



COMPETENCE CENTER HIGH PERFORMANCE COMPUTING



Der Lehrstuhl für Scientific Computing an der TU Kaiserslautern – Partner im Leistungszentrum »Simulations- und Software-basierte Innovation« – und das CC HPC arbeiten im Auftrag der MTU Aero Engines AG an der Verbesserung und Beschleunigung der Auslegungssoftware zur Optimierung von Triebwerken. Kern unseres Beitrages ist die Software GPI-Space, mit der sich »Memory Driven Computing« Architekturen realisieren lassen.



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

Die Abteilung hat mit BeeGFS, Pre-Stack 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 Big-Data-Problemen. Dahinter steht das Konzept des »Memory Driven Computing«, welches Skalierbarkeit und Performance zusammenbringt. Die im Folgenden beschriebenen Projekte ALOMA und Safeclouds entwickeln auf dieser Technologiebasis weitere branchenspezifische Lösungen.

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-/Big-Data-Markt zu verbessern.

Das Energiesystem der Zukunft wird aus Millionen von verteilten IoT-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.

Kontakt

franz-josef.pfreundt@itwm.fraunhofer.de

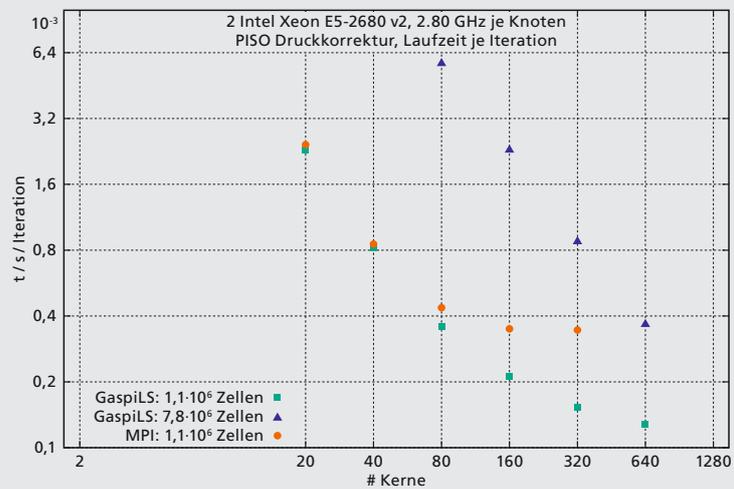
www.itwm.fraunhofer.de/hpc



SCHWERPUNKTE

- Green by IT
 - BeeGFS – Parallel Cluster File System
 - Visualisierung
 - Seismische Datenverarbeitung
 - Datenanalyse und Maschinelles Lernen
 - Skalierbare parallele Programmierung
-





1

GaspILS – SKALIERBARKEIT FÜR CFD- UND FEM-SIMULATIONEN

1 Berechnung der Druckkorrektur: Mit GaspILS (grün und blau) konnten Performanz und Skalierbarkeit entscheidend verbessert werden gegenüber der MPI-basierten Implementierung (orange).

Viele Simulationen aus dem Ingenieurbereich basieren auf CFD- und FEM-Methoden, zum Beispiel die Bestimmung aerodynamischer Eigenschaften von Flugzeugen oder die Analyse der Gebäudestatik. Ein Großteil der Rechenzeit wird dabei für die Lösung der zugrundeliegenden Gleichungen mittels iterativer Verfahren benötigt. Die Leistung der benutzten iterativen Löser hat also einen signifikanten Einfluss auf die Gesamtlaufzeit solcher Simulationen und damit auf den Erkenntnisgewinn, der mit ihnen generiert werden kann.

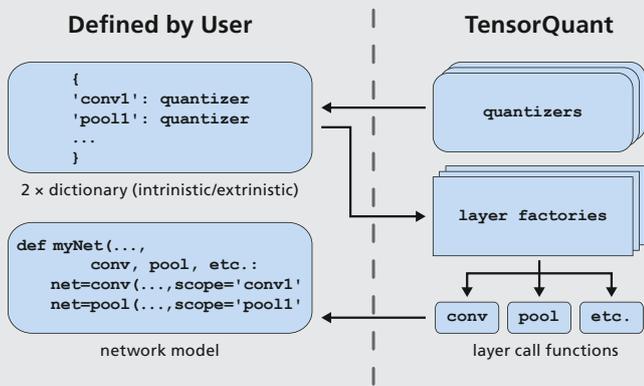
Um mehr Erkenntnisse aus den Simulationen gewinnen zu können, haben wir die lineare Löser-Bibliothek GaspILS entwickelt.

Industrie setzt auf GaspILS wegen besserer Skalierbarkeit

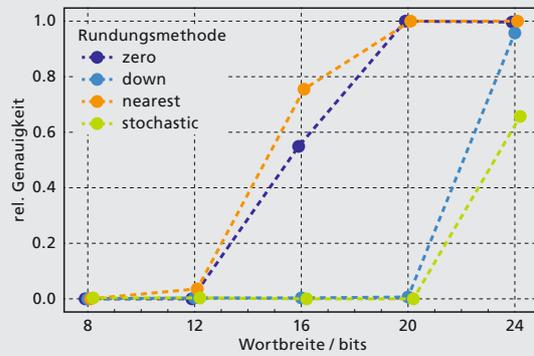
Skalierbarkeit ist ein Maß für die parallele Effizienz einer Implementierung. Das Optimum ist die sogenannte lineare Skalierbarkeit. Dies entspricht der vollständigen Ausnutzung von Rechenressourcen, wie sie die Kerne innerhalb einer CPU oder mehrerer über ein Netzwerk miteinander verbundener CPUs darstellen. Eine verbesserte Skalierbarkeit erlaubt es, mehr Rechenkapazität zur Lösung eines Problems zu nutzen und damit schneller zur Lösung zu kommen. Dies ermöglicht letztendlich detailliertere Modelle, genauere Parameter-Studien und eine kosteneffiziente Ausnutzung der Hardware-Ressourcen. Und dadurch wird GaspILS interessant für die Industrie.

Bessere Nutzung gegebener Rechenressourcen

Um gute Skalierbarkeit zu erreichen, benutzt GaspILS die von uns entwickelten Methoden und Werkzeuge zur parallelen Programmierung; dazu gehören die Kommunikationsbibliothek GPI-2 und deren zugrundeliegendes Programmiermodell. Der Algorithmus wird feingranular in Teilprobleme (sogenannte Tasks) mit gegenseitigen Abhängigkeiten aufgeteilt. Dies erlaubt zu jedem Zeitpunkt die Zuweisung ausführbarer Tasks auf freie Rechenressourcen und garantiert damit einen kontinuierlichen Strom an Rechenaufgaben für jede CPU. Der Verzicht auf globale Synchronisierungspunkte und die große Menge an erzeugten Teilproblemen erlauben es, die durch den Austausch von Daten entstehenden Latenzen zu verstecken und etwaige Imbalancen in der Rechenzeit auszugleichen. Jeder einzelne Kern kann zu jedem Zeitpunkt maximal ausgelastet werden.



1



2

TENSORQUANT BRINGT DEEP LEARNING IN DIE MOBILE ANWENDUNG

Der Trend, Methoden des Maschinellen Lernens, insbesondere tiefer künstlicher neuronaler Netze (Deep Learning – DL), zur Entwicklung intelligenter Systeme einzusetzen, hat sich weiter verstärkt. Aus der Wissenschaft kommen dabei stetig neue Deep-Learning-Ansätze, welche weit gefächerte Einsatzmöglichkeiten dieser Algorithmen in vielen verschiedenen praktischen Anwendungen ermöglichen könnten.

Eine entscheidende technologische Hürde auf dem Weg hin zur Anwendung in Massenprodukten ist jedoch der enorme Rechenaufwand, der zur Berechnung und Auswertung der DL-Modelle notwendig ist.

Daher ist in der letzten Zeit die Entwicklung spezialisierter DL-Hardware in den Fokus getreten. Neue Chip- und Speicherarchitekturen sollen zukünftig den Einsatz performanter und zugleich energiesparender Hardwaremodule ermöglichen und so den Einsatz von DL in z. B. autonomen Fahrzeugen, Mobiltelefonen oder integrierten Produktionssteuerungen ermöglichen.

Lernen erfordert keine hohe Präzision in der Zahldarstellung

Dabei nutzt man eine wesentliche mathematische Eigenschaft des DL aus: Das Lernen und Auswerten von Modellen lässt sich auf die numerische Berechnung einer recht kleinen Anzahl von Operationen aus der Tensor-Algebra (wie z. B. Matrixmultiplikationen) reduzieren. Außerdem kommt man bei der Tensor-Berechnung mit einer deutlich geringeren Genauigkeit in der Zahldarstellung aus, als man dies typischerweise von physikalischen Simulationen gewohnt ist. Diese Eigenschaften ermöglichen nun eine – im Vergleich zu allgemeinen Recheneinheiten wie CPUs und GPUs – hocheffiziente Hardwareumsetzung.

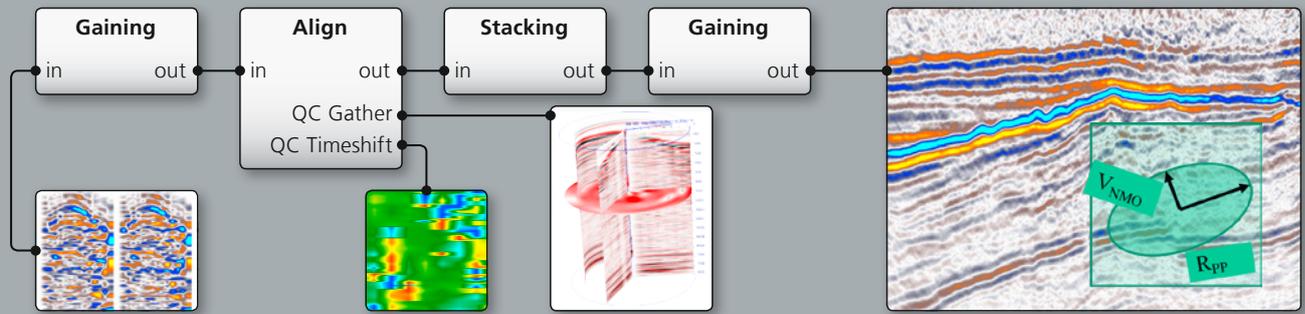
TensorQuant erlaubt die Simulation von Machine Learning Hardware

Bei der Entwicklung von DL-Anwendungen auf spezialisierter Hardware ergibt sich allerdings die Schwierigkeit, dass die Mindestanforderungen an die Rechengenauigkeit zwischen einzelnen Modellen stark variieren. Dies macht die gleichzeitige Optimierung von DL-Modellen und Hardware bezüglich Rechenperformanz, Energieverbrauch und Vorhersagegenauigkeit schwierig. Mit unserem Softwaretool TensorQuant (TQ) können Entwickler nun DL-Modelle mit beliebigen Zahldarstellungen und Rechengenauigkeiten simulieren, kritische Tensor-Operationen identifizieren und damit diesen Entwicklungsschritt deutlich beschleunigen. TQ wird bereits in Kooperationsprojekten mit der Automobilindustrie eingesetzt.

1 *TensorQuant erlaubt die automatische Simulation von gegebenen TensorFlow-Modellen mit beliebigen Zahldarstellungen einzelner Tensor-Operationen.*

2 *Das Ergebnis einer Simulation des bekannten ResNet-50 Models zeigt, dass die konkrete Wahl der Zahldarstellung erheblichen Einfluss auf die Performanz von DL-Anwendungen hat, welche ohne die Simulation in TensorQuant vorab nur schwer abzuschätzen ist.*





1

ALOMA ERLAUBT GEOPHYSIKERN FOKUSSIERUNG AUF IHRE KERNKOMPETENZ

1 Darstellung eines einfachen Workflows mit ALOMA: Eingangsgather werden korrigiert und danach gestackt. Eingabe und Ergebnisse können mit Fraunhofer XtreamView (Teil des ALOMA-Pakets) interaktiv visualisiert werden.

ALOMA befreit seine Nutzer davon, sich mit Parallelisierung, Multi-Threading und anderen Problemen des High Performance Computing beschäftigen zu müssen. Stattdessen können sich die Experten für seismische Fragestellungen auf ihre Kernkompetenz konzentrieren; um die effiziente Ausführung der Algorithmen, auch auf großen und heterogenen Systemen, kümmert sich dann ALOMA.

Das Programm ist eine spezialisierte Version von GPI-Space – einer Technologie, die auch in anderen Umgebungen wie Big Data und Machine Learning eingesetzt wird.

Komplexe Rechenoperationen auf wachsenden Datenmengen kennzeichnen Problemstellungen der Seismik. Dies zwingt Geophysiker dazu, sich HPC-Wissen anzueignen, damit ihre Software auch auf großen Systemen effizient läuft. Deshalb haben wir ein System entwickelt, das an der Schnittstelle zwischen Geophysiker und HPC-Experte sitzt. Informatiker und Geophysiker entwarfen gemeinsam Strategien für Parallelisierung, Datenaufteilung sowie Fehlertoleranz, speziell für seismische Anwendungen.

Den Kern von ALOMA, ein fehlertolerantes Laufzeitsystem zur Ausführung von Workflows auf verteilten Systemen, entwickelten HPC-Spezialisten. Für den Geowissenschaftler ist ALOMA damit lediglich eine Blackbox, in die er seine Entwicklungen über eine genau definierte Schnittstelle integrieren kann. Der Lernaufwand für das neue System ist dabei überschaubar. Ist ALOMA erst einmal installiert, können Anwender ihr erstes Modul binnen eines Tages in das neue System überführen.

Schnelle Prototypenentwicklung und -skalierung

Der Anwender kann damit nicht nur neue Algorithmen unter realen Nutzungsbedingungen testen und so rasch Prototypen für diese Verfahren entwickeln und skalieren. ALOMA erlaubt es auch, existierende Codes und Programme in verschiedenen Sprachen (z. B. C, C++, Matlab, Fortran) zu integrieren. Mit einem grafischen Editor können diese dann in Workflows zusammengefasst und automatisch parallelisiert werden.

Die Stärke dieses Konzepts konnten wir in mehreren Kooperationsprojekten mit Industriepartnern nachweisen, in denen es gelang, die Kundensoftware innerhalb weniger Tage skalierbar zu machen. Im Nachgang hat uns eine in Houston ansässige Firma aus der Öl- und Gas-Branche damit beauftragt, ihr Processing System komplett auf ALOMA umzustellen.



2

SAFECLOUDS – VERTEILTE INFRASTRUKTUR ZUR DATENANALYSE IN DER LUFTFAHRT

Der Luftverkehr nimmt weltweit zu. Das stetig wachsende Verkehrsaufkommen stellt Flugsicherheitsorganisationen, Flughäfen sowie Fluggesellschaften vor große Herausforderungen, denn sie müssen größtmögliche Sicherheit gewährleisten. Schon heute werden große Datenmengen aus verschiedenen Quellen wie Flugdatenschreibern und Radarstationen aufgezeichnet und ausgewertet.

Das EU-Projekt SafeClouds will existierende sowie neue Datenquellen in einer europaweiten Infrastruktur zusammenzuführen und anschließend mit Methoden des Maschinellen Lernens effizient auswerten; Ziel ist die deutliche Verbesserung des Luftverkehrsmanagements. Am Projekt beteiligt ist deshalb ein breites Konsortium, bestehend aus Fluggesellschaften, Flugsicherheitsorganisationen und -behörden sowie Forschungseinrichtungen.

Datenaustausch mit GPI-Space

Im CC High Performance Computing bauen wir eine mehrschichtige Hybrid-Cloud-Infrastruktur auf, die auf Amazon AWS basiert. Für die parallele Verarbeitung der Daten steht unsere Software GPI-Space zur Verfügung. Der Fokus liegt dabei auf der Daten- und Ausfallsicherheit sowie der leichten Skalierbarkeit im Hinblick auf Anzahl der Nutzer, Speicher und Rechenleistung.

Exemplarisch werden die Daten folgender Szenarien analysiert:

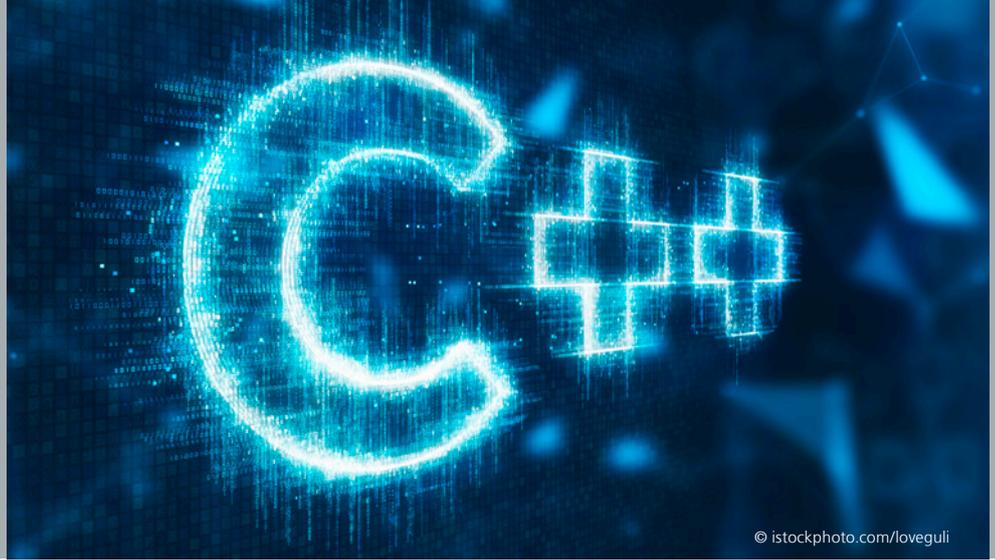
- Instabiler Anflug: Ein vordefinierter Bereich für verschiedene Parameter wie Höhe, Geschwindigkeit, Sinkgeschwindigkeit etc. wird nicht eingehalten und kann zu einer harten Landung, Landungsabbruch oder Ähnlichem führen.
- Geländebedingte Sicherheitswarnungen: Die vorgegebene Mindesthöhe wurde unterschritten aufgrund geographischer Gegebenheiten.
- AIRPROX (Aircraft Proximity Hazard): Die Sicherheit wurde durch die Unterschreitung des Mindestabstandes zwischen Flugzeugen gefährdet.

Auch die Landebahnen spielen eine Rolle: Ziel ist natürlich ihre optimale Auslastung; dabei müssen auch die Ausfahrten zu den Terminals unter Einhaltung der Mindestsicherheitsabstände berücksichtigt werden.

1 *Safe Clouds analysiert sicherheitsrelevante Vorgänge am Boden und in der Luft.*

2 *Methoden des Maschinellen Lernens erleichtern Luftverkehrsmanagement.*





MIT CONSTRAINT PROGRAMMING ZUR PASSENDEN VARIANTE

Immer komplexere Variantenkonfigurationen (VC) erfordern schnellere Algorithmen, um mit der Echtzeitfähigkeit von speicherbasierten Umgebungen wie SAP S/4HANA Schritt halten zu können. In einem Projekt für SAP nutzen wir moderne Constraint-Solving-Technologie in C++. Wir entwickeln wesentliche Teile der zugrundeliegenden Bibliothek, um den Anforderungen des SAP S/4HANA Variantenkonfigurators zu genügen und unterstützen SAP bei der Ausarbeitung von algorithmischen Lösungen und mit C++-Expertise.

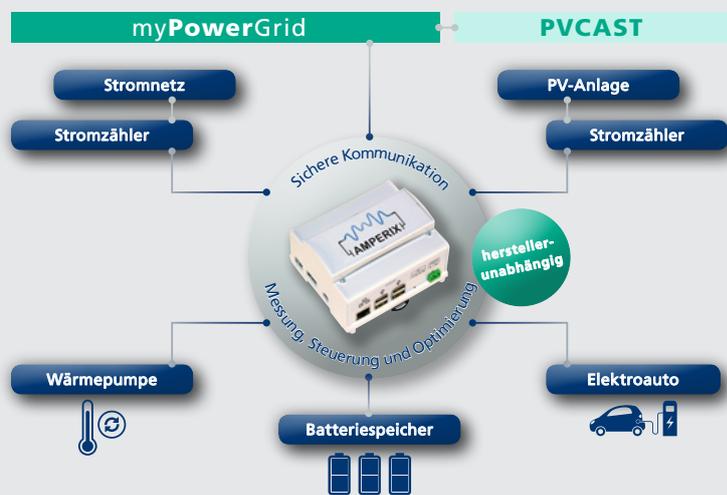
Der in SAP S/4HANA integrierte Variantenkonfigurator ermöglicht das effiziente Variantenmanagement für die konfigurierten Produktlinien eines Unternehmens. Dabei spielt es keine Rolle, ob das Unternehmen Autos oder Pizzas herstellt oder seine isolierten Software-Produkte zusammenführen will: Alle möglichen Varianten werden anhand eines Konfigurationsmodells beschrieben und die für den jeweiligen Fall passende konkrete Produktausprägung ermittelt. Die Varianten werden über alle wesentlichen unternehmerischen Prozesse hinweg unterstützt.

Constraint Solving im Backend

Um den gestiegenen Skalierbarkeits- und Performanzanforderungen gerecht zu werden, wird das VC-Backend auf eine neue C++-basierte Constraint-Solving-Technologie umgestellt. Diese ermöglicht das regelbasierte Einschränken von Variablenbelegungen. Für die Definition solcher Regeln wurden Schnittstellen entwickelt, um gültige Variablenbelegungen über Variantentabellen oder auch Boolesche Ausdrücke zu beschreiben. Des Weiteren haben wir Variablentypen für die Verarbeitung von Zeichenketten sowie hochgenauen Fließkommazahlen ergänzt.

Herausforderungen

Der Variantenkonfigurator von SAP wird weltweit auch in kritischen produktiven Abläufen genutzt. Sowohl auf algorithmischer Seite (Optimierung von Speicherbedarf, Rechenaufwand und Laufzeit, Korrektheitsbeweise) als auch bei der Entwicklung (Test-getriebene Entwicklung, hundertprozentige Code-Überdeckung durch Tests, Sechs-Augen-Prinzip, kontinuierliche Multiplattform-Integration, randomisierte Tests) wurden höchste Standards genutzt, um Skalierbarkeit und Robustheit der Lösungen sicherzustellen.



1

AMPERIX – DREHSCHEIBE FÜR ENERGIE

Das Energiemanagementsystem Amperix etabliert mit einer optimierten Steuerung von Batteriespeichersystemen, Wärmepumpen und Ladestationen für E-Fahrzeuge eine Energie-Drehscheibe für Haushalt und Industrie. In Verbindung mit der myPowerGrid-Plattform ebnet es darüber hinaus den Weg für neue Geschäftsmodelle.

Als Entscheidungsgrundlage für die Steuerung werden durch den Amperix alle Energieflüsse in einem Haushalt erfasst. Dazu zählen die Erzeugung, der Bezug oder die Einspeisung ins Netz und die Berücksichtigung größerer Verbraucher wie etwa Wärmepumpe oder Elektrofahrzeug. Die intuitive Darstellung macht die Energieflüsse für die Nutzer völlig transparent.

Sektorkopplung von Strom, Wärme und Elektromobilität

Anfangs hatte die Steuerung der Batteriespeichersysteme Priorität bei der Entwicklung des Amperix. Die Auswertungen der Pilotinstallationen zeigten jedoch, dass in den Sommermonaten trotz Batteriespeichern erhebliche Energieüberschüsse anfallen. Mittlerweile können diese genutzt werden, zum Beispiel zum Aufheizen des Warmwasserspeichers mit einer Wärmepumpe oder für das sonnengeführte Laden von Elektrofahrzeugen. Die Sektorkopplung von Strom, Wärme und Elektromobilität birgt noch ein erhebliches Potenzial zur Steigerung der Versorgung mit eigenerzeugter Energie.

Prognosebasierte Steuerung

Neben der intelligenten Verwertung aktueller Messdaten verarbeitet der Amperix Erzeugungsprognosen für Photovoltaikanlagen und Lastprognosen für den Haushaltsverbrauch und die Wärmepumpe; auch die Wünsche der Bewohner, wann das Elektrofahrzeug abfahrbereit sein soll, werden berücksichtigt. Diese Prognosen ermöglichen eine effizientere Nutzung der Flexibilität im Haushalt.

Herstelleroffen und skalierbar

Der Amperix ist herstelleroffen und kompatibel mit gängigen Produkten. Auch die Kombination bereits vorhandener Infrastruktur mit neuen Komponenten ist möglich. Mittlerweile wird der Amperix nicht nur zur Steuerung von Speichern in Privathaushalten zuverlässig eingesetzt, sondern auch in der Industrie zum Abtragen von Lastspitzen. So ermöglicht der Amperix die Integration unterschiedlichster Systeme in eine Internetplattform, die der Visualisierung und dem Management der Kundensysteme, aber auch der Schaffung virtueller Großbatterien und damit der Umsetzung neuer Geschäftsmodelle dient.

1 *Der Amperix wird vor Ort mit den Energiezählern, den steuerbaren Geräten und über das Internet sicher mit der myPowerGrid-Plattform verbunden. Intelligente Steuerungsalgorithmen ermöglichen einen optimierten Betrieb der Geräte.*



NEWS AUS DER ABTEILUNG

UNSICHERE LADEINFRASTRUKTUR AN E-TANKSTELLEN

Sehr großes mediales Echo rief Mathias Dalheimers Vortrag bei der Jahreskonferenz des CCC in Leipzig hervor – sein Thema: Sicherheitsprobleme an E-Tankstellen. Manche Ladekarten sind unsicher und ermöglichen es, auf fremde Rechnung zu laden. Schwachstelle ist die Identifikationsnummer; sie ist öffentlich einsehbar und kann beliebig kopiert werden. Auch die Kommunikation zwischen den Ladesäulen und dem Abrechnungsbackend ist schlecht geschützt. Mit geringem technischem Aufwand kann man diese Kommunikation abfangen und erhält so die Kartennummern von Kunden. Die ersten Ladenetzbetreiber haben die Schwachstellen bestätigt und ziehen nach dem großen Medienecho auch erste Konsequenzen. Geplant ist außerdem ein Experten-Konsortium, das diese Probleme systematisch angeht.

NEUES BMBF-PROJEKT ZU DEEP LEARNING

Kaum ein Fachgebiet hat sich in den letzten Jahren derart rasant und erfolgreich entwickelt wie Machine Learning, zu dem auch Deep Learning gehört. Diese für viele praktische Anwendungen erfolgreiche Vorgehensweise erfordert allerdings einen enormen Rechenaufwand und sehr viel Trainingsdaten. Darum müssen Methoden und Infrastrukturen entwickelt werden, welche auch in Zukunft eine praxistaugliche Berechenbarkeit immer komplexerer neuronaler Netze sicherstellen. Dies fördert das BMBF seit Ende 2017 im Projekt High Performance Deep Learning Framework – Softwareumgebung zum effizienten Entwurf von tiefen neuronalen Netzen auf Hochleistungsrechnern mit insgesamt 2 Millionen Euro. Koordiniert wird das Projekt, das eine Laufzeit von drei Jahren hat, am Competence Center High Performance Computing.

EUROPÄISCHER MIKROPROZESSOR EPI

Zusammen mit dem Fraunhofer IIS sind wir Teil des Konsortiums, welches bis 2021 einen europäischen Mikroprozessor entwickelt. Der Fokus der European Processor Initiative (www.european-processor-initiative.org) liegt auf dem High Performance Computing. Gefördert durch die Europäische Union soll das Projekt Kompetenzen im Bereich der Chipentwicklung vereinen und die EU zukünftig unabhängiger von außereuropäischen Mitbewerbern machen. Die Rahmenvereinbarung wurde Ende letzten Jahres unterzeichnet und die Entwicklungsarbeit startet im Sommer 2018.



Von vorne, links nach rechts: Farooq Arshad, Dr. Dimitar Stoyanov, Dominik Loroch, Dr. Tiberiu Rotaru, Dr. Rui Machado, Dr. Franz-Josef Pfreundt, Patrick Reh, Dr. Martin Kühn, Dr. Matthias Balzer, Dr. Abel Amirbekyan, Adrien Roussel, Frauke Santacruz, Tina Hill, Dr. Khawar Ashfaq Ahmed, Dr. Norman Etrich, Matthias Klein, Tobias Götz, Matthias Deller, Christian Mohrbacher, Phoebe Buckheister, Bernd Lörwald, Lukas Ristau, Dr. Jens Krüger, Delger Lhamsuren, Bernd Lietzow, Thomas Olszamowski, Julius Roob, Dr. Valeria Bartsch, Dr. Mirko Rahn, Dr. Alexander Janot, Dr. Dominik Straßel, Dr. Peter Labus, Kai Krüger, Dr. Daniel Grünewald, Javad Fadaie Ghotbi, Dr. Alexander Klauer, Raju Ram, Dr. Dirk Merten