mittel in der programmatisch-strukturellen Linie (PSL) des Förderprogramms „Forschungsbauten“ von Bund und Ländern ausgeschöpft sind (vgl. A.IV), besteht auf Ebene der Tier 2-Rechner akuter Bedarf an einer Lösung der daraus resultierenden Finanzierungsprobleme. Der Wissenschaftsrat hält daher im ersten Schritt den Beschluss der Gemeinsamen Wissenschaftskonferenz (GWK), die zeitliche Befristung und finanzielle Deckung der PSL aufzuheben, für den richtigen Weg, um diesem akuten Bedarf zu entsprechen. Der Wissenschaftsrat begrüßt, dass diese Regelung bereits für die kommende Förderphase 2013 Anwendung findet. Darüber hinaus besteht aber nach Ansicht des Wissenschaftsrates – auch aufgrund der inhärenten Differenz

zwischen einmalig zu errichtenden Forschungsbauten, die auf einen spezifischen Forschungsauftrag zugeschnitten sind, und regelmäßig zu erneuernden Rechnern, die für eine Vielzahl unterschiedlicher Anwendungen zur Verfügung stehen – weiterhin die Notwendigkeit, zu nachhaltig tragfähigen Finanzierungskonzepten im Sinne der oben formulierten Anforderungen zu kommen. Der Wissenschaftsrat empfiehlt daher Bund und Ländern, zeitnah Finanzierungsmodelle für Tier 1- und Tier 2-Rechner zu entwickeln, die diese Anforderungen erfüllen und die Weiterentwicklung von HPC-Kompetenzzentren (vgl. I.1) in Deutschland fördern.

Für die unterhalb der GCS- und GA-Zentren liegende Ebene der Tier 3-Rechner hält der Wissenschaftsrat die jetzige bedarfs- und projektorientierte Praxis der Beschaffung und Finanzierung im Rahmen von Einzelfallentscheidungen hingegen für angemessen. Eine übergeordnete Koordinierung von Beschaffungen ist hier nicht erforderlich, da diese Systeme einen thematisch oder regional sehr engen Versorgungsauftrag erfüllen.

II.2.b Betriebskosten

Der Betrieb eines HPC-Rechners ist wegen des hohen Energiebedarfs für Strom und Kühlung und der erforderlichen personellen Ausstattung mit erheblichen Kosten verbunden. Diese erreichen auf allen Leistungsstufen über die übliche Rechnerstandzeit von fünf Jahren gerechnet häufig die Höhe der Beschaffungskosten oder übersteigen diese sogar. Für eine umfassende Betrachtung der Finanzierungsnotwendigkeiten im Bereich HPC ist es daher unerlässlich, die *total-cost-of-ownership* angemessen zu berücksichtigen. Derzeit werden Betriebskosten im Fall des GCS im Rahmen der Projektförderung durch Bund und Sitzländer gemeinsam getragen. Auf den Tier 2- und Tier 3-Ebenen hingegen ist die Finanzierung der Betriebskosten alleinige Aufgabe des Sitzlandes bzw. der betreibenden Institution.

Der Wissenschaftsrat hat bereits im Jahr 2000 auf Schwierigkeiten hingewiesen, die mit einer Finanzierung von Betriebskosten über Nutzungsgebühren einhergehen. |³² Die hierfür angeführten Argumente haben weiterhin Geltung. Insbesondere stehen Marktmechanismen, die bei einer Gebührenfinanzierung greifen würden, der wissenschaftsgeleiteten Steuerung des Auf- und Ausbaus von HPC-Infrastruktur entgegen.

Einen besonderen Sachverhalt stellt die Nutzung von HPC-Ressourcen durch die Wirtschaft dar. Während der Zugang im Rahmen von Förderprojekten i. d. R. keine Schwierigkeiten verursacht, bestehen für die Wirtschaft rechtliche und

|³² Vgl.: Wissenschaftsrat (2001), S. 247-250.

organisatorische Hindernisse bei der Nutzung von Rechenzentren außerhalb einer Projektförderung. Um bessere Zugangsmöglichkeiten auch für die Wirtschaft zu schaffen, sollten Modelle erprobt werden, die eine Refinanzierung der durch die Nutzung entstehenden Kosten erlauben (bspw. durch die Gründung von Spin-offs).

Um einen offenen Zugang der Nutzer entsprechend ihrer Anforderung sicherzustellen, empfiehlt der Wissenschaftsrat den Zuwendungsgebern sowie den Betreibern von HPC-Zentren daher, sich über nicht-monetäre Wege des Ausgleichs von Betriebskosten auszutauschen. Beispiele hierfür sind das im Rahmen von PRACE auf europäischer Ebene praktizierte juste-retour-Prinzip oder das im Rahmen der „Ocean Facilities Exchange Group“ für große Forschungsschiffe etablierte „Bartering“-Verfahren |³³. Dabei ist zentral, dass ein solcher Ausgleich der wissenschaftsgeleiteten Vergabe von Rechenzeit nicht entgegensteht (vgl. II.4). Die Entscheidung für ein entsprechendes Ausgleichsmodell hängt wesentlich davon ab, wie Beschaffungen in Zukunft finanziert werden sollen (vgl. II.2.a), und sollte daher im Kontext mit den hierfür erforderlichen Abstimmungen beraten werden.

II.3 Koordination

Eine ausdifferenzierte und arbeitsteilige Struktur der Versorgung mit HPC-Ressourcen erfordert neben einem koordinierten Finanzierungskonzept eine enge Abstimmung der jeweiligen Betreiber von Rechenzentren. Diese müssen sich sowohl im Vorfeld von Beschaffungen über benötigte Rechnerarchitekturen und angemessene Erneuerungsintervalle verständigen, als auch zu gemeinsam getragenen Kriterien für die Vergabe von Rechenzeit und die Bewertung der wissenschaftlichen Qualität der Nutzungsanträge kommen. Auf nationaler Ebene wurden mit dem GCS und der GA Strukturen für diese Koordination unter den verschiedenen Betreibern von Hoch- und Höchstleistungsrechnern in Deutschland geschaffen. Wesentlich hierfür war die Empfehlung einer Expertenkommission im Auftrag des BMBF, eine „HPC-Allianz“ als Dachorganisation für die HPC-Aktivitäten in Deutschland zu schaffen. |³⁴ Vorgeschlagen wurde ein gemeinnütziger, rechtsfähiger Verein mit eigenem Haushalt, der über folgende Gremien verfügen sollte: die Mitgliederversammlung, den Aufsichtsrat als Repräsentanz der Zuwendungsgeber, einen ehrenamtlichen Vorstand inklusive einer bzw. eines Vorsitzenden, einen hauptamtlichen Geschäftsführer für das operative Geschäft sowie einen wissenschaftlichen Beirat, in dem die Hauptnutzer der Rechner eine beratende Funktion erfüllen sollten.

|³³ Vgl.: Wissenschaftsrat (2010), S. 54-57.

|³⁴ Vgl.: European Media Laboratory GmbH (2006), darin insbesondere S. 7-10.

Gauss Centre for Supercomputing (GCS): Nach Einschätzung des Wissenschaftsrates hat sich die Übertragung der Koordinierungsaufgaben in die Hände der HPC-Community bewährt. Nach Abwarten der Gründungsphase im Sinne der Empfehlung aus dem Jahr 2007 lässt sich feststellen, dass das GCS ein Beispiel für das gute Funktionieren dieses Prinzips darstellt.

Gauß-Allianz (GA): Im Fall der GA zeigen sich hingegen noch Defizite im Bereich der Koordinierung; bspw. findet keine Abstimmung im Vorfeld von Beschaffungen statt. Dies liegt aus Sicht des Wissenschaftsrates zum einen darin begründet, dass zumindest die hochschulischen Mitglieder der GA untereinander im Wettbewerb um Gelder aus dem Programm Forschungsbauten stehen. Zum anderen verfügt die GA nicht über ein nennenswertes eigenes Budget bzw. entsprechende Einkünfte aus Mitgliedsbeiträgen. Der Wissenschaftsrat empfiehlt den beteiligten Akteuren nachdrücklich zu prüfen, auf welche Weise die notwendigen finanziellen Voraussetzungen für die Selbstorganisation der GA geschaffen werden können und damit auch eine engere Koordination innerhalb der GA ermöglicht werden kann.

Auch das Fehlen einer arbeitsteiligen Governance beeinträchtigt die GA in der vermehrten Übernahme koordinierender Aufgaben innerhalb der Tier 2-Ebene. Die GA verfügt als Gremien lediglich über Vorstand und Mitgliederversammlung. Der Wissenschaftsrat ist der Überzeugung, dass neben der finanziellen auch eine strukturelle Stärkung der GA erforderlich ist. Zusätzlich zur Schaffung neuer Gremien wie bspw. Aufsichtsrat und Wissenschaftlicher Beirat sollte insbesondere die Möglichkeit einer hauptamtlichen Geschäftsführung bzw. Geschäftsstelle geprüft werden. Eine differenziertere Governance-Struktur würde auch dazu beitragen, die Transparenz von Entscheidungsprozessen, bspw. die Kriterien und Verfahren der Mitgliederaufnahme betreffend, zu erhöhen. Bei der Aufnahme neuer Mitglieder sollte generell deren Leistungsfähigkeit Priorität gegenüber formalen Kriterien eingeräumt werden.

Angesichts des guten Funktionierens des GCS und behebbarer Einschränkungen innerhalb der GA spricht sich der Wissenschaftsrat dafür aus, GCS und GA auch dauerhaft als Akteure der Koordination im Bereich HPC zu etablieren. Er wird daher die Tätigkeit des Nationalen Koordinierungsausschusses für die Beschaffung und Nutzung von Höchstleistungsrechnern, die ohnehin bereits seit dem Bericht aus dem Jahr 2007 ruht, nicht wieder aufnehmen und diesen Ausschuss auflösen.

Die strategische Koordination im Bereich HPC sollte immer im Lichte einer Stärkung der internationalen Wettbewerbsfähigkeit Deutschlands erfolgen.

PRACE: Auf europäischer Ebene bietet PRACE das Forum für eine enge Abstimmung der nationalen Höchstleistungsrechenzentren in den Mitgliedsstaaten. Dies erfolgt durch die zentrale Vergabe der von den nationalen Zentren be-

reitgestellten Rechenzeit nach gemeinsam festgesetzten Kriterien sowie durch die Identifikation von Forschungsdesideraten und deren Bearbeitung in sog. „Work Packages“. Eine Koordination von Beschaffungen zwischen den Mitgliedsländern findet hingegen bislang nicht statt, was wesentlich darin begründet liegt, dass diese rein national finanziert werden. Aus Sicht des Wissenschaftsrates wäre eine engere Koordination der in PRACE engagierten Mitgliedsstaaten unabhängig von der Frage der Finanzierung auch bei den Rechnerbeschaffungen wünschenswert. So ließe sich auch auf europäischer Ebene eine Struktur bestehend aus mehreren HPC-Zentren mit sich ergänzenden Rechnerarchitekturen und Methodenkompetenzen aufbauen, von der die europäischen Nutzerinnen und Nutzer optimal profitieren würden. Die betreffenden Mitgliedsstaaten sollten die zentrale Rolle in diesen Abstimmungsprozessen spielen.

II.4 Betrieb

Vergabe von Rechenzeit: Zentrale Aspekte im Betrieb von HPC-Infrastrukturen sind die Gestaltung der Rechenzeitvergabe und hiermit verbunden die Frage, welche Nutzergruppen Zugang zum Rechner erhalten. In diesem Punkt lassen sich wiederum Unterschiede zwischen den Höchstleistungsrechnern innerhalb des GCS und den darunter liegenden Leistungsklassen von Rechnern feststellen, die zu einem wesentlichen Teil mit den bestehenden Unterschieden in den Finanzierungsmodalitäten zusammenhängen. Während Tier 3- und Tier 2-Rechner i. d. R. nur für eingeschränkte Nutzergruppen wie bspw. Forschende der jeweiligen Sitzeinrichtung oder „Landeskinder“ zugänglich sind, steht die Nutzung der Tier 1-Rechner prinzipiell allen Nutzern bundesweit frei. Darüber hinaus gibt es dedizierte Rechenzeitkontingente, die im Rahmen von PRACE europäischen Anwendern zur Verfügung gestellt werden. Diese müssen sich wie auch die nationalen Nutzer der GCS-Systeme mit Anträgen um Rechenzeit bewerben, die durch wissenschaftliche Lenkungsausschüsse nach Kriterien wissenschaftlicher Qualität bewertet werden. Die überwiegende Mehrzahl, aber noch nicht alle Mitglieder der GA verfügen über ähnliche Verfahren und Gremien für die Vergabe von Rechenzeit. |³⁵ Aus Sicht des Wissenschaftsrates ist eine qualitätsgesteuerte Vergabe von Rechenzeit auf den obersten Leistungsebenen der Tiers 0, 1 und 2 essentiell. Auf den Ebenen Tier 0 und Tier 1 sollte wegen ihrer europäischen und nationalen Bedeutung zusätzlich die Berücksichtigung gesellschaftlicher Bedarfe in die Bewertung einfließen. Für die Vergabe von Rechenzeit haben sich entsprechend besetzte Wissenschaftliche Beiräte

| ³⁵ Ergebnis einer von der Geschäftsstelle des Wissenschaftsrates durchgeführten Umfrage unter den GA-Mitgliedern vom Februar 2011.

bewährt. Die HPC-Zentren, die noch nicht über entsprechende Gremien verfügen, sollten diese zeitnah einrichten. Wissenschaftliche Beiräte sollten nach Möglichkeit mehrheitlich mit externen Vertreterinnen und Vertretern der jeweils relevanten Anwendungswissenschaften besetzt werden.

Aus Sicht der Anwender sind zügige und transparente Vergabeverfahren zentral. Es sollte daher seitens der Betreiber von Rechenzentren Wert auf schlanke Prozesse und regelmäßige Möglichkeiten der Antragstellung gelegt werden. Hinsichtlich der Transparenz der Prozesse sieht der Wissenschaftsrat an einigen Rechenzentren noch Nachholbedarf. Die Kriterien und Verfahrensschritte der Vergabe von Rechenzeit sollten allen potentiellen Anwendern zugänglich sein. Zudem sollte öffentlich gemacht werden, mit welchen Personen die Gremien besetzt sind, die über die Vergabe von Rechenzeiten entscheiden. Damit wird das Vertrauen der Wissenschaft in diese Entscheidungen nachhaltig gestärkt.

Zugang: Bereits in seinen Empfehlungen aus dem Jahr 2000 hat der Wissenschaftsrat für HPC-Rechensysteme einen offenen „Zugang unabhängig von den geographischen Voraussetzungen und den institutionellen Zugehörigkeiten der Nutzer“ |³⁶ gefordert. Diese Forderung gilt umso mehr, wenn angestrebt wird, dass die Zentren der obersten Leistungsklassen über differenzierte Architekturen verfügen, die auf unterschiedliche Anwendungsprofile zugeschnitten sind. Dies trifft nicht nur für die Höchstleistungsrechner des GCS zu. In der Antragstellung auf Fördermittel aus dem Forschungsbautenprogramm sind auch die mehrheitlich in der GA vertretenen Betreiber von Tier 2-Systemen aufgefordert darzulegen, in welcher Weise die jeweils gewählte Rechnerarchitektur spezifische Anforderungen relevanter Anwendungsgruppen erfüllt. Angesichts der hieraus resultierenden Unterschiede in den Rechnerarchitekturen, die nach Ansicht des Wissenschaftsrates mit dem Ziel einer nationalen Struktur von sich ergänzenden HPC-Kompetenzzentren (vgl. I.1) gezielt gefördert werden sollte (vgl. II.2.a), hält der Wissenschaftsrat es daher nicht für zielführend, wenn der Zugang zu Tier 2-Rechnern auf regional oder institutionell festgelegte Nutzergruppen beschränkt ist. Er empfiehlt, auch auf dieser Leistungsstufe einen bundesweit offenen Zugang auf der Basis wissenschaftlicher Qualität zu ermöglichen.

PRACE: Auf europäischer Ebene ist im Rahmen von PRACE ein wissenschaftsgeleitetes Verfahren zur Vergabe von Rechenzeit ähnlich dem des GCS etabliert. Regelmäßig erfolgende *Calls* bieten europäischen Nutzern die Gelegenheit, sich um die bereitgestellten Kontingente von Rechenzeit zu bewerben. Dabei sehen die Regularien von PRACE die Einhaltung des „juste-retour-Prinzips“ vor, wel-

| ³⁶ Wissenschaftsrat (2001), S. 231.

ches besagt, dass die teilnehmenden Nationen von den Ressourcen relativ zu den von ihnen eingebrachten Mitteln profitieren sollen. Ein entsprechender Abgleich soll zum Ende der Projektlaufzeit erfolgen. Der Wissenschaftsrat begrüßt das Vorgehen in PRACE, über alternative Leistungserbringungen (z. B. Software-Entwicklung) auch Ländern, deren finanzielle Möglichkeiten die Bereitstellung von Rechenzeit auf diesem Leistungsniveau nicht erlauben, den Zugang zu den PRACE-Systemen zu ermöglichen.

Anhang

BIP	Bruttoinlandsprodukt
BMBF	Bundesministerium für Bildung und Forschung
BremHLR	Kompetenzzentrum für Höchstleistungsrechnen Bremen
CSE	Computational Science and Engineering
DFG	Deutsche Forschungsgemeinschaft
EFR	Europäischer Forschungsraum
EGI	European Grid Infrastructure
ESFRI	European Strategy Forum on Research Infrastructures
EU	Europäische Union
Flops	floating-point operations per second
GA	Gauß-Allianz
GCS	Gauss Centre for Supercomputing
GERD	Gross Domestic Expenditure on Research and Development
GWK	Gemeinsame Wissenschaftskonferenz
HLRS	Höchstleistungsrechenzentrum Stuttgart
HPC	High Performance Computing
JARA-HPC	Jülich Aachen Research Alliance – Section High Performance Computing
JSC	Jülich Supercomputing Centre
KONWIHR	Kompetenznetzwerk für Hoch- und Höchstleistungsrechnen in Bayern
LRZ	Leibniz-Rechenzentrum Garching
PSL	Programmativ-strukturelle Linie „Hochleistungsrechner“
PRACE	Partnership for Advanced Computing in Europe
WR	Wissenschaftsrat

Bundesministerium für Bildung und Forschung: Die Säulen des Supercomputing. Höchstleistungsrechnen gibt Antworten auf schwierigste Fragen unserer Zeit, Bonn; Berlin 2011.

Bundesregierung: Stellungnahme der Bundesregierung zum Grünbuch „Von Herausforderungen zu Chancen: Entwicklung einer gemeinsamen Strategie für die EU-Finanzierung von Forschung und Innovation“ (Stand 19.04.2011), http://www.bmbf.de/pubRD/stellungnahme_BREG_gruenbuch.pdf [zuletzt abgerufen am 07.11.2011] [= Bundesregierung (2011a)].

Bundesregierung: 2. Leitlinienpapier der Bundesregierung für das kommende Rahmenprogramm für Forschung und Innovation – Thematische Positionen (Stand 17.06.2011), http://www.bmbf.de/pubRD/zweites_leitlinienpapier__frp_deutsch.pdf [zuletzt abgerufen am 08.11.2011] [= Bundesregierung (2011b)].

Deutsche Forschungsgemeinschaft: Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme. Empfehlungen der Kommission für IT-Infrastruktur für 2011-2015, Bonn 2010.

European Media Laboratory GmbH: High Performance Computing in Deutschland. Argumente zur Gründung einer strategischen Allianz, Heidelberg 2006.

Gauss Centre for Supercomputing: Supercomputing at the Leading Edge, Berlin 2010.

Gauss Centre for Supercomputing: Applications, in: inside. Innovatives Supercomputing in Deutschland, 9 (2011) 1, S. 14-71.

Konrad-Zuse-Zentrum für Informationstechnik Berlin; Regionales Rechenzentrum für Niedersachsen, Hannover: Forschung im HLRN-Verbund 2011, Berlin 2011.

TOP500.Org: TOP500 List – June 2011, <http://www.top500.org/lists/2011/06/> [zuletzt abgerufen am 08.11.2011].

Wissenschaftsrat: Empfehlung zur Versorgung von Wissenschaft und Forschung mit Höchstleistungsrechenkapazität, in: Wissenschaftsrat: Empfehlungen zur Ausstattung der Wissenschaft mit moderner Rechner- und Kommunikationstechnologie, Köln 1995, S. 51-70.

Wissenschaftsrat: Empfehlung zur künftigen Nutzung von Höchstleistungsrechnern, in: Wissenschaftsrat: Empfehlungen und Stellungnahmen 2000, Bd. I, Köln 2001, S. 229-261.

- 39** Wissenschaftsrat: Stellungnahme zur Anmeldung des Landes Baden-Württemberg auf Errichtung eines Höchstleistungsrechners in Karlsruhe und Stuttgart zum 32. Rahmenplan, in: Wissenschaftsrat: Empfehlungen und Stellungnahmen 2002, Bd. II, Köln 2003, S. 61-82.
- Wissenschaftsrat: Empfehlung zur Einrichtung europäischer Höchstleistungsrechner, in: Wissenschaftsrat: Empfehlungen und Stellungnahmen 2004, Bd. III, Köln 2005, S. 505-538.
- Wissenschaftsrat: Stellungnahme zur Beschaffung eines Nachfolgesystems des Rechnerverbundes der Norddeutschen Länder für Hoch- und Höchstleistungsrechnen, in: Wissenschaftsrat: Empfehlungen und Stellungnahmen 2005, Bd. I, Köln 2006, S. 421-436.
- Wissenschaftsrat: Bericht über die Tätigkeiten des Nationalen Koordinierungsausschusses zur Beschaffung und Nutzung von Höchstleistungsrechnern, in: Wissenschaftsrat: Empfehlungen und Stellungnahmen 2007, Bd. III, Köln 2008, S. 277-295.
- Wissenschaftsrat: Empfehlungen zur zukünftigen Entwicklung der deutschen marinen Forschungsflotte (Drs. 10330-10), Lübeck November 2010.

Das Publikationsformat „Positionspapier“ wurde 2010 eingeführt, um mit kurzen, zugespitzt formulierten Papieren in absehbarer Zeit auf aktuelle Themen und Entwicklungen reagieren zu können. Im Positionspapier wird deshalb auch - anders als in den übrigen Publikationsformaten des Wissenschaftsrats - darauf verzichtet, umfangreiche empirische Informationen zeitaufwändig aufzuarbeiten und in den Text zu integrieren. Generell zeichnet sich das Format durch eine große prozedurale, thematische und formale Flexibilität aus.

© Wissenschaftsrat
Drs. 1838-12
Verabschiedet in Berlin, Januar 2012