Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

HANS FALK HOFFMANN, ALOIS PUTZER,
ALEXANDER REINEFELD

Vom World Wide Web zum World Wide Grid

Eine neue Informationsinfrastruktur für wissenschaftliche Anwendungen

ZIB-Report 02-39 (November 2002)

Vom World Wide Web zum World Wide Grid

Eine neue Informationsinfrastruktur für wissenschaftliche Anwendungen*

Hans Falk Hoffmann[1], Alois Putzer[2], Alexander Reinefeld[3]

[1] European Organization for Nuclear Research (CERN)
[2] Kirchhoff-Institut für Physik (KIP)
[3] Zuse-Institut Berlin (ZIB)

Zusammenfassung

In Analogie zu den Elektrizitätsnetzen (*electricity grid*), von denen die technische Revolution ausging, wurde der Begriff *Computational Grid* (kurz Grid) geprägt. Ein wichtiger Bestandteil des Systems liegt im benutzerfreundlichen Zugang und der koordinierten Nutzung der weltweit verteilten Speicherressourcen und Rechnerkapazitäten. Bei der Entwicklung dazu notwendiger Technologien und Software (Middleware) profitiert man von Kenntnissen und Erfahrungen bei der Entwicklung verteilter Algorithmen, dem Software-Engineering und dem Supercomputing.

1 Einleitung

Die Qualität von Forschung und Wissenschaft und natürlich auch die Wettbewerbsfähigkeit vieler Industrien hängen in immer stärkerem Maße von einer globalen, potenten und funktionierenden Kommunikations-Infrastruktur ab. Eine solche Infrastruktur muss nicht nur ständig wachsende Bandbreiten der Netze bereitstellen, sondern auch neue und innovative Dienste. „*Eventually, users will be unaware they are using any computer but the one on their desk, because it will have the capabilities to reach out across the national network and obtain whatever computational resources are necessary*“. Mit dieser Vision läuteten Larry Smarr und Charles Catlett [2] im Jahr 1992 eine neue Epoche des Hochleistungsrechnens ein, die nun sukzessiv in die Praxis umgesetzt wird.

Im zukünftigen *World Wide Grid* sollen Berechnungen an den günstigsten Stellen im Netz durchgeführt werden, wobei die so genannte Middleware den Datenzugang und die Zuweisung von Rechenkapazitäten organisiert. Für den einzelnen Wissenschaftler scheint es, als würden sich alle Daten an einem Ort befinden und von einem

*Erschienen in: Physikalische Blätter, 57 (Dez. 2001) Nr. 12, WILEY-VCH Verlag GmbH, Weinheim, pp. 39-44.

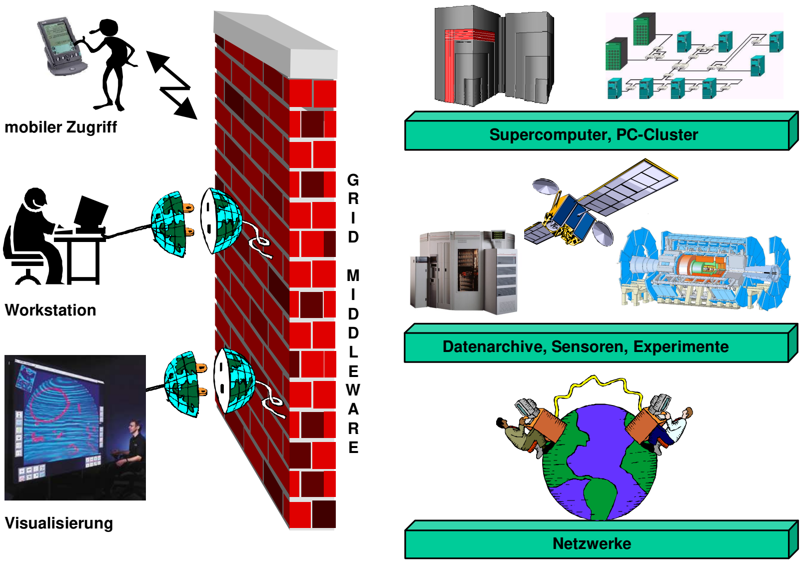

Abbildung 1: *Das Grid erlaubt es, verteilte Ressourcen zu nutzen: Die Grid-Middleware (Mauer) bietet dem Nutzer (links) den Zugriff auf die notwendigen Systemressourcen (rechts), ohne dass dieser sich um die technischen Details der koordinierten Nutzung selbst kümmern muss.*

(virtuellen) Supercomputer bearbeitet werden. Je nach Komplexität des Problems werden die Resultate nach Sekunden, Minuten oder auch erst nach Stunden dem Anwender übermittelt.

Auch das *World Wide Web* begann zuerst mit dem Austausch wissenschaftlicher Daten zwischen Computerzentren. Die Entwicklung offener Standards und frei zugänglicher, nutzerfreundlicher Software trug wesentlich zu dem heutigen Erfolg des Web mit seinen Millionen Nutzern weltweit bei. Die Wissenschaftler sehen diese Entwicklung mit einem lachenden und einem weinenden Auge. Auf der einen Seite steht der Stolz auf das Erreichte, auf der anderen erschwert die Vielzahl der Nutzer die effiziente wissenschaftliche Nutzung der Datennetze und den Austausch wissenschaftlicher Daten. Das Grid soll unter anderem dazu beitragen, die Datennetze und die verbundenen Ressourcen intelligenter zu nutzen (Abb. 1).

Die Teilchenphysik steht durch die im Bau befindlichen Experimente vor dem realen Problem, dass Wissenschaftler auf der ganzen Welt auf die ab 2006 anfallenden Daten (etwa 10^{16} Bytes pro Jahr) zugreifen und diese mit Hilfe von weltweit verteilten Clustern aus tausenden von vernetzten Kleincomputern analysieren möchten. Daher hat sie die Initiative ergriffen und in Grid-Projekten in Europa und den USA Computerspezialisten mit Forschern verschiedener Bereiche zusammengebracht, um gemeinsam die Basis für ein internationales Grid zu schaffen.

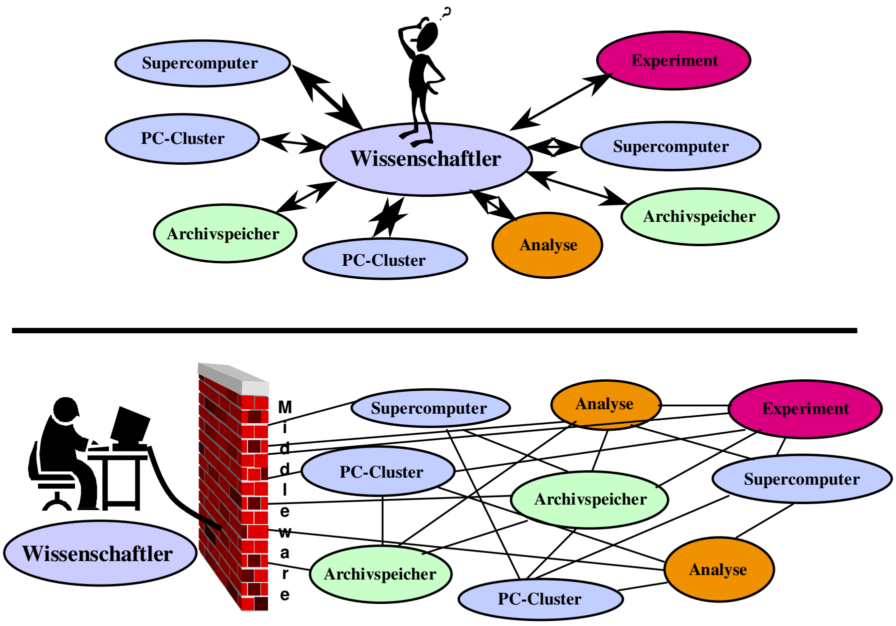

Abbildung 2: *Vorteile des Grid-Computing: In herkömmlichen verteilten Systemen (oben) muss sich der Wissenschaftler selbst um die Bereitstellung aller benötigten Computer-Ressourcen kümmern; in einer Grid-Umgebung (unten) wird diese Aufgabe von der Middleware übernommen.*

2 Was ist Grid-Computing?

Für das Grid Computing sind im Laufe der Zeit verschiedene Begriffe geprägt worden, die aber letztlich alle die koordinierte Nutzung verteilter Computer-Ressourcen beschreiben: *Metacomputing, Hypercomputing, Seamless Computing, virtuelles Supercomputing, vernetztes Höchstleistungsrechnen*, und so weiter. Sie unterscheiden sich lediglich in Nuancen: Während mit dem *Seamless Computing* der benutzerfreundliche, transparente Zugriff auf geographisch verteilte Hochleistungsrechner angepeilt wird, konzentriert man sich im *virtuellen Supercomputing* auf die Kopplung von Rechnern zur Erzielung höchster Rechenleistung für einzelne, sehr aufwendige Berechnungen.

Dem Grid-Computing liegt der Gedanke zugrunde, dass über ein weltumspannenden Netzwerks überall und jederzeit Rechenleistung gegen ein Entgelt zur Verfügung stehen soll - ähnlich wie bei einem Stromnetz, bei dem sich elektrische Leistung einfach über eine standardisierte Steckdose abgreifen lässt. Allerdings hinkt dieser Vergleich etwas, da die verschiedenen Computer im Grid häufig miteinander inkompatibel sind. Während es beim elektrischen Strom nur wenige Grundeigenschaften gibt (Spannung, Stromstärke, Frequenz), sind beim Handel und Austausch von Rechenleistung wesentlich mehr Parameter zu berücksichtigen: Die Architektur des Zielrechners (Vektor, Skalar, massiv parallel), dessen Betriebs- und Managementsystem, die Anbindung an das Netzwerk und vieles mehr. Gerade in der technischen oder auch verwaltungstechnischen Diversifikation liegen viele grundsätzliche Probleme, die noch nicht alle vollständig untersucht und gelöst worden sind.

Mit dem Grid-Computing sollen sich Ressourcen koordiniert nutzen lassen, um Aufgaben zu lösen, die sich in dynamischen, verteilten Organisationen stellen - etwa Forscher in Laboren oder Mitarbeiter in multinationalen Unternehmen, die an verschiedenen Standorten gemeinsam an einer Aufgabe arbeiten und dabei verteilte Ressourcen nutzen. Die Nutzung bezieht sich auf mehr als nur Rechenleistung und Daten, sie umfasst auch Netzwerke, Datenarchivspeicher, Visualisierungssysteme und teure Spezialrechner, die nicht an jedem Ort betrieben werden können (Abb. 2).

2.1 Arten des Grid-Computing

Wenn das Grid zur schnellen, gleichzeitigen Abarbeitung sehr vieler Programme konzipiert ist, spricht man vom *Duchsatzrechnen (high-throughput computing)*. Typische Anwendungsfelder liegen in der Bioinformatik (Sequenz-Alignment, Molekülsimulation), in der Hochenergie-Physik (Monte Carlo Simulationen, Auswertung von Beschleunigerexperimenten) und in der Satellitenbildauswertung. Derartige Anwendungen erfordern keine kostenaufwendigen Hochleistungsrechner, sondern können häufig auch auf PC-Clustern oder Computer-Farmen ausgeführt werden. Statt vieler Rechenoperationen pro Sekunde wird hier ein hoher Durchsatz parametrisierter Einzeljobs benötigt.

Eine neue Variante des Durchsatzrechnens, das *Internet Computing*, hat in letzter Zeit besonders viel Aufsehen erregt, da es ungenutzte Rechenkapazitäten für globale Rechnungen einsetzt. Im populären Seti-Projekt (www.seti.org) haben viele tausend PC-Besitzer einen kostenlosen Bildschirmschoner vom Web heruntergeladen, der im Hintergrund, während der PC gerade nicht benutzt wird, Teleskopdaten auf Merkmale außerirdischer Signale untersucht. Allein durch das Engagement der PC-Besitzer standen über einen längeren Zeitraum hinweg kostenlos mehrere Teraflop/s Rechenleistung für das Experiment zur Verfügung. Diese Art des verteilten Rechnens bleibt jedoch denjenigen Anwendungen vorbehalten, deren Problemstellung sich gut in viele Millionen kleine Pakete aufteilen lässt, die nur wenig Ein/Ausgabedaten benötigen und kaum kommunizieren. Diese Sonderklasse von Anwendungen gehört nicht zum Grid-Computing im engeren Sinne.

Auch das *Peer-to-Peer-Computing* nutzt die Rechen- und Speicherkapazität vieler Millionen PCs, die über das Internet miteinander verbunden sind. Der besondere Aspekt des Peer-to-Peer-Computing liegt im Fehlen einer zentralen Managementinstanz: Jeder PC nimmt als gleichberechtigter Partner (= peer) am Grid teil. Bislang stand die Nutzung verteilter Speicherkapazität (bei Napster oder Gnutella z.B. der Austausch von MP3-Musiktiteln) im Vordergrund, die kooperative Nutzung von Rechenkapazität ist aber auch denkbar. Erste Ansätze hierzu sind vielversprechend.

Eine im wissenschaftlichen Bereich besonders wichtige Variante des Grid-Computing ist das *verteilte Höchstleistungsrechnen (high-performance computing)*, in dem die leistungsfähigsten Supercomputer mit dem Ziel verknüpft werden, eine größtmögliche Anwendungsrechenleistung zur Lösung einzelner Probleme zur Verfügung zu stellen. Ein Beispiel hierfür ist die *Cactus*-Umgebung (www.cactuscode.org), eine Software, mit der die 3D-Einstein-Gleichungen der Allgemeinen Relativitätstheorie auf verteilten Systemen unter anderem in Berlin, Chicago und San Diego gelöst werden. Andere typische Anwendungsszenarien finden sich in *multi-physics*-Simulationen, in denen z.B. Strukturmechanik- mit Strömungssimulationscodes zur Simulation der Eigenschaften von Karosserien im Automobilbau gekoppelt werden. Auch die interaktive Visualisierung komplexer Simulationsrechnungen, die nicht vor Ort durchgeführt werden können, gehört in diese Kategorie.

Neben der höheren Rechenleistung liegt eine Hauptmotivation zum Grid-Computing in der koordinierten Nutzung und Speicherung großer, *verteilter Datenmengen*. Der enorme Leistungszuwachs moderner Prozessoren ermöglicht immer feiner aufgelöste Simulationsrechnungen und Visualisierungen, die einen exponentiell wachsenden Datenbedarf haben. Das unten beschriebene *DataGrid*-Projekt leistet Pionierarbeit auf dem Gebiet des Managements und der Verarbeitung verteilter Daten im Grid.

2.2 Grid-Architekturen

Eines der schwierigsten Probleme im Entwurf von Computational Grids liegt in der Sicherstellung der *Interoperabilität* der verteilten Systemsoftware und Anwendungsprogramme. In herkömmlichen, verteilten Systemen wirft die Interoperabilität keine größeren Probleme auf, da der Systementwickler ein einheitliches Kommunikationsschema verbindlich vorschreiben und umsetzen kann. So wird z.B. in Client/Server-Architekturen in der Regel zunächst die Schnittstelle des Servers festgelegt, wodurch die möglichen Interaktionen der Klienten definiert sind. Im Grid-Computing ist dies nicht möglich, da hier viele autonome Kommunikationspartner in nahezu beliebigen Konstellationen miteinander interagieren. Zudem sind Grid-Systeme hochgradig dynamisch: Im Laufe der Zeit werden Ressourcen, Dienste und Technologien neu in das Grid integriert oder herausgenommen.

Ein zentraler Aspekt beim Entwurf einer Grid-Architektur liegt daher in der Definition geeigneter *Protokolle* (siehe Glossar), die die Interaktionsmöglichkeiten der beteiligten Komponenten definieren. Dabei ist zu beachten, dass Protokolle lediglich die Interaktionen der Kommunikationspartner regeln; die Umsetzung der Protokolle in einen Programmcode bleibt dem Programmierer überlassen. Dies erlaubt eine flexible Erstellung und Erweiterung der Grid-Umgebung durch zukünftige Dienste. Protokolle legen Mechanismen zum Aufruf von *Diensten* in Grid-Umgebungen fest. Sie sind, genau wie die Dienste selbst, implemtationsunabhängig. Dienste werden durch Softwaremodule (z.B. UNIX-„Daemone“) erbracht, die in einer beliebigen Programmiersprache implementiert sein können. Dem konkreten Aufruf von Diensten oder einzelner Funktionen eines Dienstes dienen *Programmschnittstellen* (APIs).

2.3 Elemente einer Grid-Architektur

Wie in modernen Kommunikationsarchitekturen üblich, sind auch Grid-Systeme in stufenweise aufeinander abgestimmte Ebenen gegliedert. Dabei stellt jede Ebene einen Satz von Diensten bereit, der von der darüber liegenden Ebene genutzt werden kann.

Ebene 1: Hardware/Software-Ressourcen. Auf der untersten Ebene der Grid-Architektur befinden sich die physikalischen und logischen Ressourcen, die letztlich die Dienste im Grid bereitstellen: Rechenknoten, Speichersysteme, Netzwerke, Spezialhardware, u.s.w.. Die Inhomogenität und der dynamische Charakter der Ressourcen werfen erhebliche Probleme für das Management auf. So muss das Grid-Management-

Glossar

Ein **Protokoll** definiert die zwischen den Kommunikationsendpunkten einer Datenkommunikation gültigen Regeln zum Austausch von Nachrichten. Beispiele von Protokollen sind:

- *das Internet Protocol* (IP), das ein verbindungsloses, nicht zuverlässiges Übertragungssystem inklusive Paketfragmentierung und Zielsuche der Nachrichten definiert;
- *das Transmission Control Protocol* (TCP), das auf IP aufsetzend einen verbindungsorientierten, zuverlässigen Byte-orientierten Datenstrom zur Verfügung stellt;
- *das File Transfer Protocol* (FTP), das Regeln zum Austausch von Dateien spezifiziert.

Über die Implementierung der Dienste sagen Protokolle nichts aus, sie definieren lediglich die Regeln der Kommunikation. Dadurch können Kommunikationspartner miteinander kommunizieren, ohne dass sie notwendigerweise dieselbe Programmiersprache oder kompatible Systeme benutzen.

Ein **Dienst** (Service) stellt eine bestimmte Leistung zur Verfügung. Dienste werden über Protokolle definiert, über die sie abgerufen werden können. Auch Dienste sind implementationsunabhängig; spezifiziert werden lediglich die Leistungen. Typische Dienste sind:

- der *FTP-Server*, der das Erstellen, Löschen, Lesen und Schreiben von Dateien auf entfernten Ressourcen (z.B. auf einem Platten- oder Bandlaufwerk) ermöglicht; die Anweisungen erfolgen gemäß dem FTP-Protokoll.
- der *LDAP-Server*, der Anfragen über die aktuell lokal verfügbaren Ressourcen beantwortet; er bedient sich dazu des LDAP-Protokolls (*lightweight directory access protocol*),

Dienste werden durch sog. Daemon-Prozesse, wie z. B. dem bekannten UNIX Mailer-Daemon, erbracht. Die Verfügbarkeit der Dienste ist nicht immer garantiert, da ggf. Zugangsbeschränkungen bestehen oder der betreffende Daemon erst gestartet werden muss.

Eine **Programmschnittstelle** (API, Application Program Interface) definiert eine Schnittstelle zum Aufruf von Diensten, d.h. genauer der vom Dienst zur Verfügung gestellten Funktionen oder Objekte mit den darauf arbeitenden Methoden. Ein typisches Beispiel einer Programmschnittstelle ist

- das *Message Passing Interface (MPI)*, das Funktionen zum Nachrichtenaustausch zwischen Prozessen definiert; es ist für C und Fortran definiert.

Üblicherweise existieren Programmschnittstellen für alle Programmiersprachen oder Skripte, von denen aus der Aufruf des zugrunde liegenden Dienstes sinnvoll ist. Es kann also mehrere Programmschnittstellen für einen Dienst geben.

system flexibel auf lokale Betriebseingriffe, wechselnde Leistungsfähigkeit oder gar auf Systemausfälle reagieren, indem z.B. der Programmablauf kurzfristig neu geplant wird.

Ebene 2: Verbindungsebene. Die Verbindungsebene regelt den sicheren Zugriff auf die Ressourcen über die Netzverbindung hinweg. Neben den üblichen Netzprotokollen (TCP/IP, UDP) sind hier auch Methoden zum sicheren Benutzerzugriff und zur Weiterleitung von Zugriffsrechten angesiedelt.

Ebene 3: Ressourcenzugriff. Auf der Ressourcenebene wird der Zugriff auf einzelne Ressourcen initiiert und kontrolliert. Der aktuelle Zustand einzelner Ressourcen kann abgefragt werden, Anforderungen können gestellt und mit dem lokalen Managementsystem ausgehandelt werden und der Programmablauf kann kontrolliert/gesteuert werden.

Ebene 4: Kooperatives Ressourcen-Management. Aufbauend auf der Ressourcenzugriffsebene wird hier die gemeinsame Reservierung zusammengehöriger Ressourcen für Grid-Anwendungen gesteuert. Dies umfasst die zeitliche und örtliche Planung und Durchführung komplexer Systemanforderungen, die mehrere Komponenten benötigen. Die dafür grundlegenden Fragen des sogenannten Co-Scheduling sind prinzipiell geklärt, allerdings existieren erst wenige Implementierungen, da viele lokale Ressourcen-Management-Systeme die Reservierung von Komponenten noch nicht unterstützen. Auch die Vermittlung zwischen Angebot und Nachfrage liegt auf dieser Ebene.

Ebene 5: Anwendungsebene. Auf dieser Ebene wird schließlich die Anwendung ausgeführt. Dazu gehört u.a. das Datenmanagement, die Applikationssteuerung, die Prozesserzeugung und die Mechanismen zur verteilten Kommunikation. Für einige typische Anwendungsfelder (z.B. entfernte Visualisierung) sind spezielle Grid-Umgebungen erstellt worden.

3 Die Herausforderung der LHC-Experimente

Seit Jahrzehnten werden in der Teilchenphysik Experimente durchgeführt, in denen hunderte von Physikern enorme Datenmengen analysieren. Zu Beginn des neuen Jahrhunderts machte man erneut einen Schritt in neue Größenordnungen: In der Zahl der beteiligten Wissenschaftler, in der zu analysierenden Datenmenge und in der Dauer der Experimente. Die Teilchenphysik steht hier aber nur als Beispiel. Ähnliche Anforderungen treten in anderen Wissenschaftsbereichen und nicht zuletzt auch in der Industrie auf.

In der Teilchenphysik werden Bündel von Teilchen (Elektronen, Protonen, Ionen) beschleunigt und entweder auf feste Materie („Targets“) oder aufeinander geschossen. Dabei wird die Wechselwirkungsenergie in neue Teilchen umgewandelt, die in riesigen Detektoren registriert werden. Unzählige solcher Ereignisse müssen analysiert werden, um einen Schritt weiter auf dem Weg zu neuen Erkenntnissen zu gehen.

Ereignisse von fundamentalem wissenschaftlichen Interesse wie z.B. die Erzeugung von Higgs-Teilchen mit dem anschließenden Zerfall in beobachtbare Teilchen werden höchst selten, nämlich nur etwa einmal bei 10^{12} Proton-Proton-Wechselwirkungen erwartet. Der Nachweis für diese Teilchen, von dem man sich ein tieferes Verständnis des Ursprungs der Masse verspricht, ist daher noch schwieriger als die sprichwörtliche Suche nach einer Stecknadel in einem Heuhaufen; eine Suche in 20 Millionen Heuhaufen würde der Realität schon näher kommen. Um Antworten auf die physikalischen Fragen bekommen zu können, müssen viele Petabytes (10^{15} Bytes) an Daten aufgezeichnet werden, die eine enorme Zahl von Wechselwirkungen beschreiben. Dazu werden neben Beschleunigern auch Detektoren benötigt, die Tausende von Tonnen wiegen und mehrere hunderttausend Messkanäle haben. Wenn jede Person auf der Welt gleichzeitig 20 Telefongespräche führen würde, entspräche das in etwa dem Datenfluss in einem Detektor.

Im Jahre 2006 soll der *Large Hadron Collider (LHC)* am Europäischen Labor für Teilchenphysik (CERN) in Genf in Betrieb genommen werden und Proton-Proton-Wechselwirkungen bei 14.000 GeV Schwerpunktenergie erzeugen. Die Protonenbündel werden bis zu 40 Millionen mal pro Sekunde kollidieren, wobei sich jeweils etwa 25 Wechselwirkungen mit der Produktion von im Mittel jeweils 100 neuen Teilchen ereignen. Als Schwerionenbeschleuniger wird der LHC Wechselwirkungen bei 5700 GeV/Nukleon (im Schwerpunktsystem) erzeugen. In diesem Fall werden pro Ereignis viele Tausend geladene Teilchen im Endzustand registriert.

Zum Bau der vier LHC-Detektoren *ALICE, ATLAS, CMS und LHCb* haben sich internationale Kollaborationen mit insgesamt mehr als 5000 Physikern aus etwa 300 wissenschaftlichen Einrichtungen gebildet, die über mehr als 50 Länder weltweit verteilt sind. Jedes der vier LHC-Experimente zeichnet die Daten nach verschiedenen Triggerstufen mit einer Rate von etwa 150 Hz und einer Ereignisgröße von 1-2 MB auf. Nach der Online-Datenreduktion und Datenkomprimierung um einen Faktor von mehreren Millionen stehen etwa 10^9 Ereignisse und eine Gesamtdatenmenge von mehreren Petabytes pro Jahr für die Analyse zur Verfügung. Würde man die Daten eines Jahres auf CDs speichern und diese übereinander legen ergäbe sich ein viele Kilometer hoher Turm. Diese Rohdaten werden im CERN rekonstruiert und dann von Instituten auf der ganzen Welt analysiert. Um sicherzustellen, dass die notwendigen Ressourcen für die Datenanalyse vorhanden sind, benötigt jedes Experiment ein weltweit verteiltes Datenbank- und Rechnersystem mit

1. großen Datenspeichern (viele Petabytes),
2. einer großen Zahl von Rechnern (einige zehntausend Prozessoren) und
3. einem Zugangssystem für weltweit tausende von Nutzern, die gleichzeitig mit den Daten arbeiten können.

Diese Anforderungen sind (Abb. 3 und 4) seit einigen Jahren bekannt. Deshalb nahmen alle vier LHC Experimente an dem *MONARC*-Projekt (Models of Networked Analysis at Regional Centres for LHC Experiments, monarc.web.cern.ch/monarc) teil. Ein hierarchisches System mit verschiedenen Ebenen (*Tiers*) bildet die Grundlage des LHC-Computing (Abb. 5). In diesem Model findet die Speicherung der Experimentdaten und die Rekonstruktion am Tier-0-Zentrum am CERN statt. Die Analyse und die Durchführung von Ereignissimulationen werden vor allem Aufgabe der regionalen Tier-1-Zentren sein, gefolgt von nationalen Tier-2-Zentren, Institutsrechnern (Tier-3) und lokalen Arbeitsplatzrechnern (Tier-4). Die wesentlichen Voraussetzungen, die zu diesem Vorschlag führten, sind:

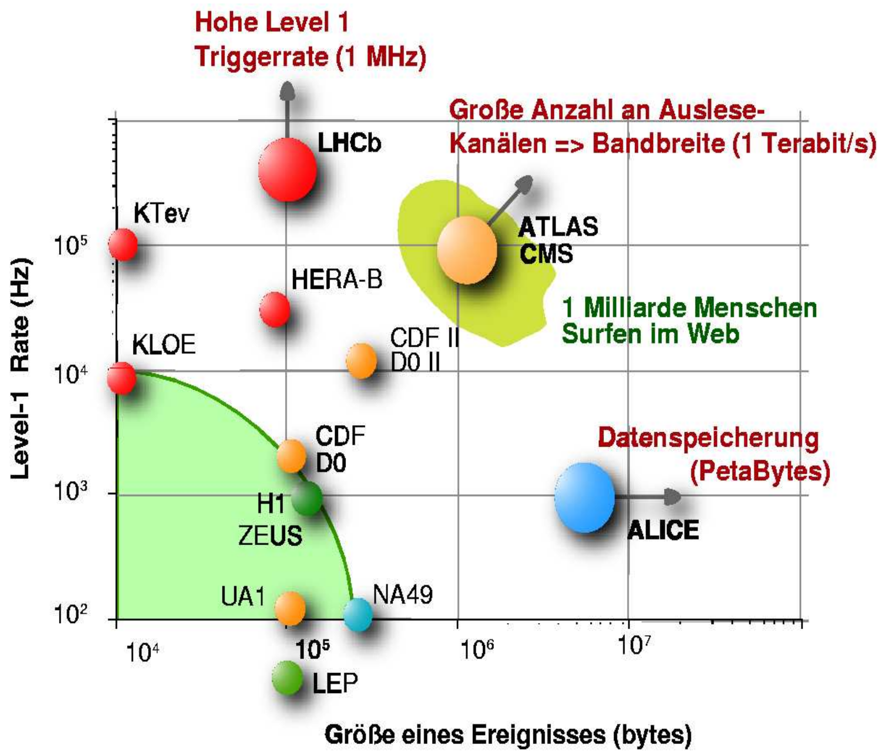

Abbildung 3: *Vergleich der Datenmenge und des Datenflusses der LHC-Experimente mit einigen laufenden und abgeschlossenen Experimenten der Teilchenphysik. Die eingezeichnete Wolke kennzeichnet grob die Anforderungen, wenn eine Milliarde Menschen gleichzeitig im Web surfen würden.*

- Die intellektuellen Beiträge der Physiker überall auf der Welt sollen eingebunden werden, ohne dass alle am selben Ort (in diesem Fall CERN) sein müssen.
- Alle Wissenschaftler müssen, zumindest im Prinzip, den gleichen Zugang zu den Daten haben, unabhängig von ihrem Standort.
- Der Zugang zu den Daten muss transparent und so effizient wie möglich sein. Dies muss auch für den Zugang zu den tausenden von Prozessoren gelten, die benötigt werden, die Daten zu selektieren und weiterzubearbeiten.

Die Idee des Grid-Computing kam gerade zur rechten Zeit, um die in MONARC aufgeworfenen Fragen zu beantworten. Es scheint alle Elemente zu enthalten, die notwendig sind, die Computing Probleme der zukünftigen Teilchenphysik-Experimente zu lösen. Andererseits ist die Teilchenphysik ein ideales Feld, um zu zeigen, dass die Grid-Technologie sowohl für daten- als auch rechenintensive Anwendungen nutzbar ist.

Obwohl die eigentliche Datenaufnahme erst im Jahr 2006 beginnt, werden schon jetzt Prototypen *(Testbeds)* der Regionalzentren benötigt, um die Testdaten auszuwerten, die Triggerbedingungen zu optimieren und die kritischen Parameter des weltwei-

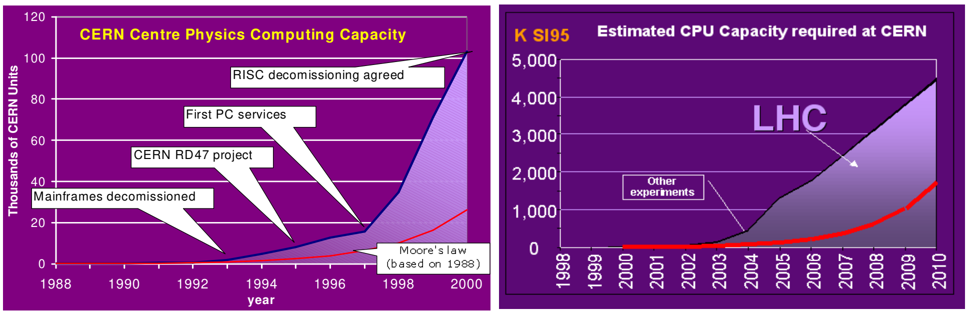

Abbildung 4: *Vorhandene CPU-Leistung des CERN-Computerzentrums in der Vergangenheit (links) und Extrapolation der zukünftig für die LHC-Experimente erforderlichen Leistung (rechts). Die Linie markiert die Vorhersagen des Mooreschen Gesetzes für die CPU-Leistung, die sich aufgrund von Technologie- und Preisentwicklungen bei konstantem Budget erwarten lässt. Zwar ist die Entwicklung der Rechnerkapazitäten bisher viel günstiger verlaufen als erwartet, die Datenexplosion bei den LHC-Experimenten wird diese in jedem Fall abhängen. Hier könnte das Grid Abhilfe schaffen, indem es die verteilten Computer-Ressourcen bündelt.*

ten Computing-Modells zu testen. An diesen Tests *(Data Challenges)* werden neben CERN die meisten Tier-1- und einige Tier-2-Zentren teilnehmen. Um die Kosten für die Hardware möglichst niedrig zu halten, werden die LHC-Experimente die vorhandenen Ressourcen gemeinsam nutzen.

An den LHC-Experimenten nehmen mehr als 300 Wissenschaftler aus 25 deutschen Instituten teil, die wesentlich zum Bau der vier LHC-Detektoren beitragen und vor allem danach an der Analyse der Daten mitarbeiten werden. Deshalb wird dafür ein deutsches Regionales Daten- und Rechenzentrum am Forschungszentrum Karlsruhe (FZK) unter wissenschaftlicher Beteiligung von Informatikern anderer Institute in Deutschland eingerichtet.

Während der Aufbau- und Testphase des „LHC Computing Grids“ wird dieses Regionalzentrum schon von den bereits laufenden datenintensiven Experimenten (*BABAR* im Stanford Linear Accelerator Centre *(SLAC)*, *CDF* und *D0* im Fermi National Accelerator Laboratory *(FNAL)* und *Compass* im *CERN*) genutzt werden. Deren Anforderungen stellen wohldefinierte Zwischenschritte auf dem Weg zur Durchführung der LHC-Experimente dar und sind daher hervorragend geeignet, die Funktionalität des Regionalzentrums und die Zusammenarbeit mit anderen Tier-1-Zentren unter realistischen Bedingungen zu testen.

4 Internationale Grid-Projekte

Nachdem das Potenzial des Grid-Computing zunehmend sichtbar wurde, sind in den letzten Jahren viele Grid-Projekte (www.computingportals.org) in den unterschiedlichsten Anwendungsdisziplinen initiiert und erste Grid-Umgebungen entwickelt worden. Schnell hat sich herausgestellt, dass die umfangreichen Forschungs- und Entwicklungsaufgaben nur in enger internationaler Kooperation erfolgreich bewältigt werden können. Im europäischen *E-Grid-Forum* (www.EGrid.org) und im weltweiten *Global Grid Forum* (www.Gridforum.org) treffen sich die Wissenschaftler mehrmals im Jahr zum Austausch von Ideen, Ergebnissen und Software.

Am europäischen *DataGrid*-Projekt (www.eu-datagrid.org) sind unter der Federführung des *CERN* mehr als 170 Wissenschaftler aus 21 Institutionen beteiligt. Aus Deutschland sind dies das Zuse-Institut Berlin und das Kirchhoff-Institut für Physik an der Universität Heidelberg. Das Ziel des Projektes ist, der nächsten Generation von Experimenten die zur Auswertung der Daten notwendige Computing-Infrastruktur zur Verfügung zu stellen und zu zeigen, dass Produktionsqualität über Hochgeschwindigkeitsnetze erreicht werden kann. Dies wird den Zugang einer großen Zahl von Nutzern zu einem im Prinzip unbeschränkt umfangreichen Satz experimenteller Daten ermöglichen. Dabei kann ein Nutzer entweder direkt auf die Daten zugreifen oder die Daten erst aufbereiten. Das lokale und globale Management der Ressourcen und Methoden, sowie die Zugangsverwaltung sind daher ein wichtiger Bestandteil der im Projekt zu entwickelnden Software.

Der erste Meilenstein dieses Projektes wurde mit dem Aufbau eines internationalen Testbeds im September 2001 erreicht. Da die Funktionalität des Grid kritisch von der Qualität und der Zuverlässigkeit der Datennetze abhängt, wird DataGrid das europäische (*GEANT*, www.dante.net/geant) und die nationalen Wissenschaftsnetze nutzen, um ein virtuelles privates Netzwerk zwischen den am Test beteiligten Regionalzentren aufzubauen.

Mit den analogen Projekten in den USA *GriPhyN* (Grid Physics Network, www.-griphyn.org) und *PPDG* (Particle Physics Data Grid, www.ppdg.net) wurde eine enge Zusammenarbeit vereinbart.

Die Middleware wird neben der Teilchenphysik auch in anderen Bereichen ihre Anwendung finden, z.B. in der Biologie (Genomprojekte), Medizin (Gehirnforschung) und in der Umweltforschung (Erdbeobachtung). Gemeinsam ist diesen Projekten

1. die kollaborative Datenanalyse durch sehr große Forschergruppen,

2. die enorme Komplexität der verwendeten Software und Hardware und

3. die lange Zeitskala.

Das derzeit prominenteste Grid-Projekt in Deutschland ist das vom BMBF geförderte *UNICORE*-Verbundprojekt (www.unicore.org). Hier arbeiten neun Partner unter Leitung des Forschungszentrums Jülich an der Realisierung eines intuitiven und sicheren Zugangs zu verteilten Rechner-Ressourcen im Internet.

Während in UNICORE noch an der Vervollständigung der Middleware gearbeitet wird, ist bereits ein Nachfolgeprojekt von der EU gestartet worden, das *EuroGrid* (www.eurogrid.org). Es beschäftigt sich primär mit drei wichtigen Anwendungsbereichen, die auf der Grid-Infrastruktur laufen sollen: Biologie- und Chemieanwendungen, Codes für die Wettervorhersage und Computer-unterstütztes Engineering (CAE).

Weitere von der EU geförderte Grid-Projekte in den Bereichen Tools, Infrastruktur und Anwendungen werden in Kürze starten.

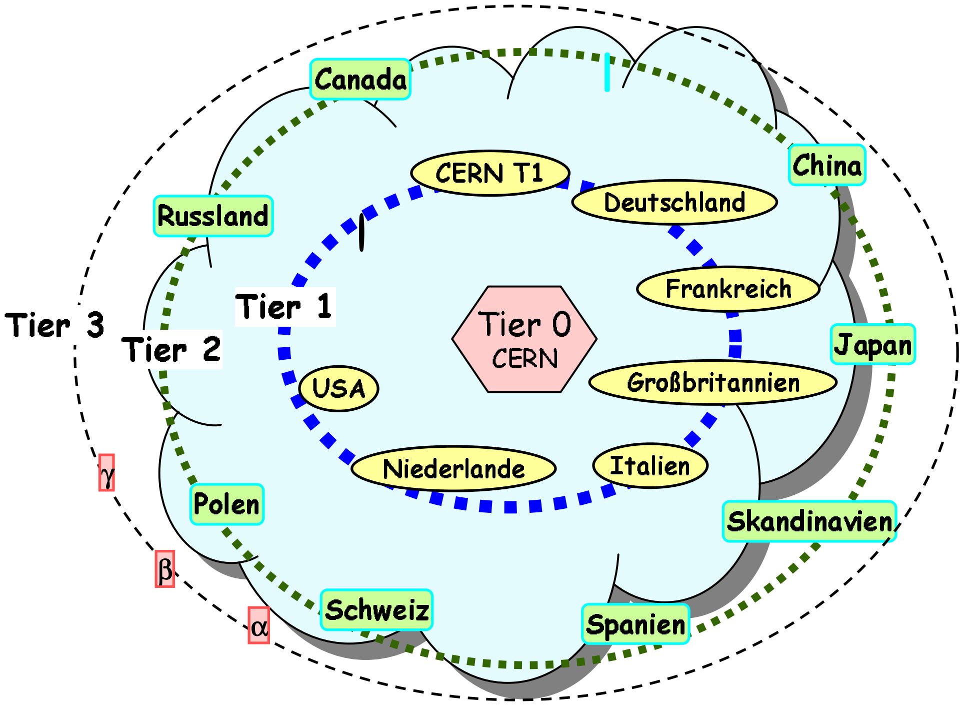

Abbildung 5: *Das LHC Computing Model: Die Grundlage des Models sind weltweit verteilte Regionalzentren, wobei jede der Ebenen (Tiers) eine bestimmte Funktion hat. Die Zentren bilden ein Grid (virtuelles LHC Rechenzentrum), auf das alle Mitglieder einer Kollaboration Zugriff haben, gesteuert von der Grid Middleware.*

5 Perspektiven

Vor etwas mehr als 10 Jahren kamen aus dem CERN die entscheidenden Anstöße zur Entwicklung des *World Wide Web*, die erst das Internet allgemein und intuitiv zugänglich gemacht haben. Die hier skizzierten neuen Anforderungen und Lösungsansätze bei der Suche nach dem Ursprung der Masse am CERN und die Ansätze zum Grid-Computing können nun vielleicht den Anstoß zum *World Wide Grid*, einer neuen Infrastruktur für Wissenschaft und Industrie, geben.

Neben der Rechenleistung stehen die Daten im Vordergrund: In vielen Wissenschaftsgebieten führt das exponentiell anwachsende Volumen und der zunehmende Detaillierungsgrad der zu verarbeitenden Daten dazu, dass der Überblick und das Verständnis der wesentlichen Züge eines Wissensgebietes verloren gehen. Hinzu kommt die Tatsache, dass man nur zögernd schwierige Aufgaben der Wissenschaft im internationalen Verbund angeht, der den Einsatz einer „kritischen Masse“ an Forschern und Mitteln erlauben könnte.

Der Direktor des englischen „Research Councils“, Dr. John Taylor, hat den Begriff „e-science“ geprägt, der diese Problematik sehr gut beschreibt und einen Ausweg aufzeigt:

> *„e-science" („enhanced science")* ist mehr als nur die Bereitstellung einer leistungsfähigen Europa- oder weltweiten Rechen-, Daten- und Anwendungsprogramminfrastruktur. Die darüber hinaus notwendige Einigung über Datendarstellung, Computernutzung und Anwendungsumgebung setzt globale, wohlorganisierte, gemeinsame Absprachen und Zusammenarbeit über Wissenschaftsfelder voraus, die die gemeinsame Forschungsleistung verstärken und zu kritischer, den Problemen angemessener „Größe" führen.

Der Aufbau einer intuitiv nutzbaren, vernetzten, Rechner-, Daten-, anwendungsgestützten, dynamischen Infrastruktur (Grid) ist dabei vorrangig und grundlegend. Das LHC-Problem mit seiner kohärenten, weltweiten Gemeinde von Forschern, die weitgehend auch „computer literate" ist, wird als günstiger Testfall angesehen und soll als solcher richtungsweisend in den Aufbau der Infrastruktur eingebunden werden.

In der Genomforschung konnte in den letzten Jahren eine ähnliche, weltweite Konzentration der Mittel und Forscher beobachtet werden, die schließlich in den beiden kompletten Publikationen des Humangenoms resultierte. Das geplante „virtuelle Observatorium" in der Astronomie ist ein weiteres Beispiel, das dem Astronomen zu jedem Objekt am Himmel alle Beobachtungen im gesamten zugänglichen Spektrum von Infrarot bis zu höchsten Energien an seinem Desktop zur Verfügung stellt.

Echte e-Science, und in der Folge dann auch erweitertes *e-Business*, bedingt zuverlässige, widerstandsfähige, sichere und hochflexible Lösungen der Echtzeit-Transaktionen. Diese werden in den bisherigen lokalen „Client-Server"-Architekturen im Netz nicht ohne weiteres unterstützt. Das Ziel besteht darin, heterogene Ressourcen, auch weit verteilt, von Fall zu Fall neu zusammenzustellen und zu nutzen – vorher schon angesprochen unter dem Begriff „Interoperabilität". Dies wird erlauben, Daten, Anwendungen und Rechenleistungen zu ad hoc-Lösungen zusammenzuschnüren, wobei Anwendungsumfeld, Lesbarkeit der Daten und Bereitstellung der Rechenleistung vom Grid und seinen offenen Standards ermöglicht wird. Der Vergleich mit dem Stromnetz bietet sich wieder an: Das Geschehen hinter der Steckdose ist „transparent", im Sinne von unerheblich für die beabsichtigte Nutzung wie Staubsauger, Laptop, Kühlschrank oder Lampe.

Zusammenfassend soll deshalb, auch mit deutscher Beteiligung,

- ein neuer Wissenschaftsverbund aufgebaut werden, der die Entwicklung der Komponenten ermöglicht, die gebraucht werden, um ein weltweites Grid zu installieren;
- die Effektivität solcher neuen Technologien durch Anwendungen gezeigt werden, die den Anforderungen der europäischen Wissenschaft entsprechen;
- demonstriert werden, dass man große Cluster und große Datenbanken aus billigen kommerziellen Komponenten zusammenbauen und in das Grid integrieren kann;
- gezeigt werden, dass der Zugang zu fundamentalen wissenschaftlichen Daten nicht länger auf den Erzeuger der Daten beschränkt ist;

- gezeigt werden, dass die Grid-Technologie einen positiven Einfluss auf zukünftige industrielle und kommerzielle Aktivitäten haben kann.

Für die konkreten nächsten Schritte und in der zunächst geplanten F&E Phase des „LHC Computing Grids“ werden zusätzliche Naturwissenschaftler und Informatiker benötigt. Gedacht ist vor allem an junge Leute nach dem Hochschulabschluss. Das CERN und die weltweit beteiligten Institute, die im engen Verbund mit extrem großen Datenmengen und anspruchsvollen Klienten arbeiten, werden dabei Ausbildungsplätze für Grid-Computing sowie für alle Ausbildungsaspekte des Hochleistungsrechnens anbieten.

Literatur

[1] I. Foster, C. Kesselman, *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann Publ., 1999.

[2] L. Smarr, C. E. Catlett, *Metacomputing*, Communications of the ACM 35,6 (1992), 45-52.

[3] Report of the Steering Group of the LHC Computing Review. CERN/LHCC/2001-004. `http://lhc-computing-review-public.web.cern.ch/lhc-computing-review-public/Public/Report\_final.PDF`

[4] MONARC Phase 2 Report CERN/LCB 2000-001. `http://monarc.web.cern.ch/MONARC/docs/phase2report/Phase2Report.pdf`