



INNOVATION, DISRUPTION UND GANZHEITLICHES DENKEN IN DER WELT DES VERTEILTEN RECHNENS

Die Abteilung hat mit BeeGFS, PreStack-Pro, dem Global Address Space Programmiermodell (GPI) sowie dem Big Data Framework GPI-Space innovative, weltweit anerkannte Technologien zur Lösung von Large-Data-Problemen entwickelt. In den vergangenen Jahren haben wir diese Technologiebasis sehr erfolgreich mit Deep-Learning-Methoden kombiniert und internationale Sichtbarkeit gewonnen. Im Kern geht es dabei immer um die skalierbare automatische Parallelisierung von Daten- und Compute-Problemen. Asynchrone Kommunikation und »Memory Driven Computing« sind die Technologien, mit denen es gelingt, Skalierbarkeit und Performance zusammenzubringen. Besondere Erfolge konnte die Abteilung im Machine Learning vorweisen, vor allem in den Projekten Deep Fake Detection und DeTol.

Wir engagieren uns in der EU-geförderten HPC-Forschung mit dem Ziel, europäische Technologien zu stärken und die Marktfähigkeit europäischer HPC-Softwareprodukte zu verbessern. Darüber hinaus ist es unser Anliegen, in Co-Design-Projekten die Mikroelektronikentwicklung und die Anwendungsentwicklung zusammenzubringen. Wir sehen in der anwendungsspezifischen Entwicklung von Compute-Hardware einen Weg, Europas Position auf dem stark wachsenden HPC- und Big-Data-Markt zu verbessern.

Das Energiesystem der Zukunft wird aus Millionen von verteilten Internet-of-Things-Rechnern bestehen. Diese optimieren den Eigenverbrauch von PV-Strom, regeln den Aufbau von Community Grids, steuern große und kleine Stromspeichersysteme und koordinieren den Energiefluss in unseren Energienetzen. In unseren Projekten entwickeln wir Technologien und Lösungen, um diese verteilte Rechnerwelt zu beherrschen. Dabei gilt unser Engagement intelligenten Lösungen, die die Energiewende voranbringen.

SCHWERPUNKTE

- Skalierbare parallele Programmierung
 - Deep Learning Tools und Anwendungen
 - Automatische Parallelisierung (ALOMA, GPI-Space)
 - BeeGFS – Parallel Cluster File System
 - Green by IT
 - Hardware Software Co-Design
-

Kontakt

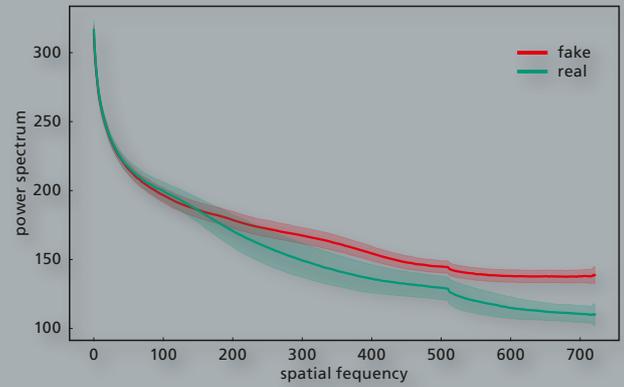
franz-josef.pfreundt@itwm.fraunhofer.de

www.itwm.fraunhofer.de/hpc





1



2

©Karras, Tero, Samuli Laine, Timo Aila

DEEP FAKE DETECTION: MATHEMATIK ENTLARVT BILDMANIPULATIONEN

1 *Synthetische Gesichter; erzeugt wurde sie mit dem Ansatz von Karras, Tero, Samuli Laine, and Timo Aila, vorgestellt in »A Style-Based Generator Architecture for Generative Adversarial Networks«*

2 *ITWM Deep Fake Detektor: nach der mathematischen Abbildung lassen sich echte und gefälschte Bilder leicht unterscheiden.*

In den letzten Jahren hat die KI-Forschung beeindruckende Fortschritte erzielt. Insbesondere im Bereich generativer Modelle, die nicht nur Daten auswerten, sondern auch realistisch anmutende synthetische Daten erzeugen können. Diese bergen leider auch das Risiko des Missbrauchs. Wie diese Deep Fakes genannten visuellen Fälschungen aufgedeckt werden können, zeigt der Deep Fake Detektor, den wir mitentwickelt haben.

Mit den generativen Modellen, auch »Generative Adversarial Networks« (GANs) genannt, wurde in der KI-Forschung ein Durchbruch erreicht, der auf Deep-Learning-Algorithmen beruht. Leider hat die GAN-Technologie auch ihre negativen Seiten: Schon kurze Zeit nach ihrer Einführung wurden GANs verwendet, um Bild- und Audiodaten zu manipulieren. Im Internet kursieren gefälschte, aber täuschend echte Bilder und Videos von Prominenten und Politikerinnen und Politikern. Diese Manipulationen wurden unter dem Begriff »Deep Fake« bekannt. Unsere Abbildungen zeigen einige Beispiele für synthetische Gesichter, die für menschliche Betrachter kaum von echten Gesichtern zu unterscheiden sind.

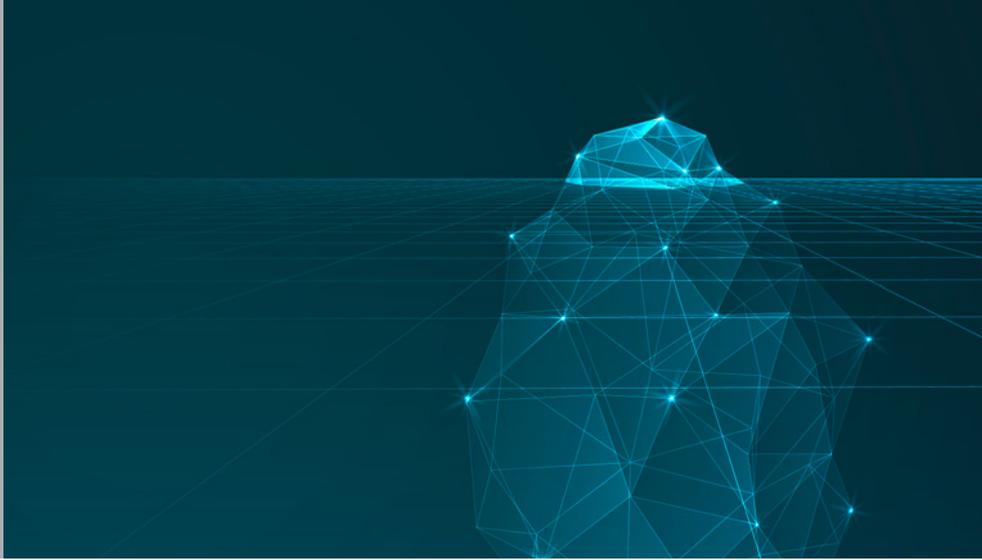
Gemeinsame Forschung gegen den Vertrauensverlust

Aufgrund der zu erwartenden gesellschaftlichen Auswirkungen, welche mit einem vollständigen Vertrauensverlust in potenziell manipulierte Bild- und Audiodaten einhergehen würden, haben sich weltweit Forscherteams an die Entwicklung von Algorithmen gemacht, welche »Deep Fakes« automatisch erkennen können. Während die meisten derzeitigen Ansätze versuchen, wieder Lernalgorithmen einzusetzen, um die Manipulationen zu erkennen, sind wir mit einer Gruppe von Wissenschaftlerinnen und Wissenschaftlern der Universität Mannheim sowie der Hochschule Offenburg einen anderen Weg gegangen: Bei Analysen fiel auf, dass GANs beim Generieren von Bildern inhärente Fehler machen. Diese sind zwar für das menschliche Auge kaum sichtbar, lassen sich mathematisch aber sehr leicht im Fourier-Raum abbilden.

Der Deep Fake Detektor ermittelt

Die neue, am ITWM entwickelte Methode hat zu bestehenden Verfahren gleich mehrere Vorteile:

- Der entdeckte GAN-Fehler ist systematisch bedingt; daher ist es theoretisch unmöglich, dass derzeitige GAN-Architekturen lernen können, den Detektor zu umgehen.
- Es werden nur sehr wenige Beispieldaten benötigt, um Deep Fakes zuverlässig zu erkennen.
- Die Methode ist einfach zu implementieren und benötigt vergleichsweise wenig Rechenleistung. Bei einer ersten Auswertung auf öffentlichen Testdaten erreichte der neue Ansatz eine Genauigkeit von nahezu einhundert Prozent.



DEEP TOPOLOGY LEARNING: AUTOMATISCHES DESIGN VON TIEFEN NEURONALEN NETZEN

Die Entwicklung der letzten Jahre zeigt, dass Maschinelles Lernen und insbesondere der Teilbereich Deep Learning sich zukünftig als maßgeblicher Baustein sowohl im wissenschaftlichen als auch im industriellen Bereich darstellen. Im BMBF-Projekt »Deep Topology Learning« arbeiten wir gemeinsam mit anderen Fraunhofer-Instituten an der Beschleunigung von Entwurfsalgorithmen.

Von der Spracherkennung über die automatische Bildanalyse bis hin zu Prototypen autonom fahrender Autos oder »Go«-spielender Algorithmen auf Weltmeister-Niveau: Fast immer stehen hinter den Erfolgsmeldungen sogenannte Deep-Learning-Verfahren. Diese Familie von Lernverfahren verwendet typischerweise überparametrisierte und meist sehr große künstliche neuronale Netze (DNN) zur Modellierung der Lernprobleme. Das Training solcher Netze bedarf nicht nur sehr großer Datenmengen, sondern auch enormer Rechenleistung. Trotz der teilweise beeindruckenden Ergebnisse, die mit DNNs erreicht werden, haben diese noch einige Nachteile, die aktuell oft noch den breiten Einsatz in der Praxis behindern. Neben den typischerweise benötigten sehr großen Datenmengen, ist dies vor allem der aufwändige Entwicklungsprozess.

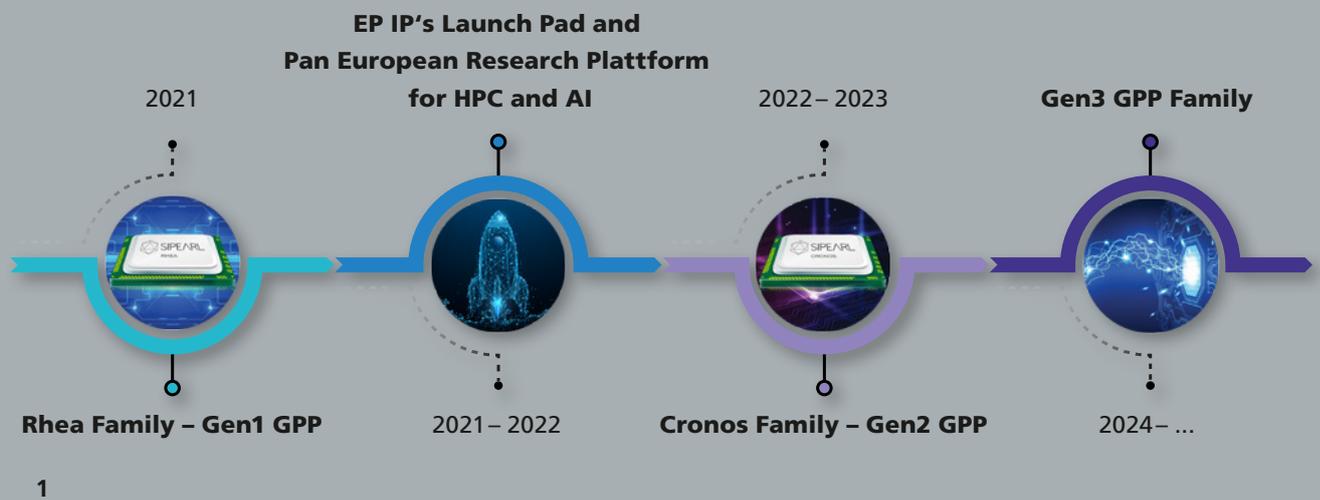
Entwurfsalgorithmen ersetzen Trail and Error

Der Entwurf neuer, problemspezifischer Netztopologien ist ein sehr zeit- und rechenaufwändiger Prozess. Bisher erfolgte die Entwicklung neuer Deep-Learning-Lösungen in einem heuristischen und erfahrungsgesteuerten »Trial-and-Error«-Ansatz. Ziel des BMBF-Projekts »Deep Topology Learning« (DeToL) ist es, dieses Entwurfsverfahren für Deep-Learning-Lösungen durch automatische, datengetriebene Entwurfsalgorithmen entscheidend zu beschleunigen und zu vereinfachen.

Nächster Schritt: Automatisierung

Das Anwendungsgebiet von Deep Learning ist breit gefächert. Vom maschinellen Sehen über autonomes Fahren bis hin zu Spracherkennung, Musikgenerierung oder Kunst – Deep Learning trägt immer mehr wesentliche Anteile zur Entwicklung bei. Gleichzeitig ist die Menge an Expertinnen und Experten, die tiefe neuronale Netze für ein bestimmtes Anwendungsgebiet designen könnten, begrenzt. Ein logischer nächster Schritt ist, den Automatisierungsgrad bei der Entwicklung von Netzarchitekturen zu maximieren. Da diese Automatisierung sehr rechenaufwändig ist, kommen an dieser Stelle HPC-Systeme ins Spiel. Die menschliche Interaktion wird auf das Design eines Suchraumes aller für ein bestimmtes Problem möglichen Topologien begrenzt. Anhand einer gegebenen Suchstrategie wird dann im nächsten Schritt die Architektur optimiert.





EUROPÄISCHE PROZESSOR INITIATIVE EPI

1 Die EPI Roadmap

High Performance Computing gehört in vielen Bereichen zu den Schlüsseltechnologien der zukünftigen Wertschöpfung. So wird es zum Beispiel für Simulationen von Wetter und Klima sowie neuen Materialien verwendet oder auch beim Einsatz von Künstlicher Intelligenz. Entwicklungen aus dem High Performance Computing finden sich wieder in autonomen Fahrzeugen, im Kontext von Industrie 4.0 und im Cloud Computing.

In all diesen Bereichen werden leistungsfähige Prozessoren gebraucht, um die enormen Datenmengen verarbeiten zu können. Es gibt nur wenige Hersteller, die solche erfolgreich produzieren. Der internationale Markt ist beherrscht von Firmen aus den USA und neuerdings auch aus China.

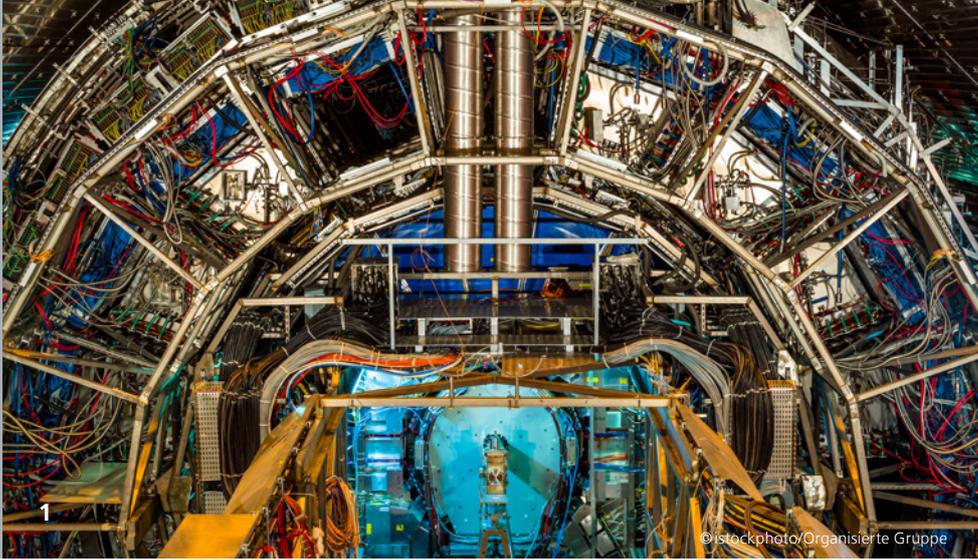
Europäische Unabhängigkeit stärken

Um die Unabhängigkeit von Europa in dieser Schlüsseltechnologie zu gewährleisten, vereinigt die Europäische Prozessor Initiative (EPI) 26 Partner aus Industrie und Wissenschaft, um europäische Prozessoren zu entwerfen und ein Hardware- und Software-Ökosystem in Europa aufzubauen. Ziel ist es, einen General-Purpose-Prozessor sowie Beschleunigerprozessoren zu entwickeln, die für bestimmte Anwendungsgebiete optimiert wurden. In diesem Bereich engagieren wir uns in Zusammenarbeit mit dem Fraunhofer IIS und der ETH Zürich und entwerfen einen Beschleuniger für Stencil- und Tensor-Berechnungen (STX). Stencil-basierte Applikationen sind sehr häufig bei HPC-Simulationen anzutreffen, Tensor-Berechnungen sind die Grundlage des Deep Learning.

Hard- und Software zeitgleich entwickeln

Wir wollen eine höchst energie-effiziente Hardware bauen, welche sich für die Endnutzer komfortabel programmieren lässt und dabei nicht nur auf ein Anwendungsgebiet, z.B. Deep Learning, beschränkt ist. Um dies zu erreichen, verwenden wir das sogenannte Hardware Software Co-Design. Hierbei wird die Software parallel zur Hardware entwickelt und optimiert. Während dieses Prozesses können in das Hardware-Design Anforderungen der Software einfließen und die Software an die Hardware optimiert werden. Die Auswahl der Software besteht aus wichtigen mathematischen Rechenkernen wie zum Beispiel die Matrix-Matrix-Multiplikation, Faltungsfiler und verschiedenen 2D- und 3D-Stencils. Für die Entwicklung nutzen wir die offene RISC-V Instruction Set Architecture, sodass keine Lizenzkosten oder Abhängigkeiten entstehen. Der erste Chip soll Anfang 2021 produziert werden, um Energieeffizienz und Performanz zu testen. Des Weiteren optimieren wir unseren HPC Software-Stack mit der Kommunikationsbibliothek GPI und unserem parallelen Dateisystem BeeGFS auf die EPI Hardware und unterstützen die Entwicklung des General-Purpose Prozessors mit einem weiteren Co-Design Prozess.





ENORME BESCHLEUNIGUNG MIT GPI-SPACE

Die immer genauere Interpretation der durch Experimente am Large Hadron Collider (LHC) produzierten sehr großen Datenmengen erfordert den Einsatz neuer mathematischer Werkzeuge, die Gegenstand aktueller Forschung sind. Ein erfolgreicher Ansatz setzt auch auf symbolische Rechnungen zur analytischen Reduktion komplexer Strukturen. Dabei stößt die Mathematik in den innersten Bereich des wissenschaftlichen Rechnens vor, sowohl bei der schieren Größe der Aufgaben als auch bei Fragen der Nachvollziehbarkeit, der Reproduzierbarkeit und der Automatisierung.

1 Teilchendetektor ATLAS am Large Hadron Collider des europäischen Kernforschungszentrums CERN

Das in unserer Abteilung entwickelte Werkzeug »GPI-Space« ist für diese Anforderungen konzipiert. Es ermöglicht, Daten auf sehr großen Maschinen effizient zu verwalten und zu bearbeiten. Virtualisierung des Speichers, erprobte Scheduler, robuste Fehlerbehandlung, skalierbares Ressourcenmanagement und Petri-Netz basierte Abhängigkeitsgraphen sind in GPI-Space integriert.

Antworten auf offene Fragen

Gemeinsam mit der der AG Algebra, Geometrie und Computeralgebra des Fachbereichs Mathematik der TU Kaiserslautern wurde das dort entwickelte und weltweit führende Computeralgebra-System »Singular« an GPI-Space angebunden und das entstandene Konglomerat eingesetzt, um Antworten auf bisher unbeantwortete Fragen zu erhalten. Dafür hat die Gruppe der TU in mehreren Projekten die Struktur ihrer mathematischen Verfahren in Petri-Netzen dargestellt und diese Darstellung benutzt, um mit GPI-Space die konkreten Rechnungen auf große Maschinen zu skalieren. Beschleunigungen um Faktor 400 bei gleichzeitiger paralleler Effizienz von 80 Prozent sind Resultate, die bisher als unerreichbar galten.

Wir konnten aus dieser Zusammenarbeit neue Anforderungen an GPI-Space generieren und zum Teil bereits Lösungen in GPI-Space bereitstellen. Z. B. basieren viele Teilaufgaben auf Berechnungen von Gröbner-Basen, deren Laufzeit sich für konkrete Eingaben nur sehr schlecht vorhersagen lässt. Gleichzeitiges Beschreiten verschiedener Pfade zu möglichen Lösungen hilft, Lösungen schneller zu finden. Danach werden die Resultate einiger der bereits gestarteten Rechnungen nicht mehr benötigt; GPI-Space unterstützt in der neuesten Version das kontrollierte Abbrechen dieser Rechnungen.

Aufbauend auf den sehr positiven Erfahrungen haben wir gemeinsam mit der AG Algebra, Geometrie und Computeralgebra die nächsten Schritte zur tieferen Integration der beiden Systeme GPI-Space und Singular geplant und mit der Umsetzung begonnen.

GPI  SPACE





1



2

© Isabel Nabuurs Fotografie

ENERGIEGENOSSENSCHAFT BETREIBT EIGENES MICROGRID MIT AMPERIX

1 Peak-Shaving im Einsatz.
Durch den Einsatz der verteilten Speicher überschreitet der Netzbezug der Energiegemeinschaft (orange) die gegebene Grenze nicht.

2 Die schwimmende Wohnsiedlung Schoonschip in Amsterdam Nord

Im Norden von Amsterdam wird eine neue Form der Energieversorgung realisiert: Dreißig schwimmende Häuser sind nahezu energieautark und bilden gemeinsam das nachhaltigste schwimmende Wohnquartier Europas. Zum Einsatz kommt dort auch unser Energiemanagementsystem Amperix®.

Die Idee: 30 Häuser mit 47 Wohnungen bilden eine Energieeinheit, die ihren Strom mittels Solarenergie weitgehend selbst erzeugt, mit Wärmepumpen und Batterien auch selbst speichert und unter den Bewohnern gegenseitig zur Verfügung stellt. Die Häuser sind also untereinander vernetzt und benötigen nur einen einzigen Anschluss an das kommunale Stromnetz für das gesamte Wohnprojekt. Dieser Netzknoten dient der Absicherung von Schoonschip in Spitzenverbrauchszeiten, auch bei leeren Batteriespeichern und wenig Sonne. Der Netzanschlusspunkt ist mit einer maximalen Leistung von 175 kW verglichen mit der Anzahl an vollelektrischen Wohneinheiten relativ klein.

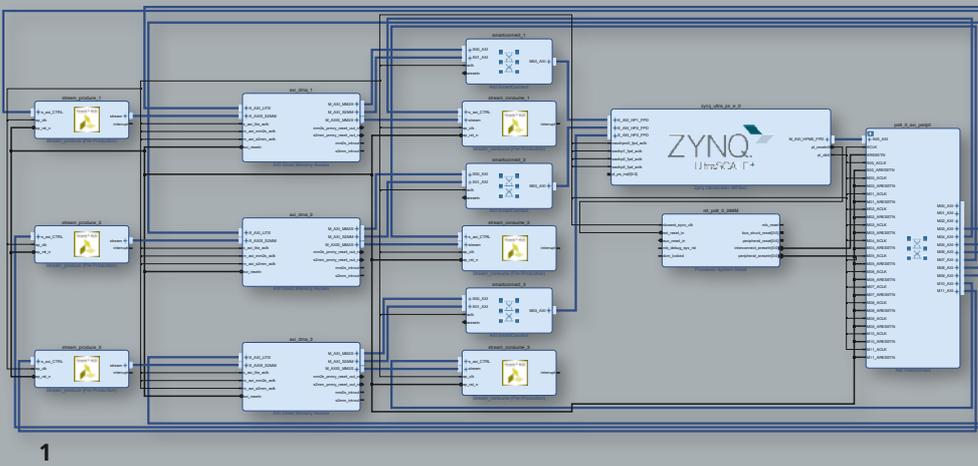
Amperix ermöglicht Sektorenkopplung

Hier kommt nun das am Competence Center High Performance Computing entwickelte Amperix® Energiemanagementsystem ins Spiel. Der Amperix® ist ein Werkzeug für Microgrids und Energiecommunitys. Neben der Steuerung der Stromspeicher implementiert der Amperix® hier auch eine Sektorenkopplung. Das heißt Wärmespeicher in Kombination mit Wärmepumpen und Ladestationen für Elektrofahrzeuge können flexibel entsprechend der Verfügbarkeit erneuerbarer Energien gesteuert werden. So wird die Lastverschiebung innerhalb der Wohngemeinschaft optimiert und Verbrauchsspitzen werden abgeflacht (sogenanntes Peak-Shaving).

Die gemeinschaftliche Steuerung der Energiegemeinschaft konnte sich im Winter 2019/2020 bereits bewähren. Gleichzeitig aufgetretene Lastspitzen der Wärmeversorgung und der Stromversorgung konnten mittels der koordinierten Ausspeisung die Batteriespeichersysteme reduziert und somit eine stabile Energieversorgung gewährleistet werden.

Weiterführung als H2020-Leuchtturmprojekt

Wegen des großen Erfolgs wird die Forschung und Entwicklung rund um das innovative Wohn- und Energiekonzept im ATELIER Smart City Projekt fortgeführt. Ziel des Projekts ist die Umsetzung von bürgerorientierten »Positive Energy Districts« in den Leuchtturm-Städten Amsterdam und Bilbao. Das ATELIER-Projekt wird von der Europäischen Kommission im Rahmen von Horizon 2020 finanziert.



FPGAs ALS BESCHLEUNIGER: DER ALGORITHMUS ALS DIGITALE SCHALTUNG

Field Programmable Gate Arrays (FPGAs) erleben zurzeit eine Renaissance als Beschleuniger-Hardware. Auf einem FPGA lassen sich komplexe Funktionen direkt als maßgeschneiderte Digitalschaltung realisieren. Damit verknüpfen FPGAs die Leistungsfähigkeit spezialisierter Hardware mit der Flexibilität von Software und werden immer häufiger als Beschleuniger in Hochleistungsrechner eingesetzt.

1 Blockdarstellung eines FPGA-Designs

Als eine Art Baukastensystem für digitale Schaltungen ermöglichen FPGAs, elementare elektronische Bausteine miteinander zur gewünschten Funktionalität zu kombinieren. Die Verschaltung der physischen Elemente auf dem FPGA wird als Konfiguration geladen und kann so nach Belieben in Sekundenschnelle geändert werden. War die Anwendungsentwicklung für FPGAs lange Zeit Hardware-Spezialisten vorbehalten, erleichtern heutige Werkzeuge einer breiteren Schicht von Software-Entwicklern den Einstieg.

Energieeffiziente Auswertung seismischer Daten

Im Rahmen des Projekts EuroEXA portieren wir eine Software zur Auswertung seismischer Daten auf FPGAs. Seismische Verfahren berechnen aus den auf der Oberfläche gemessenen Daten ein Abbild des Erduntergrunds, das es dem Geowissenschaftler erlaubt, Erdöl-/Erdgasvorkommen zu entdecken. Wir erwarten durch die Berechnung auf FPGAs eine Verbesserung der Performanz bzw. Energieeffizienz im Vergleich zu CPUs oder GPUs. Zur Portierung der rechenintensiven Programmteile stimmen wir Algorithmen und Datenstrukturen auf die spezifischen Möglichkeiten der Hardware ab. Die Performanz kann durch maximale Parallelität innerhalb eines FPGAs und durch die parallele Berechnung auf mehreren FPGAs optimiert werden. Parallelisierung über mehrere FPGAs setzt eine effiziente Datenkommunikation voraus, die mithilfe des von uns im Competence Center High Performance Computing entwickelten Programmiermodells GPI umgesetzt wird.

FPGAs ermöglichen spezifische Auslegung der Schaltkreise auf beliebige Zahlenformate

Eine weitere Anwendung von Interesse sind tiefe neuronale Netze (Deep Learning), wie sie im Maschinellen Lernen verwendet werden. Deep-Learning-Algorithmen sind besonders gut für Beschleuniger geeignet, da die Berechnungsmuster sehr homogen sind. Zudem zeigen sie sich robust gegenüber geringer Präzision in der Zahlendarstellung. Während die Hardware in CPUs und GPUs Berechnungen nur mit vordefinierten Zahlenformaten direkt ausführen kann, lassen sich die Schaltkreise auf dem FPGA spezifisch auf beliebige Zahlenformate auslegen. Üblicherweise werden FPGAs zur Ausführung bereits trainierter Netze genutzt, wir beschäftigen uns mit dem Training tiefer neuronaler Netze auf FPGAs.





IDEEN UND PRODUKTE FÜR DIE ENERGIEWENDE

Auf der E-world Anfang Februar zeigten unsere Forschenden aktuelle Technologien und Projekte zur Energiewirtschaft, auch das Energiemanagementsystem Amperix inclusive der dazugehörigen Web-Plattform myPowerGrid und der KI-basierte Solarvorhersage PVCAST: Der Amperix vernetzt Erzeuger und Speicher unterschiedlicher Hersteller und vollzieht die Sektorkopplung von Strom, E-Wärme und E-Mobilität vom privaten Haushalt bis hin zur Industrie. Energiedienstleistungen wie prognosebasierte Eigenverbrauchsoptimierung und Peak-Shaving des Netzbezugs (physikalisch und wirtschaftlich) können über drei Sektoren gleichzeitig betrieben werden.



BEST STUDENT PAPER AWARD

Ricard Durall Lopez hat den Best Student Paper Award der IVCNZ für die Präsentation »Semi Few-Shot Attribute Translation« gewonnen. In seinem Paper entwickelte er eine Deep Learning-Methode weiter. Der neuartige Ansatz funktioniert auf der Grundlage von GANs (Generative Adversarial Networks) und demonstriert empirisch das Potenzial des Trainings einer GAN für die Bild-zu-Bild-Übersetzung mit wenigen Aufnahmen.

EPI: SUPERCOMPUTER-INFRASTRUKTUR MADE IN EU

Die European Processor Initiative (EPI) ist ein Zusammenschluss von 27 Partnern aus zehn EU-Ländern, die gemeinsam Prozessortechnologie für einen Europäischen Supercomputer entwickelt. Innerhalb der nächsten Jahre soll so eine EU-eigene Supercomputer-Infrastruktur entstehen. Wir steuern zu EPI einen einfach zu programmierenden Co-Prozessor bei, der insbesondere Simulationscodes (Stencil-Algorithmen) beschleunigt. Erste Hardware wird 2021 zum Testen zur Verfügung stehen.

