

Inhalt

Editorial	2
Iterationsschleife	6
Getting to the bottom of modelling large-scale parallel applications	9
KI-System der nächsten Generation für bayerische Spitzenforschung	13
Partnerschaftlich in die Zukunft	18
Zwei flotte Franken unter den 500 Besten	22
Der rätselhafte Fall des Programms, das zu schnell lief	26
preCICE v2.4	27
KONWIHR: New projects from spring 2022	29
SIAM UQ & ECCOMAS: Die neue Normalität oder zurück zur Normalität?	31
Historischer Erfolg beim ENB Elite-Cup 2022	34
Ausblick Ferienakademie 2022	37

Quartl* - Impressum

Herausgeber:

Prof. Dr. A. Bode, Prof. Dr. H.-J. Bungartz, Prof. Dr. U. Rüde

Redaktion:

S. Herrmann, S. Reiz, Dr. S. Zimmer

Technische Universität München, Fakultät für Informatik

Boltzmannstr. 3, 85748 Garching b. München

Tel./Fax: ++49-89-289 18611/18607

e-mail: hermasa@in.tum.de,

<https://www.cs.cit.tum.de/sccs/startseite/>

Redaktionsschluss für die nächste Ausgabe: **01.09.2022**

* Quartl: früheres bayerisches Flüssigkeitsmaß,

→ das Quart: 1/4 Kanne = 0.27 l

(Brockhaus Enzyklopädie 1972)

Das Quartl erhalten Sie online unter <https://www.cs.cit.tum.de/sccs/weiterfuehrende-informationen/quartl/>



Das Quartl ist das offizielle Mitteilungsblatt des Kompetenznetzwerks für *Technisch-Wissenschaftliches Hoch- und Höchstleistungsrechnen in Bayern* (KONWIHR) und der *Bavarian Graduate School of Computational Engineering* (BGCE)

Editorial

In dieser Ausgabe gibt es mal wieder etwas aus der Reihe „Wenn einer eine Reise tut ...“ – diesbezüglich war die letzten Jahre ja eher Funkstille. Jüngst (als diese Zeilen entstanden, war ich dem folgenden Reisechaos noch nicht entronnen) musste ich drei Tage nach Manchester, mit dem Flugzeug. Ja, ich weiß, muss das wirklich sein, kann man nicht auch hinrollern? Und überhaupt – Manchester, UK, Brexitland?? Nehmen wir es der Einfachheit halber mal als Arbeitshypothese, dass es sein musste. Nun hört man zwar allerlei Gruselgeschichten von instantanen Flugausfällen, aber irgendwie glaubt man doch immer, an einem selbst ginge das alles vorüber.

Es ging damit los, dass ich eine Woche vor dem geplanten Hinflug um 9 Uhr, Mittwoch vormittags eine Email erhielt mit „nützlichen Hinweisen zu meiner Flugreise“. Da gab es Informationen zum prognostizierten Wetter in Manchester, zu Shopping, Mietwagen oder kulturellen Angeboten etc. Ganz am Ende der Email hieß es dann noch „hier nochmal Ihre Flugdaten“ – und da stand unter „Abflug“ nicht mehr 9 Uhr, sondern auf einmal 16 Uhr. Immerhin nett, dass mich die Lufthansa überhaupt informierte, wenn auch nur in einem unscheinbaren Nebensatz. Das Dumme war nur, dass die aufoktroyierte Verschiebung alle meine Pläne hinfällig machte – denn von 12-14 Uhr war die wichtigste Sitzung in Manchester, und ab 14:15 musste ich dann von dort aus eine VC-Sitzung in München leiten. Also nix mit Abflug um 16 Uhr. Flug schon am Vortag ging nicht wegen der Vorlesung, die schon oft genug verschoben oder vertreten werden muss.

Die Reiseagentur meinte „sonst gibt's nix“, worauf ich selbst schnell die Umsteigeverbindung über Frankfurt raussuchte. Das hieß zwar um 4 Uhr in der Früh aufstehen, aber es ging wenigstens. „Ach ja“ meinte die Agentur, bedankte sich artig und buchte. Kundenfreundlichkeit und Kompetenz sehen anders aus. Seiteneffekte des Streiks bei British Rail (ja ja, der war auch noch) konnten dadurch abgedefert werden, dass ein kundiger Sekretariatsengel mir dankenswerterweise online ein Taxi vorbestellte. Und so klappte tatsächlich noch alles – am ersten Tag meiner Reise ins schöne Manchester zumindest.

Ferienakademie

2022

18.9. - 30.9.2022

Sarntal (Südtirol)

Program

Topic	Lecturers
1 Modern Approaches to Optimization and Verification in Computer Science	J. Kleinig, München F. Weira, Erlangen
2 Decentralized Decision Making and Control in Smart City Infrastructures	B. Bittig, München J. Nölke, Siemens AG (GD) A. Bauer, Siemens AG (GD)
3 Dark Matter and the Cosmos	A. Weber, München
4 Deep Learning in Computational Science and Engineering	S. Köttemann, München D. Pfleger, Stuttgart
5 Let's play! Stochastic Physics for Games	H.-J. Bungartz, München
6 Climate Change: Social, Technical, and Engineering	M. Scheuers, München W. Köttemann, Erlangen
7 Physical Models meet Deep Learning	G. von Borstel, Stuttgart (GD) S. Köttemann, München
8 Perceptual Audio Coding - From Theory to Software	B. Eber, Erlangen
9 Computational Medical Imaging	T. Lasser, München C. Rees, Erlangen

Organization:
T. Nölke, München, noelke@informatik.uni-erlangen.de
A. Bauer, Erlangen, abauer@informatik.uni-erlangen.de
F. Weira, Stuttgart, fweira@informatik.uni-erlangen.de

Universitätsbeauftragte:
G. Müller, München
A. Kemp, Erlangen
B. Schmitt, Stuttgart

Director:
H.-J. Bungartz, Institut für Informatik, TU München

www.ferienakademie.de

FAU Friedrich-Alexander-Universität Erlangen-Nürnberg

Technische Universität München

Universität Stuttgart

Abbildung 1: Das Poster der Ferienakademie 2022.

In der Folge beobachtete ich natürlich argwöhnisch Emails der Lufthansa, könnte doch auch mit dem Rückflug, für 17:45 am Freitag Nachmittag vorgesehen, entsprechend Ärgerliches geschehen. Bis 14:30 an besagtem Freitag tat sich aber nix, und so machte ich mich auf den Weg zum Flughafen. Dort angekommen, prangte auf der Anzeigetafel allerdings ein „cancelled“; und in der Tat, um 14:32 war eine entsprechende Email der Lufthansa eingetrudelt – man suche nach einer Lösung; immerhin. Anderthalb Stunden später war eine vermeintliche Lösung gefunden – Rückflug am Folgetag. Das war aber wiederum blöd, weil ich an jenem Samstag ab 8:30 Programm hatte. Am Schalter ... aber halt, um das Folgende zu verstehen, bedarf es einer kurzen Schilderung der Zustände am Flughafen Manchester.

Man möge mindestens drei Stunden vor Abflug da sein, hieß es. Die Fast Lane bewegte sich langsamer als eine Schnecke; im Gegensatz zur normalen Schlange bewegte sie sich aber zumindest. 40 Sicherheitsschalter, davon gefühlte 4 besetzt – zur Rush Hour am Freitag Nachmittag. Totales Chaos eben. Dementsprechend großes Gebrüll, und als dann noch die Flugstreichungen dazukamen, brachen alle Dämme. Zur Ehrenrettung des Flughafens von Manchester sei gesagt, dass es Ähnliches momentan leider von ganz vielen Airports zu berichten gibt.

Also: Am Lufthansa-Schalter dann, nach einigem Warten, erklärte ein völlig überforderter Mitarbeiter, es gebe nicht anderes. Seine etwas weniger überforderte Kollegin meinte zumindest, sie könne mich noch nach Frankfurt bringen, aber weiter müsse ich dann schon selbst schauen. Die Lufthansa Hotline anrufen – aussichtslos, Schleifen ohne Ende: man möge für das erhöhte Anfrageaufkommen Verständnis haben. Dieses erhöhte Aufkommen ist allerdings seit über zwei Jahren Dauerzustand.

Der Flug nach Frankfurt hatte dann 90 Minuten Verspätung, aber er fand statt. Wobei das Ausrufesystem in Manchester offenkundig statisch ist – alle Ausrufe, auch der Final Call, erfolgen zu den ursprünglich geplanten Zeiten. Anpassungen an die Realität Fehlanzeige. Und so dröhnte um 17:45,

anderthalb Stunden vor dem tatsächlichen Boarding, ein „very final call for LH flight ... to Frankfurt, we will close the flight“ durch die Gänge. Alle Betroffenen bekamen Herzflattern, sprangen auf, nahmen die Beine in die Hand, rannten zum Gate und kamen dort völlig außer Atem an – derweil die Maschine grad den Ärmelkanal passierte und das Gate noch nicht einmal mit Personal bestückt war..

Kurz nach 22 Uhr treffen wir schließlich in Frankfurt ein. Als Flughafen von Welt waren alle Service Centers der Lufthansa natürlich bereits geschlossen, und im Check-in- Bereich standen gefühlte 9.999 gestrandete Passagiere vor Schaltern an, an denen „nur Hotel“ prangte. Alle Priorisierungen hatten sich aufgelöst – aber mei, da ist schließlich Solidarität gefragt. Nur wo man sich den Anschlussflug organisieren sollte, blieb unklar: Schalter Fehlanzeige, Hotline Fehlanzeige, online Fehlanzeige. Dann die Idee mit dem Mietwagen: Was kann schöner sein, als in der Nacht von Freitag auf Samstag auf A5, A6 und A9 von FRA nach MUC zu brausen. Tja, wenn es denn am Flughafen Frankfurt noch einen Mietwagen gegeben hätte. Unisono Ansage sämtlicher präserter Firmen: „Wir haben keine Autos mehr“ – im Autoland Deutschland. Das ist mir fürwahr noch nie passiert, an keinem der von mir heimgesuchten Orte dieser Welt. Da sage noch ein Ostdeutscher, wir hätten nach der Wiedervereinigung nix von der DDR übernommen!

Also kein Hotel, kein Auto ... ah, die Bahn! Und tatsächlich, es gab einen Nacht-ICE. Vorher noch schnell zu McDonald's. Um 23:40 gaben wir die Bestellung auf, um 23:50 setzten wir uns hin, um 23:58 kam ein generiver Mitarbeiter und meinte, wir sollten verschwinden – um Mitternacht würden sie schließlich zumachen, er sei mit Kräften und Nerven am Ende und werde uns keine Minute länger sitzen lassen. Also schnell die angebissenen Burger und die Fritten eingepackt und draußen irgendwo hingesetzt. Und danach auf zum Zug. Besagter ICE sollte um 0:57 von Gleis 4 vom Fernbahnhof abfahren. Er fuhr dann schließlich um 01:27 von Gleis 7 ab. Das allein war noch nicht spektakulär. Aber dies hier schon: Geschlagene drei Mal sprang die Anzeige zwischen 00:50 und 01:15 von Gleis 4 auf Gleis 7 um und zurück, und immer wieder setzte sich die gesamte am Bahnsteig wartende Meute

Ferienakademie 2022

FERIENAKADEMIE

Im Herbst 2022 wird zum 38.ten Mal die Ferienakademie⁸ im Samtal (Südtirol) stattfinden, die von der TU München, der FAU Erlangen-Nürnberg und der Universität Stuttgart gemeinsam organisiert wird.

In diesem Jahr wird es insgesamt neun Kurse zu ganz unterschiedlichen Themen geben:

1. Modern Approaches to Optimization and Verification in Computer Science
2. Decentralized Decision Making and Simulation Models in Smart City Infrastructures
3. Dark Matter and the Cosmos
4. Deep Learning in Computational Science and Engineering
5. Let's play! Simulated Physics for Games
6. Climate Change: Social, Technical, and Political Dimensions
7. Physical Models meets Deep Learning
8. Perceptual Audio Coding - From Theory to Software
9. Computational Medical Imaging

Details zu den Kursen sind unter <https://www.ferienakademie.de/kurse-2022/> sowie unter <https://youtu.be/hDzz30KTRMM> zu finden.

Tobias Neckel

⁸www.ferienakademie.de

ab. Nach unglücklicher Schiedsrichterentscheidung ging CDTM in Führung⁷. Jedoch ließen wir uns nicht unterkriegen und erzielten kurz vor Ende der regulären Spielzeit den mehr als verdienten Ausgleich zum 1:1. So ging es also ins Neumeterschießen. Auch hier war es ein hin und her, die Stimmung war angespannt ob der Wichtigkeit des Ergebnisses und es gab verschossene Elfmeter auf beiden Seiten. Trotz starker Torwartleistungen unseres BGCE-Titanen, Steffen Seckler, mussten wir uns dann letztlich mit 3:4 geschlagen geben.

Enttäuscht, aber erhobenes Hauptes gingen wir daher in das Spiel um Platz 3. Wir hatten noch historisches vor: das beste Ergebnis eines BGCE-Teams, ein vierter Platz, liegt schon einige Jahre zurück. Daher konnten wir an diesem Tag Geschichte schreiben und mit dem Gewinn des Spiels die beste Platzierung aller Zeiten für BGCE erreichen. Und nicht nur das machte es zu einem besonderen Spiel: wir bekamen auch die Chance auf Revanche, trafen wir doch auf den Widersacher unserer einzigen Niederlage aus der Gruppenphase, den Wirtschaftlern aus Regensburg. Diese Motivation nahmen wir mit ins Spiel und gaben unseren Gegnern keine Luft zu atmen. Hungrig wie wir waren, sah die Mannschaft aus Regensburg kein Land und wir konnten den dritten Platz mit einem deutlichen 3:0 für uns gewinnen.

Das Finale gewann im Anschluss die Technologischen und Mathematischen Physiker der LMU deutlich mit 3:1 gegen CDTM. Daher kehrt der Elite-Cup auch im Jahr 2023 wieder nach München zurück. Mit der Motivation und dem Erfolg aus diesem Jahr sind wir bereit, nach den Sternen zu greifen und den Pokal im nächsten Jahr nach Hause zu holen.

Friedrich Menhorn

über Rolltreppen und Brücke über die Gleise in Bewegung. Alle suchten nach der versteckten Kamera, so absurd war das – aber nein, das war kein Slapstick, sondern blütenreine Realsatire.

Aber egal. Irgendwann saßen wir im Zug, hatten sogar drei Sitzplätze ergattert, und tuckerten mit der genannten Verspätung los. Und was niemand für möglich gehalten hatte, geschah doch – wir kamen nahezu pünktlich am Münchener Hauptbahnhof an. Jetzt noch schnell mit der S-Bahn nach MUC gedüst (schließlich warteten dort ja unsere Autos auf uns), und schon neigte sich die Reise ihrem Ende zu, und meinem Einsatz am Samstag ab 8:30 stand nichts im Wege.

Alles gut also? Nee, nicht wirklich. Und es ist noch nicht zu Ende – schließlich müssen die Mehrausgaben noch bei der Lufthansa geltend gemacht werden. Mal schauen, was da so passieren wird. Aber jetzt war erst mal Wochenende. Mal ehrlich – wenn das zum Dauerzustand wird, wird die Reiserei von allein abnehmen, ohne Zwangsmaßnahmen. Das Klima wird sich freuen.

Doch damit genug für heute. Die gesamte Quartl-Redaktion wünscht Ihnen einen schönen Sommer – mit viel Erholung und wenig Flugausfällen! Zunächst aber natürlich wünschen wir Ihnen viel Vergnügen mit der neusten Ausgabe Ihres Quartils!

Hans-Joachim Bungartz.

⁷Der Ball war zuvor klar im Seitenaus gelandet, jedoch ließ der Schiedsrichter weiterspielen.

Iterationsschleife

N=43 08. Mai 2022

Die Wahrheit

Im Krieg – so heißt es – stirbt als erstes die Wahrheit. Der Grundgedanke lässt sich niemandem als Zitat zuordnen. Weder findet er sich – wie oft behauptet wird – bei Aischylos^a noch in einer Rede des amerikanischen Senators Hiram Johnson^b – wie gerne behauptet wird. Die Behauptung, dass die Wahrheit im Krieg als erstes stirbt ist also von Unwahrheiten über ihren Ursprung umgeben. Trotzdem erscheint sie uns als Wahrheit. Wir halten es für wahr, dass im Krieg das Wahre verschwindet – logischerweise wäre dann der Verlust der Wahrheit das einzige was an Wahrheit zurückbliebe. Das Fehlen der Wahrheit wäre die einzige Wahrheit.

Das ist paradox. Wenn im Krieg die Wahrheit stirbt, so gibt es keine Möglichkeit mehr festzustellen, ob es wahr ist, dass die Wahrheit stirbt. Tatsächlich löst der Krieg die Wahrheit ab und macht damit die Wahrheit erst verständlich und sichtbar. Denn was mit dem Krieg stirbt ist nicht die Wahrheit, sondern unsere Fähigkeit, uns eine Wahrheit auszuhandeln und zu konstruieren. Belügen wir uns also ständig selber? Nicht wirklich. Aber wir müssen als Menschen ständig unserer Vorstellung von Wahrheit neue Informationen, Erkenntnisse, Erschütterungen, Überraschungen, Enttäuschungen und vieles mehr hinzufügen. Die Welt, wie wir sie wahrnehmen, fordert von uns eine ständige Neujusterung unseres Weltbildes. Das ist der Fluch des modernen Menschen, der sich das Recht zur eigenen Meinung erkauft mit der Verpflichtung, sich diese Meinung zu bilden. Es ist der Fluch des digitalen Menschen, der mit der Möglichkeit alles jederzeit wissen zu können, auch die Pflicht erbt – oder erleidet – alles jederzeit wissen zu müssen.

Was wahr ist, ist nur in einer völlig rationalen Welt einfach zu klären. Lesen wir die aktuelle Temperatur von einem Thermometer ab, so sind wir uns über den Wert normalerweise einig^c. Die „harten“ Wissenschaften sind also auf der sicheren Seite. Und um sie soll es hier auch nicht gehen.

^aGriechischer Dichter (525 v. Chr. – 456 v. Chr.) bekannt für seine Tragödien

^bAmerikanischer Politiker (1866 – 1945)

^cSofern wir die gleiche Skala (z.B. Celsius, Kelvin, Fahrenheit) verwenden



Abbildung 1: Das BGCE-Team 2022.

2018, den Honors Wirtschaftswissenschaften aus Regensburg. Wie eine typische Turniermannschaft verbesserte sich aber unsere Leistung von Spiel zu Spiel, und somit konnten in den darauffolgenden Gruppenspielen klare Siege gegen Advanced Signals and Communication und dem Master Program in Advanced Optical Technologies, beide aus Erlangen, sowie Global Change Ecology aus Bayreuth erzielt werden. Da die Wirtschaftler aus Regensburg ungeschlagen blieben, zogen wir als Gruppenzweite in die Playoffs.

Im Halbfinale mussten wir gegen einen Koloss des Elite-Cups in den Ring steigen. Wir trafen auf das CDTM, den Turnierorganisator und eine Größe des ENB Fußballs. Doch wir nahmen die Außenseiterrolle an und warfen uns mit allen Kräften in den Kampf. Und was für ein Kampf es war. Bei glühender Mittagshitze, kurz nach dem Genuss von Schweinebraten, Klößen, Käsespätzle und isotonischen Getränken, verlangten wir uns gegenseitig alles

Ein Tag für die Geschichtsbücher – Historischer Erfolg beim ENB Elite- Cup 2022



Der Elite-Cup des Elitenetzwerk Bayerns (ENB), das nach der Herren-Fußballweltmeisterschaft zweitwichtigste Ereignis Fußball-Deutschlands 2022, kehrte im Juni nach einem Ausflug nach Regensburg im Jahr 2019 und einer zweijährigen Corona-Auszeit bei prachtvoller Wetter wieder zurück nach München. Traditionell lädt das ENB seine Elitestudiengänge, Doktorandenkollegs und Max-Weber-Stipendiaten im Sommer zum Kampf um Ruhm und Ehre im Fußball ein. Auf dem schön gelegenen Rasenplatz an der U-Bahn Station Kiefergarten kämpften also 10 Mannschaften aus ganz Bayern um die Krone des (bayrischen, akademischen) Fußballs. Organisiert wurde das Turnier vom Sieger des Jahres 2019, dem Center for Digital Technology and Management (CDTM).

Auch wir, die Bavarian Graduate School of Computational Engineering (BGCE), waren wieder mit einem schlagkräftigen Team vertreten. Mit zehn Spielern aus den Programmen Computational Science and Engineering, Computational Mechanics und Alumnus, und ohne gemeinsames Training hatten wir Großes vor. Und soviel sei schon einmal vorweg genommen: es ist uns gelungen.

Das Turnier startete mit einer Gruppenphase. Je zwei Fünfer-Gruppen spielten parallel auf zwei Plätzen im Round-Robin-Modus gegeneinander. Die ersten beiden Teams aus jeder Gruppe trafen sich dann im Halbfinale wieder, um darauf folgend noch ein Spiel um Platz 3 und das große Finale auszuspielen. Jedes Spiel wurde im Format 7 gegen 7 auf Kleinfeld gespielt und ging über 20 Minuten.

Im ersten Spiel hatten wir noch Anlaufschwierigkeiten und verloren so gleich nach kurzfristigem Ausgleich mit 1:2 gegen die Sieger aus dem Jahr

Auch die Mathematik ist, wenn sie Gödel in den blinden Fleck der Absurditäten verschiebt – wie es die Theologie mit ihrem „Credo, quia absurdum est“^a tut – eine Wissenschaft in der alle Aussagen in die Kategorien des Wahren oder Falschen fallen, womit die Wahrheit als solche als Kompendium mathematisch bewiesener Aussagen zusammengefasst werden kann, deren Eindeutigkeit nur durch „Vermutungen“ gestört wird für deren Beseitigung mitunter Preise ausgerufen werden.

Der größere Teil der Welt die uns umgibt ist aber nicht messbar oder auch nur rational beschreibbar. Die Aussage „Quidquid id est, timeo danaos et dona ferentes“^b lässt sich nicht als wahr oder falsch qualifizieren. Wir mögen dem Priester Laokoon glauben, dass er die Griechen fürchtet, aber nur wenn wir diese Furcht teilen, können wir auch die Wahrheit, die sich für Laokoon daraus ergibt, teilen. Auf diese Art wird die Wahrheit jenseits der messbaren Welt immer zu einem zwischen den Individuen und der Allgemeinheit ausgehandeltem Konstrukt, das sich nicht zu 100% auf reine rational überprüfbare Fakten stützen kann.

Es wäre absurd anzunehmen, dass es damit keine wahren und falschen Dinge mehr gibt, oder dass alle Wahrheit relativ sei. Im Gegenteil, ist alle Wahrheit immer absolut und ist unser Ziel auch jederzeit, die Sicherstellung der Absolutheit der Wahrheit. Nichts liegt uns als westlichen Menschen ferner als einen Relativismus in Bezug auf die Wahrheit zu wollen. Wir sind deshalb bereit – und darin geübt – unsere Wahrheit jederzeit in Frage zu stellen, wenn sich neue Hinweise auf die wirkliche Wirklichkeit oder die wahre Wahrheit ergeben. Wissend, dass wir irren können, nehmen wir Täuschungen wahr und sind nicht enttäuscht, wenn wir enttäuscht werden, weil die Enttäuschung eben den Schleier der Täuschung von der Wahrheit nimmt und die Wahrheit unverhüllt zeigt. Die Enttäuschung ist also der Triumph der Wahrheit und damit unser Ziel.

^aZu übersetzen etwa als „Ich glaube es, weil es widersinnig ist“. Auch für dieses Zitat findet sich kein echter Beleg, auch wenn es als „gefügelteres Wort“ gilt

^b„ich fürchte die Griechen auch wenn sie Geschenke bringen“. Aussage des Priesters Laokoon nach Vergil, Aeneis, xxxix

Da Wahrheit aber Rationalität voraussetzt ist jeder Bruch in der Rationalität auch ein Bruch in unserer Wahrheit. Damit wird der Krieg zur Gefahr für die Wahrheit. Einen Krieg zu beginnen ist nie rational. Es widerspricht unserem Denken das zu tun. Wir sehen Krieg als Pflicht, als Notwehr oder als moralisch geboten an. Die Logik des Rationalen sagt uns aber, dass der Krieg mitnichten nur der Vater aller Dinge^a, sondern auch der Zerstörer aller Dinge ist. Nach dem Krieg muss die Wahrheit neu gesucht, gegründet und aufgebaut werden.

M. Resch

^aDieses Zitat von Heraklit lässt sich tatsächlich in seinen Fragmenten (B 53) finden. Vollständig zeigt Heraklit aber auf mehr, als nur den Krieg als Vater aller Dinge, wie wir im westlichen Sinn das meist technisch verstehen. Heraklit bemerkt: Der Krieg ist der Vater aller Dinge und der König aller. Die einen macht er zu Göttern, die anderen zu Menschen, die einen zu Sklaven, die anderen zu Freien.

vermutlich behilflich sein können, um agil darauf reagieren und sie online einbinden zu können.

Nach den Eindrücken beider Konferenzen stellt sich mir die Frage, wo die Reise wohl hingehen wird? Wird es weiterhin ein Onlineangebot geben oder werden wir zurückkehren zu Konferenzen, die nur eine Teilnahme vor Ort erlauben? Wie eigentlich immer, gibt es meiner Meinung nach keine klare Antwort und beide Angebote haben ihre Vor- und Nachteile. Für mich hat das hybride Angebot immer noch einen großen Reiz, da es grundsätzlich viel agiler ist und die Erreichbarkeit vergrößert. Auch die unattraktiven Aspekte des Reisens, wie die Kosten, der Zeitaufwand und auch die Umweltbelastung, fallen dadurch weg. Das große Manko sind meiner Meinung nach aber die technischen Herausforderungen, die sich nie ganz unterbinden lassen werden, sowie die fehlende Interaktion für Onlineteilnehmer:innen; einer der Hauptaspekte für die Teilnahme an einer Konferenz und etwas, für das sich online noch keine wirklich gute Alternative herauskristallisiert hat. Schlussendlich ist es aber vielleicht auch einfach das monetäre Argument, das für die Konferenzorganisatoren entscheidet: mit welchem Angebot lässt sich mehr Geld verdienen? Wir werden sehen, was die zukünftigen Konferenzen anbieten. Ein Blick in meinen Konferenzkalender zeigt noch keinen klaren Gewinner — reine Onlinekonferenzen sehe ich allerdings auch nicht mehr.

Friedrich Menhorn

häufiger technische Probleme. Das klassische „Can you hear me?“ oder „Can you see my screen?“ zur Begrüßung wurde des Öfteren vernommen. Auch Verbindungsabbrüche waren überraschenderweise relativ häufig, entweder auf Seiten der Vortragenden, aber auch auf der Seite des Hotels, vermutlich wegen Überlastung. Es ist im ersten Moment doch überraschend, das nach all den Jahren und dem technischen Fortschritt, ein stabiles System immer noch schwer zu erreichen ist.⁶ Auch war die Interaktion nicht so gegeben, wie wenn alle Teilnehmer:innen vor Ort gewesen wären. Es kam häufiger vor, dass selbst die Minisymposia Organisatoren nur online zugeschaltet waren, was dazu führte, dass die Teilnahme vor Ort deutlich abnahm. (Teilweise sah man komplett leere Räume, in denen ein reines Zoommeeting ablief). So wurde, überspitzt gesagt, aus einem „das Beste aus beiden Welten“ eher ein „nichts halbes und nichts ganzes“. Nichtsdestotrotz war vor allem das Treffen und die Interaktion mit Kollegen, die man zwei Jahre nicht gesehen hatte, ein Highlight.

Der ECCOMAS Kongress 2022 hingegen kehrte zurück zur Normalität. Als klassische Konferenz wie zu Vor-Corona-Zeiten gab es kein Onlineangebot und alles fand vor Ort statt. Die einzige Ausnahme gab es für Invited Speakers, deren Video teilweise abgespielt wurde. Das ergab sich aber eher aus der Not heraus, da manche Vortragende kurzfristig nicht anreisen konnten. Norwegen (oder Skandinavien) war ja seit Beginn der Pandemie relativ lax im Hinblick auf ihre Coronaregeln, und so gab es keine Vorgaben mehr hinsichtlich Masken oder gar Impfung. Und tatsächlich hatte man sich schnell wieder an die alten Abläufe gewöhnt, saß in vollgefüllten Vortragsräumen, stand dicht an dicht in den Schlangen bei der Kaffeepause oder genoß ein 3-Gänge-Menü im gefüllten Saal beim Konferenzdinner. Auffallend war nur die fehlende Teilnahme vieler US-Kolleg:innen, da die meisten wohl wegen der strengen Rückreiseregeln in die Vereinigten Staaten auf eine Anreise verzichteten. Das war natürlich schade und hier hätte ein hybrides Angebot

⁶Wenn man dann jedoch miteinbezieht, dass jede ihren eigenen Laptop verwenden möchte, mit verschiedenen Betriebssystemen und Anschlüssen, ist es dann natürlich auch wieder zu erwarten.

ISC'21: First Place PhD Forum Award

**KONWIHR Erlangen:
Getting to the bottom of modelling
large-scale parallel applications**



Ayesha Afzal, doctoral student at the professorship for High Performance Computing (Prof. G. Wellein) and at the Erlangen National High Performance Computing Center (NHR@FAU) at Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany, was awarded first place in PhD Forum Award at the ISC event 2021.

White-box analytic performance models provide insight into performance limitations of highly parallel applications. However, coupling analytic models across the system hierarchy (socket, node, network) fails to encompass the intricate interplay between the program and the hardware, especially when execution and bottlenecks are involved. To encounter this great challenge, an in-depth investigation was performed in KONWIHR project Erlangen on why simply adding communication modelled time to computation modelled time does not often yield an adequate estimate of parallel run time. This is due to deviations from the expected bulk-synchronous execution; processes get out of sync even for extremely regular programs comprising homogeneous compute-communicate phases and a perfect translational symmetry across processes. This modelling work has been summarized as “Noise-Driven Cluster-Level Performance Modelling and Engineering”, and was awarded the “PhD Forum Award - 1st Place” at the ISC 2021.



ISC 2021 DIGITAL

PHD FORUM AWARD - 1st Place

Noise-Driven Cluster-Level Performance
Modelling and Engineering

Ayesha Afzal, Georg Hager, Gerhard Wellein

Friedrich-Alexander-University
Erlangen-Nuremberg (FAU) / Germany



CONGRATULATIONS!
ISC 2021 DIGITAL | JUNE 28, 2021

Olga Pearce
Olga Pearce
PhD Forum
Program Chair

Abhinav Bhatnagar
Abhinav Bhatnagar
PhD Forum
Program Deputy Chair

Martin Meurer
Martin Meurer
ISC General Co-Chair

Thomas Heuer
Thomas Heuer
ISC General Co-Chair

ISC HIGH PERFORMANCE 2021 DIGITAL | JUNE 24 - JULY 2, 2021 | isc-hpc.com

Figure 1: The “First Place PhD Forum Award” certificate. In addition, Ayesha Afzal presented the results clearly and concisely in her presentation, which is available at [swapcard](https://www.swapcard.com) link and is accessible to the registered participants of the ISC-HPC Digital event. For those interested, an extended version of the lecture is publicly available on the YouTube channel of the NHR@FAU.

Afzal et. al. developed a validated analytic model of the propagation speed of *idle waves*.

**SIAM UQ & ECCOMAS:
Die neue Normalität oder
zurück zur Normalität?**



Nach zwei Jahren mit Onlinekonferenzen, späten Abend- oder Nachtvorträgen am Laptop wegen Zeitverschiebung, Kaffeepausen in Gathertown und Vorträgen mit technischen Problemen oder gleich als Videoaufzeichnungen, ist es nun endlich wieder möglich, Konferenzen stattfinden zu lassen. So konnte ich im April 2022 die SIAM Conference on Uncertainty Quantification (UQ22) in Atlanta besuchen. Zwei Monate darauf kam dann gleich die nächste Konferenz mit dem ECCOMAS Kongress 2022, der in Oslo im Juni stattfand. Während UQ22 noch als hybride Konferenz angeboten wurde, verzichteten die Veranstalter von ECCOMAS komplett auf ein Onlineangebot und waren somit meine erste „fully in-person“ Konferenz, wie es jetzt distinktiv bezeichnet wird, seit zwei Jahren.

SIAM startete den Versuch einer hybriden Konferenz: auf der UQ22 waren alle Vorträge, ob Plenary, Semi-Plenary, Minisymposia, wie auch Poster komplett hybrid. Jedes Symposium hatte ein Zoommeeting zugeordnet und vom Konferenzhotel gab es Angestellte, die sich in jedem Raum um die Technik, also Video- und Audioübertragung, kümmerten. So konnte sowohl die Vortragenden vor Ort, als auch die Vortragenden online, ihre Präsentationen an alle Zuhörer:innen im Hotel oder verteilt auf der Welt übertragen. Das hatte doch einige Vorteile: so erreichte man ein deutlich breiteres Publikum. Durch die sehr angespannte Coronalage im April in vielen Ländern, war es vielen nicht möglich zu reisen. Auch vor Ort konnte man sich, wenn man etwas Ruhe wollte, auch einmal remote in eine Runde einwählen und damit auch leichter zwischen verschiedenen Symposia springen. Und als Teilnehmer vor Ort boten sich die gewohnten Möglichkeiten zur Interaktion und Netzwerken; treffen zwischen den Meetings, am Abend oder während den Kaffeepausen waren endlich wieder möglich. Andererseits ist es dann aber auch nicht so, dass hybrid das beste aus Online und Präsenz vereint. Natürlich gab es

- *Improving FAIRness of HPC research data* – Prof. C. Stemmer, N. Hoppe, B. Farnbacher (NFDI4Ing, TUM)
- *HPC-Stützpunkt at the University of Regensburg* – Prof. T. Wettig (High Energy Physics, Univ. Regensburg)

You can find more details about these projects at

<https://www.konwihhr.de/konwihhr-projects/>

We would like to invite you to our online workshop on October 12, 15:00–18:00, in which new projects will present their goals and challenges. For more details, please watch our konwihhr-announcements mailing list.

About KONWIHR

The main objective of KONWIHR is to provide technical support for the use of high performance computers and to expand their deployment potential through research and development projects. Close cooperation between disciplines, users and participating computer centres as well as efficient transfer and fast application of the results are important. Read more on konwihhr.de.

Contact KONWIHR

For any KONWIHR inquiries, you only need one address:

info@konwihhr.de

Your email will be read carefully and answered by Katrin Nusser or Gerasimos Chourdakis, KONWIHR’s current contact people in the Bavarian North and South. Together with Prof. Gerhard Wellein and Prof. Hans-Joachim Bungartz (who you can also reach using the same address), we collect and process your proposals two times per year (1st of March and 1st of September). Learn more about how you can apply for funding at:

<https://www.konwihhr.de/how-to-apply/>

Katrin Nusser, Gerasimos Chourdakis

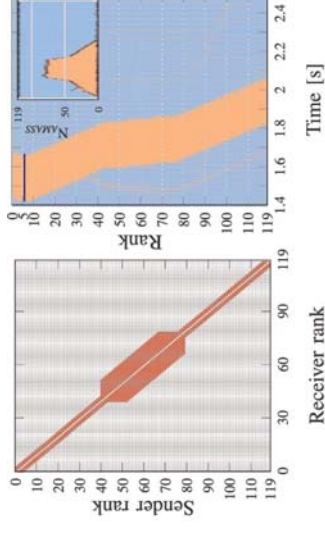


Figure 2: *Idle waves* emerge when disturbances emanating from different sources on individual MPI processes propagate across the others depending on the execution and communication properties of the program. This propagation is much like a train delay that causes other trains to wait and thus “ripples” through the schedule.

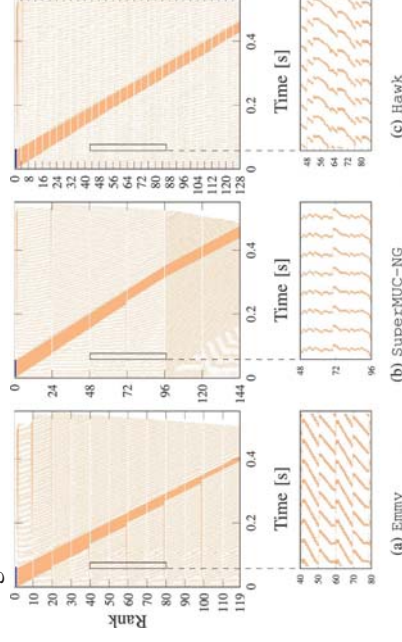


Figure 3: *Idle wave* decay on the systems running scalable workload using non-blocking MPI distance-1 communication (1 MB messages, computation in white color and injected extra work in blue color at rank 0). Topological boundaries exist at every 10, 24, and 4 cores on Emmy, SuperMUC-NG and Hawk.

Using a spectrum of HPC resources and widespread application scenarios, we further explored how *idle waves* interact nonlinearly and decay within a parallel code on a cluster. Other than imbalanced application workload and system noise, the primary factors in *idle wave* decay include the differences in communication characteristics among parts of the system (i.e., intrasocket vs. intersocket vs. internode communication); see Figure 3.

We developed insight into how typical hardware bandwidth bottlenecks among parallel processes impact parallel execution. Despite causing increased per-process waiting time, the desynchronized state via bottleneck evasion may exhibit higher performance as it can boost the per-process memory bandwidth and facilitate an automatic communication and computation overlap. We discovered the decisive factors for this effect, such as the minimum number of processes per memory domain required to achieve full memory bandwidth on memory-bound programs. We have shown that some MPI collectives, depending on their particular implementation, are permeable to *idle waves* and thus allow out-of-lockstep behavior to persist. This result extends the scope of the research beyond collective-avoiding parallel programs. We developed a lightweight message-passing simulation toolkit based on first-principles analytic performance models and a domain-specific language (DSL). It simulates large-scale applications by taking the socket-level performance properties of the code into account to reproduce and explore the complex dynamics of parallel programs on current and future supercomputers in a well-controlled environment.

Ayesha Afzal

New projects from spring 2022



The competence network for scientific high-performance computing in Bavaria welcomes the new projects that succeeded in the application round of spring 2022. In every round, we accept proposals for “normal” (up to 12 months) and “small” (up to 3 months) projects, as well as “basis” projects to establish contact points. In this round, the following projects were funded:

- *Optimization of the ALF implementation of the auxiliary-field quantum Monte Carlo algorithm: porting to GPUS and symmetry considerations* – Prof. F. Assaad, Dr. M. Ulybyshev, Dr. J. Portela, J. Schwab (Theoretische Physik, Univ. Würzburg)
- *HPC mixed precision quantization of encoder-decoder deep neural networks* – Prof. A. Kist (Artificial Intelligence in Communication Disorders, FAU)
- *Continuous Benchmarking for the GHODDESS framework* – Prof. H. Köstler, Prof. U. Rüde (LSS, FAU), Prof. V. Aizinger, Prof. M. Bebandorf (Scientific Computing, Univ. Bayreuth)
- *Massively parallel solvers for geophysical flow problems with strong viscosity variations* – Prof. U. Rüde (LSS, FAU), Dr. M. Mohr (Earth and Environmental Sciences Geophysics, LMU)
- *Porting of Lattice QCD simulation software to GPUs* – Prof. T. Wettig (High Energy Physics, Univ. Regensburg)
- *“ALPACA in Florence – first steps towards the Ponte Vecchio”*: *Porting ALPACA to Intel’s upcoming GPU architecture Ponte Vecchio* – Prof. N. Adams, Dr. S. Adami, N. Fleischmann, J. Winter (AER, TUM)

identified and deprecated features that were once useful and have become an unnecessary burden in the meantime. That's often complicated, because we don't want to break our user's code. But preCICE v3.0 is slowly approaching and this will allow us to (gently) break things. This means: If you have a favourite preCICE function that is barely used by anybody else, you should check whether it is in our list of candidates for deprecation.⁴

On the other hand, we have renamed parts of the user interface to make it easier to understand and future-proof: preCICE helps two (or more) independent participants (simulation programs, if you are from outside the preCICE bubble) to communicate with each other. This is especially complicated, if the participants are running on multiple ranks in parallel.

Here, preCICE uses a master-slave pattern to initialize the communication channels between the individual ranks of the participants. And there we are: The term *master-slave* is problematic and, therefore, we decided to deal with the hard problem of renaming it. So, what terminology do we want to use in the future? Here, wikipedia is a great help for finding alternatives⁵ and Markdown helped us to create a nice table (see Figure 1). We discussed and finally solved this hard problem of computer science in issue #1026.

The alternative term *pilot-rower*, inspired by team sports, definitely deserves an honorable mention – renaming definitely needs some creativity. However, we concluded to stick to one of the wikipedia-inspired alternatives, to not confuse us and our users too much, and use the terminology *primary-secondary* instead of *master-slave* in the future. Is the story over? Unfortunately not. There was still the task of actually replacing all occurrences of *master-slave* in the codebase, which also means a lot of effort. For the full story with renamed function names, log messages and everything else: see pull request #1253 that finally ended up in v2.4. But now we are finally there: we now use an *intra-comm* and not a *master-slave* communication scheme anymore and preCICE v2.4 is released! One of the next big tasks is to rename our *master* branch, but we are well prepared for the next challenge!

Benjamin Rodenberg

⁴Currently: issue #1248 and issue #1251.

⁵[https://en.wikipedia.org/wiki/Master/slave_\(technology\)](https://en.wikipedia.org/wiki/Master/slave_(technology))

KI-System der nächsten Generation für bayerische Spitzenforschung



Leibniz-Rechenzentrum
der Bayerischen Akademie der Wissenschaften

Als erstes wissenschaftliches Rechenzentrum in Europa bietet das LRZ Forschenden Zugang zu einem Cerebras CS-2-System mit HPE Superdome Flex Servern – einer wegweisenden Computerarchitektur für die Verarbeitung größter Datenmengen mit KI- und Machine Learning-Methoden.

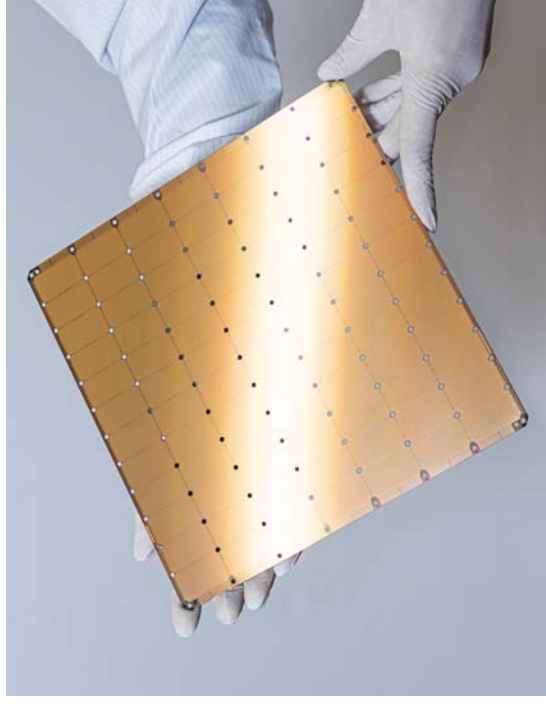


Abbildung 1: Die Cerebras Wafer Scale-Engine 2 - der derzeit größte Chip weltweit (Foto: Cerebras Systems)

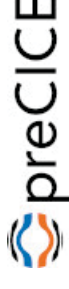
Das Leibniz-Rechenzentrum (LRZ) der Bayerischen Akademie der Wissenschaften erweitert seine Hochleistungsinfrastruktur um ein Cerebras CS-2-System mit HPE Superdome Flex Servern. Es bietet damit seinen Nutzerinnen

und Nutzern – allen voran der bayerischen Spitzenforschung im Bereich KI – die nötigen Rechenkapazitäten, um große Datenvolumina mit Methoden der Künstlichen Intelligenz (KI) und des Maschinellen Lernens (ML) auszuwerten. Als erstes wissenschaftliches Rechenzentrum in Europa setzt das LRZ damit auf die Wafer Scale Engine 2 (WSE-2), den bis dato größten weltweit gefertigten Chip mit der höchsten Rechenleistung für smarte Datenverarbeitung. „Manchmal kommt es eben doch auf die Größe an – vor allem, wenn es um die Verarbeitung größter Datenmengen geht. Methoden von KI und das maschinelle Lernen brauchen viel Rechenkraft. Der neue Superchip des LRZ ist der bis dato größte weltweit und wird der bayerischen Spitzenforschung neue Wege öffnen. Mit ihm stärken wir nachhaltig den KI-Standort Bayern. Finanziert wird der Chip über Mittel aus der Hightech Agenda Bayern, der milliardenschweren Technologieoffensive des Freistaats“, so der Bayerische Staatsminister für Wissenschaft und Kunst, Markus Blume.

Turbo für bayerische KI-Forschung

Das LRZ erweitert kontinuierlich sein Portfolio heterogener Recheninfrastruktur, das neben Haupt- und Grafikprozessoren auch Spezialhardware wie FPGAs oder ASICs umfasst. Mit dem Cerebras-CS-2-System mit HPE Superdome Flex setzt das LRZ auf ein hoch leistungsfähiges System für die Verarbeitung von Big Data in Wissenschaft und Forschung: „Aktuell beobachten wir, dass sich der KI-Rechenbedarf unserer Nutzerinnen und Nutzer alle drei bis vier Monate verdoppelt. Mit der hohen Integration von Prozessoren, Speichern und On-Board-Netzwerken auf einem einzigen Chip ermöglicht das neue System Hochleistung und Tempo. Das verspricht deutlich mehr Effizienz bei der Datenverarbeitung und somit schneller bahnbrechende wissenschaftliche Erkenntnisse“, beschreibt Prof. Dr. Dieter Kranzlmüller, Leiter des LRZ, die Vorteile. „Als akademisches Rechen- und nationales Supercomputingzentrum stellen wir Forschenden hochperformante, zuverlässige IT-Dienstleistungen bereit. Um eine effiziente Nutzung sicherzustellen und Spitzenforschung auf dem System zu ermöglichen, werden wir gemeinsam mit unseren Nutzerinnen und Nutzern sowie unseren Partnern Cerebras und Hewlett Packard Enterprise ideale Anwendungsszenarien identifizieren und an deren Umsetzung feilen.“ Zu den möglichen Use Cases zählen die

preCICE v2.4



There are 2 hard problems in computer science: cache invalidation, naming things, and off-by-1 errors.² We will see in a moment what this has to do with preCICE, but first things first: We just released preCICE v2.4! There are many new features and several bugfixes and, as far as we know, no issues with off-by-1 errors. To just name a few new features: We added Julia to the list of programming languages supported by preCICE and there is the new experimental features time interpolation that is waiting for its first users.³

Nr	master	slave	Opinion
1	master	follower	👍👍
2	leader	follower	👍👍
3	primary	secondary	👍👎
4	main	secondary	👍👎
5	agency	agent	👍👍
6	coordinator	coordinated	👍👍
7	coordinator	node	👍
8	controller	worker	👍👍
9	supervisor	worker	👍👍
10	parent	child	👍👍

Figure 1: Alternatives to *master-slave* from issue #1026 discussed by the preCICE developers.

We are also constantly cleaning up the codebase to keep it tidy. There were two main things in this direction this release: On the one hand, we

²The original statement without the *off-by-1 errors* is by Phil Karlton, but I prefer this version by Leon Bambrick, <https://twitter.com/secretGeek/status/7269997868>

³Refer to our post on Discourse for a full list: <https://precice.discourse.group/t/highlights-of-the-new-precice-release-v2-4>

Code-Beschleunigung durch CUDA – ohne GPU: Der rätselhafte Fall des Programms, das zu schnell lief



Ich hätte nicht GPUs zu meinem Spezialgebiet machen müssen, wenn die Verwendung von CUDA nicht in der Regel auch zu Speed-ups führen würde. Wenn allerdings, wie die Autoren der Meeresströmungssimulation GHODESS feststellten, bereits ein strategisch platzierter Aufruf der Funktion *cuCtxCreate*¹ zu speed-ups von bis zu 70% führt, so muss an einer derart universellen Wirksamkeit von GPUs doch gezweifelt werden.

Eine Messung des L2-Cache-Datenvolumens mit LIKWID liefert das erste Indiz: Wird der (fehlerhafterweise übriggebliebene) Funktionsaufruf entfernt, steigt das L1- zum L2-Cache- Datenvolumen an und die L1-Cache Hit Rate sinkt – aber nur, wenn der Funktionsaufruf vor den Speicherallokationen stattfindet. Eine genauere Betrachtung der Adressen die *malloc* für die etwa 50 Allokationen zurückgibt, liefert den nächsten Hinweis: Der *cuCtxCreate*- Funktionsaufruf ändert, ohne klar erkennbaren Grund, das Standardverhalten von *malloc*, welches normalerweise jede Allokation mit jeweils gleichem Abstand zum Beginn einer 4kB-Speicherseite platziert.

Das führt leider auch dazu, dass Elemente mit der gleichen Position innerhalb eines Buffers auch in das gleiche Cache Set des L1-Caches fallen. Wird auf genügend viele Arrays direkt nacheinander zugegriffen, ist die achtfache Assoziativität moderner x86-CPU's schnell erschöpft, und die Conflict Misses reduzieren die L1-Cache Hit Rate. Ein selbst geschriebener Speicherallokator konnte dieses unglückliche Verhalten dann, sorry Jensen, auch ganz ohne teure GPU beheben.

Dominik Ernst

Verarbeitung natürlicher Sprache (Natural Language Processing, NLP), die medizinische Bildverarbeitung mit innovativen KI-Algorithmen zur Analyse medizinischer Bilder sowie die numerische Strömungsmechanik (Computational Fluid Dynamics, CFD) zur Verbesserung des Verständnisses in Bereichen wie Luft- und Raumfahrttechnik sowie Fertigung.

KI-System der nächsten Generation mit skalierbarer und beschleunigter Rechenleistung

Das neue KI-System am LRZ wurde speziell für die Verarbeitung großer Datenmengen und zur Bewältigung komplexer wissenschaftlicher Fragestellungen entwickelt. Es besteht aus dem HPE Superdome Flex Server und dem Cerebras CS-2-System, das in dieser Kombination in Europa zum ersten Mal verfügbar gemacht wird. Der HPE Superdome Flex Server bietet eine modulare, skalierbare Lösung, um den Rechenanforderungen gerecht zu werden, und verfügt über spezielle Funktionen für große In-Memory-Verarbeitungen, die für die Auswertung riesiger Datenmengen erforderlich sind. Darüber hinaus sind die spezifischen Vor- und Nachbearbeitungsfunktionen des HPE Superdome Flex Servers für das Training und die Inferenz von KI-Modellen ideal. Es unterstützt so das Cerebras CS-2-System, das die Deep-Learning-Leistung von Hunderten von Grafikprozessoren (GPUs) mit der einfachen Programmierung eines einzelnen Knotens bietet.

Der zurzeit größte Chip mit 850.000 Rechenkernen

Herzstück des Cerebras-Systems ist daher die Wafer Scale Engine der zweiten Generation (WSE-2), ein matt golden schimmernder Prozessor, der so groß ist wie ein Essteller. Auf den etwa 46.000 Quadratmillimetern Silizium sammeln sich 2,6 Billionen Transistoren und 850.000 Rechenkern, die speziell für Verfahren von KI entwickelt wurden, außerdem gleichmäßig verteilte Speicher, die bis zu 40 Gigabyte Daten fassen, sowie schnelle Verbindungen für deren Transport über die Platte mit 220 Petabyte pro Sekunde. So kann WSE-2 während der Ausführung alle Parameter von vielschichtigen neuronalen Netzen auf einem Chip halten, was wiederum die Rechenzeit und Datenverarbeitung verkürzt. Bisher kommt das Cerebras CS-2-System in einigen US-amerikanischen Forschungseinrichtungen zum Einsatz. „Wir

¹Erstellt einen CUDA Kontext, ist aber normal gar nicht nötig

haben Cerebras gegründet, um die Zukunft des Computing zu revolutionieren“, sagt Andrew Feldman, CEO und Mitbegründer von Cerebras Systems. „Wir sind stolz darauf, mit dem LRZ und HPE zusammenzuarbeiten, um der bayerischen Forschungscommunity Zugang zu KI der Spitzenklasse zu verschaffen. Diese wird es ihnen ermöglichen, neue Hypothesen zu testen, große Sprachmodelle zu trainieren und letztlich die wissenschaftliche Forschung voranzutreiben.“

Supportsystem und Software für die Entwicklung von KI-Methoden

Der HPE Superdome Flex Server komplettiert die CS-2 und bietet zusätzlich große Speicherkapazitäten und eine enorme Skalierbarkeit der Rechenleistung. So ist das integrierte System ideal abgestimmt auf die riesigen, datenintensiven ML-Projekte, die zu den klassischen Anwendungsfällen des Cerebras Systems zählen. Außerdem plant und verwaltet der HPE Superdome Flex Server das System entsprechend der Anforderungen verschiedener Anwendungen und ermöglicht den Zugriff via Cloud. Darüber hinaus enthält der HPE Superdome Flex Server einen Softwarestack mit Programmen zur Erstellung von KI-Verfahren und -Modellen.

„Wir freuen uns, unsere Zusammenarbeit mit dem Leibniz-Rechenzentrum zu erweitern mit Computer-Technologie der nächsten Generation für die wissenschaftliche Gemeinschaft“, sagte Justin Hotard, Executive Vice President und General Manager, HPC & AI, HPE. „Zusammen mit dem Leibniz-Rechenzentrum und Cerebras unterstützen wir die nächste Welle wissenschaftlicher und technischer Innovationen in Deutschland. Mit der zunehmenden Verbreitung von KI und maschinellem Lernen werden hochoptimierte Systeme wie dasjenige des LRZ wissenschaftliche Durchbrüche zum Wohle der Menschheit beschleunigen.“

Zusätzlich zu Anwendungsfällen aus dem KI-Umfeld wird das Cerebras/HPE-System auch für traditionellere High-Performance Computing (HPC) Use Cases in Betracht gezogen, wenn die spezielle Rechenarchitektur den Eigenschaften der Anwendung entgegenkommt. „Die Zukunft des Computings wird komplexer, die Systeme werden heterogener und auf spezifische Anwendungen abgestimmt. Wir sollten aufhören, in HPC- oder KI-Systemen zu

Natürlich sollen „die beiden Franken“ nicht nur LINPACK rechnen und in den Listen gut aussehen. Sie stehen in erster Linie den Forschenden zur Verfügung – und das nicht nur in Franken. Das NHR@FAU nimmt seit Q2/2022 auch Anträge auf Rechenzeit für diese Systeme aus ganz Deutschland und natürlich auch aus den nicht-fränkischen Teilen Bayerns entgegen. Informationen zum Antragsprozess finden sich unter <http://tiny.cc/NHRFAU-Application>.

Gerne unterstützen NHR@FAU und KONWIHR Forschende in Erlangen bei der Portierung und Optimierung ihrer Codes für Fritz und Alex!

Die Kosten von Fritz und Alex werden von der FAU, der DFG und dem NHR weitgehend gemeinsam getragen. Für Alex hat die Hochschule Coburg 40 A100-GPUs beigesteuert und der Freistaat Bayern fördert diese Kooperation mit der Übernahme der anfallenden Betriebskosten. Hierzu vielleicht mehr im nächsten Quartl.

Gerhard Wellein

und verbessert, sodass Fritz und Alex es richtig gemütlich kalt bzw. warm haben. In den am 30.5.2022 veröffentlichten Rankings der weltweit schnellsten (TOP500) und energieeffizientesten (Green500) Rechner konnten sich Fritz und Alex getrennt platzieren:

- Mit einer LINPACK-Rechenleistung seiner 256 NVIDIA-A100-Karten (NVIDIA-A40-Karten können leider nur MD-Anwendungen aber kein LINPACK) von 2,9 PF/s wird Alex auf Platz 184 der TOP500 gelistet. Sehr erfreulich ist auch der 16. Platz in der GREEN500 der energieeffizientesten Rechner – hier nimmt Alex immerhin den Spitzenplatz in Deutschland ein.
- Fritz dagegen konnte für seinen LINPACK-Lauf nur 612 Knoten (Stichwort: tröpfchenweise eintrudelnde Mellanox-IB-Karten) verwenden und erzielte damit 2,3 PF/s LINPACK-Leistung was zu Platz 324 reichte. Da es zwischenzeitlich auch die letzten IB-Karten gerechnet hat, wird sich Fritz auch im November – dann mit deutlich mehr als 3 PF/s – gut in der TOP500 platzieren können.



Abbildung 3: Rechenknoten von „Alex“ (li.) und „Fritz“

denken“, sagt Laura Schulz, Head of Strategy am LRZ. „Auf CPU-basierten Systemen wie dem SuperMUC-NG funktionieren KI-Verfahren, und umgekehrt können High Performance Computing-Algorithmen große Leistungssteigerungen auf Systemen wie Cerebras erzielen. Wir arbeiten kontinuierlich daran, dass sich unsere Nutzerinnen und Nutzer auf ihre For-schung konzentrieren können – ganz egal, ob die zugrundeliegende Infra-struktur auf HPC-, KI- oder Quantentechnologie basiert.“

Sabrina Schulte

Partnerschaftlich in die Zukunft



Leibniz-Rechenzentrum
der Bayerischen Akademie der Wissenschaften

Er soll modernste Technik enthalten, Komponenten, die (noch) nicht verkauft werden und die der Forschung neue Möglichkeiten offerieren: Bei der Beschaffung seines nächsten Supercomputers setzt das Leibniz-Rechenzentrum (LRZ) auf ein neues Verfahren, die Innovationspartnerschaft. Die Architektur des neuen Höchstleistungsrechners am LRZ wird bis 2024 gemeinsam mit Technologie-Unternehmen entwickelt, dazu werden Prototypen gebaut oder bestehende Komponenten an spezifische Bedürfnisse angepasst. „Schon seit den 2010er Jahren ist klar, dass die nächste Generation von HPC-Systemen, die Exascale-Rechner, im Co-Design entstehen werden, das heißt, Hard- und Software werden in Abstimmung für den gewünschten Einsatzzweck optimiert. Im wissenschaftlichen Umfeld bauen bereits viele National Labs der USA auf die gemeinsame Entwicklung von Hard- und Software“, erklärt Prof. Dr. Dieter Kranzlmüller, Leiter des LRZ. „Wir wollten die Beschaffung unserer Systeme verbessern, die Innovationspartnerschaft bot sich dazu an.“ Auch die Geldgeber, das Bundesministerium für Bildung und Forschung (BMBF) sowie das bayerische Staatsministerium für Wissenschaft und Kunst, gaben grünes Licht. Und so kann der nächste nationale Supercomputer für Garching als erster in Europa im Rahmen einer Innovationspartnerschaft geplant werden. „Passend zu unseren Bedürfnissen“, erwartet Kranzlmüller, „können wir gemeinsam mit Herstellern innovative Computing-Ansätze erforschen, evaluieren und schließlich im neuen Exascalesystem umsetzen.“

Werkzeug zur Modernisierung und Wirtschaftsförderung zugleich

In Deutschland sind Innovationspartnerschaften seit 2016 möglich. Bund und Länder unterstützen diese Beschaffungsmethode, weil sie die Entwicklung von neuen Angeboten und innovative Unternehmen unterstützt, laut Bundeswirtschaftsministerium sogar die Verwaltung und öffentliche Institutionen modernisieren hilft. „Die öffentliche Hand gibt pro Jahr in Deutschland rund 500 Milliarden Euro aus“, stellt Matthias Berg, Leiter des Kompetenzz-

die beliebten Bereiche Künstliche Intelligenz (AI) und Maschinelles Lernen (ML) 256 NVIDIA A100 GPUs bereitstellt. Die 304 NVIDIA-A40-GPUs richten ihren Blick hingegen auf die ökonomische und schnelle Abarbeitung von Molekulardynamik-Workloads.

- Fritz dagegen vereint die Rechenleistung von fast 70.000 Intel-CPU-Kernen in 944 Knoten und verfügt (bzw. verfügte irgendwann) über ein schnelles Infiniband-Netzwerk. Mit seinen 72 Kernen und 256 GB Speicher pro Knoten ist er das Arbeitspferd für parallele Anwendungen aus allen Wissensbereichen.

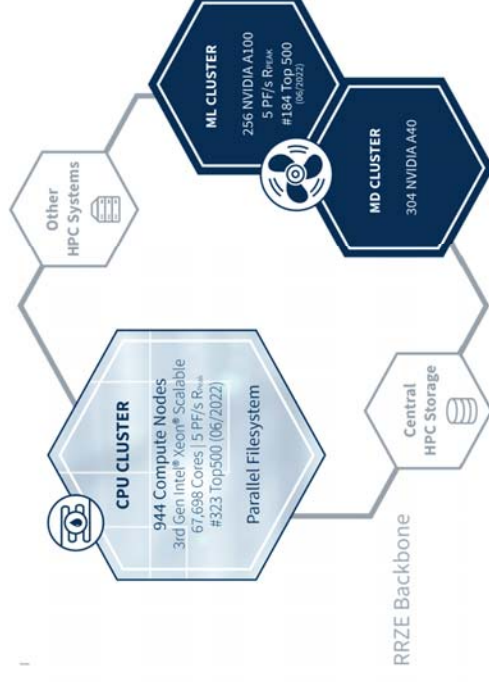


Abbildung 2: Parallel-GPU-Cluster-Megware

Die Lieferung und Installation der Systeme startete im zweiten Halbjahr 2021 und zog sich bis in den Juni 2022 hinein – Stichwort: homöopathische Liefermengen bei Infiniband-Karten der Firma Mellanox/NVIDIA. Parallel dazu wurde 2021 die Kühlinfrastruktur in rekordverdächtiger Zeit erweitert

Zwei flotte Franken unter den 500 Besten



Die Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) hat ihre High-Performance-Computing-(HPC)-Ressourcen signifikant ausgebaut. Gleich mehrfach platzieren sich die neuen Supercomputer am Zentrum für Nationales Hochleistungsrechnen Erlangen (NHR@FAU) in den Rankings der weltweit schnellsten (TOP500) und energieeffizientesten (Green500) Rechner.

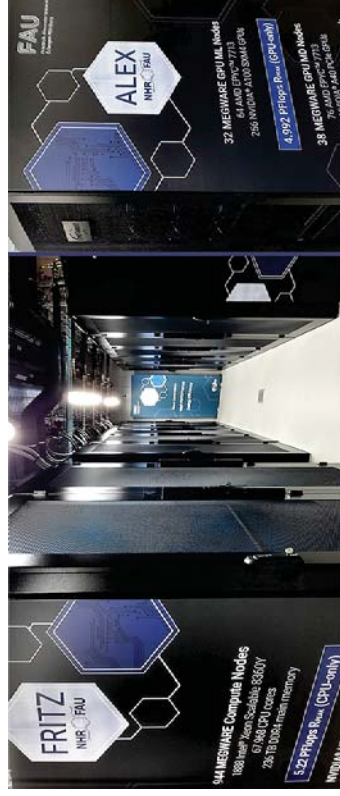


Abbildung 1: Parallel-Cluster „Fritz“ und GPU-Cluster „Alex“

Bereits im Laufe des Jahres 2020 wurde mit den Planungen für Beschaffung und Installation des ersten NHR-Systems in Erlangen begonnen. Nach Aufnahme der FAU in den NHR-Verbund zum 1.1.2021 gingen dann Ausschreibung und Auswahl des neuen Rechners sehr schnell. Im Mai 2021 wurde mit MEGAWATT ein Vertrag zur Lieferung nicht eines monolithischen Systems, sondern zweier Module mit komplementärer Rechnerarchitektur abgeschlossen:

- Alex, ein System, das mit fast 600 leistungsfähigen Grafikkarten der Firma NVIDIA und großem lokalen Speicher vollgepackt ist und für

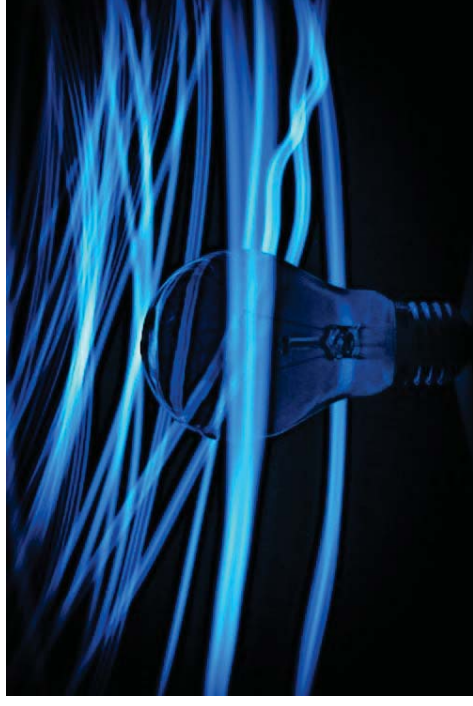


Abbildung 1: Innovation entstehen auch im Wettbewerb – die wichtigste Grundidee bei der neuen Beschaffungsmethode Innovationspartnerschaft. Foto: Nejk Soklic/Unsplash

trums für innovative Beschaffung, KOINNO, fest. Er begleitete als Berater bereits einige solcher Partnerschaften. „Wenn dabei nur ein Prozent in neue Produkte und Dienstleistungen fließt, dann ist das eine enorme Wirtschaftsförderung.“ Bei der Entwicklung des Exascale-Computers geht es zwar weniger um Modernisierung und Wirtschaftsförderung als vielmehr um die Entwicklung zukunftsreicher Computertechnologien. Wie andere Innovationspartnerschaften wird auch die des LRZ in mehreren Phasen verlaufen. Etwa zehn Spezialist:innen aus unterschiedlichen Bereichen begleiten die Beschaffung, die bereits Anfang 2022 mit einem Teilnahmewettbewerb begann. In mehreren Runden und bis Mitte April konkurrierten Hardware-Unternehmen nach den Vorgaben des LRZ mit Konzepten für das nächste High-Performance Computing- oder HPC-System. Aktuell sind die Bieterfirmen aufgefordert, dem LRZ offizielle Angebote mit detailliert ausgearbeiteten Co-Design-Schritten zu übermitteln. Im Juli soll das Vertragswerk stehen

und werden die praktischen Entwicklungsarbeiten am neuen System mit den Unternehmen starten. Bis 2024 hat sich dann herausgestellt, welches Unternehmen den nächsten Supercomputer der Exascale-Generation aufbauen und betreiben wird.

„Die Innovationspartnerschaft ermöglicht es uns, mit mehreren Anbietern gleichzeitig Technologien für das nächste System zu entwickeln und zu optimieren“, nennt Herbert Huber Vorteile. Der promovierte Physiker leitet seit zehn Jahren die Abteilung Hochleistungssysteme am LRZ und hat bereits mehrere Supercomputer mitgeplant und beschafft. „So können wir die Unternehmen und ihre Arbeitsweise schon weit vor der Lieferung des Höchstleistungsrechners kennenlernen.“ In die Innovationspartnerschaft des LRZ fließen die Erfahrungen aus dem Future Computing und mit der Testumgebung Bavarian Energy, Architecture and Software Testbed (BEAST) ein. Auch das Big Data & Artificial Intelligence Team (BDAI) sowie die Gruppe Computational X Support (CXS), die Forschende bei der Optimierung von Algorithmen und deren Implementierung in HPC-Systemen unterstützt, arbeiten an Vorschlägen und Spezifikationen mit. Die Hoffnungen sind hoch, dass das Zusammenspiel der Komponenten des Supercomputers schon im Vorfeld getestet und verbessert werden kann. Außerdem sollen Anlaufschwierigkeiten abnehmen, wenn das System in Betrieb genommen wird. „Sind die ersten technischen Schritte und Verträge geklärt“, plant Huber, „werden wir daher auch Anwender:innen in die Entwicklungsarbeiten einbeziehen.“

Aufwand, Diskussionen und Risiken

Innovationspartnerschaften sind aufwändiger als andere Beschaffungsverfahren: „Sie lohnen bei Produkten mit hohem Investitionsbedarf und bei Neuentwicklungen“, beobachtet Berg. „Allerdings gibt es meist nur eine vage Idee von der zu entwickelnden Technik oder Dienstleistung, die Verträge werden über Forschungs- und Entwicklungsarbeiten sowie Prototypenbau geschlossen.“ Diese berühren Urheber-, Patent- und auch Vermarktungsrechte und werden daher individueller und detaillierter ausgefertigt. Außerdem sind viele Vorsichtsmaßnahmen zu treffen. Das LRZ muss unter anderem sicherstellen, dass alle Unternehmen mit den gleichen Informationen arbeiten können und keiner von den technischen Neuerungen der anderen erfährt.

Die Prototypen von Prozessoren, Acceleratoren und anderen Komponenten werden daher in unterschiedlichen und abgeschlossenen Racks installiert. Und das Team ist zu Schweigen verpflichtet, wer was und wie zum nächsten Supercomputer beisteuert, darf aktuell nicht an die Öffentlichkeit kommen. Innovationskennner Berg empfiehlt außerdem, den Zeitaufwand nicht zu unterschätzen. „Für die einzelnen Entwicklungsschritte sollten die Konstruktionsvorgaben und vor allem die Prüfkriterien sorgfältig definiert und ausgearbeitet werden“, sagt er. „An den Diskussionen sollten möglichst viele beteiligt sein, damit unterschiedliche Sichtweisen und Bedürfnisse berücksichtigt werden.“ Auch wenn gleich mehrere Unternehmen in den ersten Runden an einem Problem arbeiten und dafür honoriert werden, helfen Innovationspartnerschaften durchaus auch beim Sparen. Erfahrungen zufolge sinken im IT-Bereich die Kosten um bis zu 30 Prozent, weil Hardware oder Software gezielter auf reale Bedarfe entwickelt werden. Vor allem aber soll im Fall des neuen Garching Supercomputers durch technologischen Wettbewerb das bestmögliche Exascale-System für Forschung und Wissenschaft entstehen – und werden hoffentlich nebenbei auch neue Komponenten entwickelt, die in der Wirtschaft IT-Prozesse besser gestalten helfen. Teilnehmende Unternehmen gehen bei der Innovationspartnerschaft zwar das Risiko ein, dass ihre Lösungen letztlich nicht ausgewählt und bestellt werden. Dafür können sie aber die während der Kooperation entstandenen Produkte selbst vermarkten und auf weitere Einsatzmöglichkeiten hin anpassen. „Gut möglich, dass Vorstellungen scheitern, aber diese Erfahrungen sind für alle Beteiligten wertvoll. Scheitern gehört zu Innovationen, der Öffentliche Dienst tut sich damit schwer, weil er über seine Ausgaben Rechenschaft ablegen muss“, sagt Berg zuversichtlich: „Obwohl das Verfahren komplex und die Risiken hoch sind, habe ich bisher noch keine Innovationspartnerschaft scheitern sehen.“

Sabrina Schulte