Verwendungsnachweis ThinKIsense NXP (Inhaltlicher Teil)

Teil II: Eingehende Darstellung (öffentlich)

Einleitung

Industrielle Anlagen erzeugen heute große Mengen an Sensordaten. Viele moderne Multisensoren messen Vibrationen, Akustik und Umgebungsgrößen mit hohen Abtastraten. Diese Daten werden häufig vollständig in die Cloud übertragen, wo sie verarbeitet und gespeichert werden. Das ist technisch einfach, jedoch energieintensiv. Die Übertragung großer Datenmengen verkürzt die Batterielaufzeit deutlich und verursacht laufende Kommunikations- und Infrastrukturkosten. In der Prozessautomatisierung ist die vorausschauende Wartung inzwischen ein zentrales Thema. Anlagenbetreiber möchten den Zustand von Pumpen, Ventilen oder Motoren frühzeitig erkennen, um ungeplante Stillstände zu vermeiden. Smarte Sensoren spielen dabei eine wichtige Rolle. Sie erfassen mechanische Veränderungen oft lange bevor ein Fehler sichtbar wird. Viele dieser Systeme arbeiten jedoch noch als reine Datensammler. Die eigentliche KI-gestützte Auswertung erfolgt anschließend in der Cloud. Das führt zu langen Kommunikationswegen, hohem Energiebedarf und einer starken Abhängigkeit von Netzverbindungen.

Genau an diesem Punkt setzte das Verbundprojekt ThinKIsense an. Die Idee entstand aus der Beobachtung heraus, dass lokale KI-Verarbeitung direkt im Sensor zahlreiche Vorteile bringen kann. Die Projektpartner wollten prüfen, ob moderne KI-Modelle in stark reduzierter Form auf energieeffizienten Mikrocontrollern oder spezialisierten Beschleunigern lauffähig sind. Nur wenige Kennwerte (KPIs) sollten anschließend über ein schmalbandiges Funknetz übertragen werden. Dadurch sollte der Energiebedarf sinken, die Batterielaufzeit steigen und die gesamte Lösung robuster gegenüber Netzschwankungen werden. ThinKIsense konzentrierte sich auf ein praxisnahes Beispiel: eine Exzenterschneckenpumpe. Pumpen erzeugen charakteristische Vibrationsmuster im Normalbetrieb und verändern ihr Verhalten je nach Last oder Störung. Die ausgewählte Pumpe wurde im Projektverlauf mehrfach instrumentiert, um Körper- und Luftschall, Drehzahl, Druck und Durchfluss aufzuzeichnen. Auf Basis dieser Daten sollten KI-Modelle entwickelt werden, die Betriebszustände zuverlässig erkennen. Verschiedene Fehlerbilder wie Überdruck oder Kavitation wurden in den Messkampagnen bewusst erzeugt, um realistische Trainingsdaten zu erhalten.

Eine zentrale Grundlage war die Definition einer modularen Systemarchitektur. Die Architektur sollte sicherstellen, dass unterschiedliche KI-Ansätze miteinander vergleichbar sind. Das betraf klassische neuronale Netze, RISC-V-basierte Spezialinstruktionen und neuromorphe Verfahren wie Spiking Neural Networks. Die Modelle mussten so gestaltet werden, dass sie auf begrenzter Rechenleistung ausgeführt werden können. Gleichzeitig sollten sie ausreichend Genauigkeit erreichen, um im industriellen Umfeld nutzbar zu sein. Diese Ziele wurden in mehreren Workshops gemeinsam abgestimmt und dokumentiert. Eine Herausforderung bestand außerdem darin, geeignete Sensorik und Datenerfassungssysteme zu finden. Sicherheit spielte von Anfang an eine große Rolle. Sensoren im Edge-Bereich müssen gegen Manipulation geschützt sein, da sie kritische Prozessdaten verarbeiten. Das Projekt untersuchte mögliche Bedrohungen systematisch. Dabei wurden Angriffspunkte wie

Firmware-Manipulation, unautorisierte Netzkommunikation oder das Abgreifen sensibler Daten betrachtet. Die Ergebnisse flossen in ein Sicherheits- und Resilienzkonzept ein, das später für den Demonstrator umgesetzt wurde. Zusammengefasst entstand die Idee zu ThinKIsense aus einem klaren industriepraktischen Bedarf. Viele Unternehmen möchten datenbasierte Wartungslösungen einsetzen, stoßen aber bei Energieverbrauch, Datenmengen und Kommunikationsinfrastruktur an Grenzen. Das Projekt sollte zeigen, wie intelligente Edge-Sensoren diese Herausforderungen lösen können. Es verband Forschungsthemen wie KI-Modelloptimierung, neuromorphe Verarbeitung und Edge-Security mit der praktischen Umsetzung in einem Demonstrator. Die Kombination aus realen Messdaten, industrieller Sensorik und KI-Methoden machte das Vorhaben zu einem technisch anspruchsvollen, aber praxisnahen Forschungsvorhaben.

Ziele

Das Vorhaben ThinKIsense verfolgte das übergeordnete Ziel, Sensorknoten für industrielle Anwendungen deutlich energieeffizienter und zugleich intelligenter zu machen. Im Mittelpunkt stand die Verlagerung von KI-basierten Auswertungen aus der Cloud direkt in den Sensor. Damit sollte die Menge der zu übertragenden Daten massiv reduziert werden. Durch diesen Ansatz wollte das Projekt zeigen, dass sich die Batterielaufzeit von Sensorknoten erheblich verlängern lässt, ohne auf präzise Zustandsüberwachung verzichten zu müssen. Diese Grundidee wurde bereits zu Projektbeginn klar definiert und im Konsortium abgestimmt.

Ein zentrales Ziel war die Entwicklung einer Systemarchitektur, die unterschiedliche KI-Verfahren unterstützen kann. Dazu gehörten klassische neuronale Netze, spezialisierte RISC-V-Prozessorvarianten sowie neuromorphe Ansätze auf Basis von Spiking Neural Networks. Alle drei Technologien sollten in einem Sensorknoten lauffähig sein, der nur sehr begrenzte Rechen- und Speicherressourcen hat. Die Modelle mussten daher robust, kompakt und effizient sein. Die Architektur sollte modulare Schnittstellen bieten, damit die Varianten vergleichbar bewertet werden können.

Ein weiterer Schwerpunkt lag auf der Energieeffizienz. Das Projekt verfolgte das Ziel, den Energiebedarf sowohl der Datenübertragung als auch der Berechnung zu reduzieren. Durch lokale KI-Inferenz sollte der Sensor nur noch wenige Kennwerte senden müssen. Ergänzend dazu sollte die Firmware so gestaltet werden, dass sie längere Schlafphasen nutzt und nur die notwendige Hardware aktiviert. Auf diese Weise sollte eine deutlich längere Batterielaufzeit ermöglicht werden, als sie bei heutigen industriellen Multisensoren üblich ist.

Darüber hinaus enthielt das Projekt klare Ziele im Bereich Sicherheit und Resilienz. Da Sensordaten in industriellen Anlagen sicherheitsrelevant sein können, sollte der Sensorknoten vor Manipulation geschützt werden. Zu den Zielen gehörten ein sicherer Systemstart, signierte Firmware, verschlüsselte Kommunikation und eine eindeutige Authentifizierung des Geräts. Eine umfassende Bedrohungsanalyse sollte die Grundlage bilden, um gezielte Gegenmaßnahmen zu definieren. Dieser Aspekt war ein eigenständiges technisches Ziel und wurde als integraler Teil des Gesamtsystems betrachtet.

Ein grundlegendes Ziel war die Erstellung eines Demonstrators, der die Technologie im realen Betrieb zeigt. Der Demonstrator sollte eine industrielle Pumpe überwachen und die komplette Prozesskette abbilden: Datenerfassung, KI-Inferenz im Sensor, sichere Übertragung der Ergebnisse und Darstellung in einer Cloud-Umgebung. Damit sollte die technische Machbarkeit aller zuvor definierten Ziele nachgewiesen werden.

Zusammengefasst konzentrierten sich die Ziele des Vorhabens auf:

- die deutliche Reduktion der übertragenen Datenmengen,
- die Entwicklung energieeffizienter und robuster KI-Modelle für Edge-Geräte,
- die Absicherung der gesamten Datenverarbeitungskette,
- die Demonstration eines realen, funktionsfähigen Systems,
- und die Schaffung einer technologischen Grundlage für zukünftige industrielle Lösungen.

Projektorganisation und Koordination

Die Organisation des Verbundprojekts ThinKIsense basierte auf einer klaren Rollenverteilung und einer eng abgestimmten Zusammenarbeit der beteiligten Industrie- und Forschungspartner. NXP übernahm im Projekt die Rolle des Konsortialleiters und war damit verantwortlich für die inhaltliche und organisatorische Gesamtkoordination. NXP strukturierte das Projekt in regelmäßigen zweiwöchentlichen Status- und Fachmeetings, die über die gesamte Laufzeit hinweg durchgeführt wurden. In diesen Terminen wurden Arbeitspakete synchronisiert, technische Fragen diskutiert sowie Fortschritte, Risiken und nächste Schritte abgestimmt. NXP führte dabei die Sitzungen, dokumentierte die Ergebnisse und stellte sicher, dass Abhängigkeiten zwischen den Partnern frühzeitig erkannt und adressiert wurden. Diese kontinuierliche Koordination war notwendig, da mehrere technische Linien (klassische KI, RISC-V-Beschleuniger und neuromorphe Verarbeitung) parallel entwickelt wurden und schließlich in einem gemeinsamen Sensorsystem zusammengeführt werden mussten.

Ein wichtiger Bestandteil der Projektorganisation waren neben den regelmäßigen Meetings mehrere Workshops zu zentralen Themen. Dazu gehörten Workshops zur Definition des Use Cases und der Systemanforderungen, zur Systemarchitektur, zu Security-Themen, KI-Modellen, Datensätzen sowie zur Demonstrator-Planung in späteren Projektphasen. NXP bereitete diese Workshops teilweise vor, stellte Inhalte zusammen und moderierte die Diskussionen. Dadurch entstand früh ein gemeinsames Verständnis der architektonischen und funktionalen Ziele des Projekts.

Neben der Moderation fachlicher Abstimmungen übernahm NXP auch Aufgaben im administrativen Bereich. Dazu gehörte die Kommunikation mit dem Projektträger, die Klärung formaler Fragen sowie die termingerechte Bereitstellung der erforderlichen Berichte. Bereits vor dem offiziellen Projektstart koordinierte NXP die Erstellung und Abstimmung zentraler Unterlagen und sorgte dafür, dass alle Partner ihre Beiträge fristgerecht einlieferten. Auch die Vorbereitung des Projekt-Kick-Offs, der Meilensteintreffen, sowie des Projektabschlusstreffens lag organisatorisch bei NXP.

Wesentlich war auch die strukturierte Kommunikation zwischen den Partnern in den späteren Projektphasen. In dieser Zeit fanden mehrere Fachtreffen statt, darunter das Meilensteintreffen MS2 in Karlsruhe, das FZI Open House und weitere Abstimmungen zum Demonstratoraufbau. NXP stellte sicher, dass die Ergebnisse dieser Treffen vollständig dokumentiert wurden und in die nächsten Projektphasen einflossen. Die enge Zusammenarbeit ermöglichte es, technische Herausforderungen bei Integration, Datenverarbeitung oder Firmwareentwicklung zügig zu lösen.

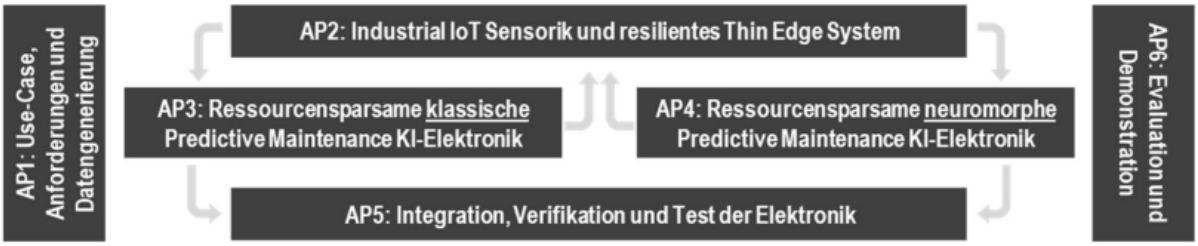
Die Projektorganisation war damit nicht nur geprägt von einer Rolle NXPs als Konsortialleiter, sondern vor allem auch einer engen Abstimmung im Konsortium und einem klar strukturierten Kommunikationsprozess. Dies trug wesentlich dazu bei, dass das Projekt trotz komplexer technischer Inhalte und der Beteiligung mehrerer Partner zielgerichtet voranschritt.

Im Laufe des Projekts kam es zu personellen Engpässen im Konsortium, die sich auf mehrere Arbeitspakete auswirkten. Dadurch verzögerten sich insbesondere Integrations- und Evaluierungsarbeiten sowie die finale Umsetzung des Demonstrators. Um die Projektziele vollständig und in der geplanten Qualität zu erreichen, wurde eine kostenneutrale Verlängerung um sechs Monate beantragt.

Die ursprüngliche Projektlaufzeit war vom 01.07.2022 bis 30.06.2025, durch die bewilligte Verlängerung änderte sich das Projektende auf den 31.12.2025. Dies gab allen Partnern ausreichend Zeit, die offenen technischen Arbeiten abzuschließen, die KI-Modelle final zu testen und den Demonstrator wie geplant umzusetzen und zu präsentieren.

Arbeitspakete und deren Ergebnisse

Das Vorhaben ThinKIsense war in sechs Arbeitspakete unterteilt. Die Anwendungsfälle und Anforderungen wurden in AP1 definiert. Außerdem wurden hier die Daten für die KI-basierte Predictive-Maintenance-Anwendung generiert. In AP2 wurden Konzepte für die IIoT-Sensorik und für einen effizienten und robusten Betrieb als Teil eines Edge-Computing-Gesamtsystems erforscht. Auch die systemweite Energiebilanz und das Powermanagement wurden betrachtet. Darauf aufbauend wurden in AP3 und AP4 Elektroniklösungen für klassische KI und neuromorphe Verarbeitung im Sensor entwickelt. Es wurden Ansätze für den Einsatz von Ultra-Low-Power RISC-V basierten Prozessoren (AP3) und die Beschleunigung von Spiking Neural Networks (AP4) untersucht. Diese Ansätze verprachen ein hohes Energie-einsparungspotenzial. Für die Verifizierung und Erprobung in AP5 wurden Sensor-Demonstratoren gebaut und Labortests sowie Benchmark-Messungen durchgeführt. In AP6 wurden die Konzepte und Ergebnisse des Gesamtprojekts evaluiert und demonstriert. Die folgende Abbildung zeigt den Aufbau der Arbeitspakete:



Nachfolgend sind die einzelnen Arbeitsergebnisse und deren jeweiligen Ergebnisse aufgeführt.

AP1: Use Case, Anforderungen und Datengenerierung

Im Arbeitspaket 1 wurden die Grundlagen des Gesamtvorhabens geschaffen. Dazu gehörten die Analyse der relevanten Anwendungsfälle, die Definition der funktionalen und technischen Anforderungen, die Festlegung der Grobarchitektur sowie die Planung und Durchführung der Datenerfassung zur späteren Entwicklung der KI-Modelle. Das Arbeitspaket wurde über mehrere Workshops und Abstimmungsrunden im Konsortium erarbeitet und 2023 abgeschlossen.

T-AP 1.1: Analyse von Anwendungsfällen und relevanten Messdaten

Zu Beginn des Projekts wurde eine umfassende Analyse möglicher Anwendungsfälle durchgeführt, um ein industriell realistisches Szenario für die Entwicklung eines Thin-Edge-Sensorsystems auszuwählen. Dabei betrachtete das Konsortium verschiedene Anlagen und typische Problemstellungen der Zustandsüberwachung. Die Analyse zeigte, dass Pumpensysteme ein besonders geeignetes Umfeld darstellen, da sie klar unterscheidbare Betriebszustände und typische Fehlerprofile aufweisen, die sich gut über Vibrations- und

Akustikdaten abbilden lassen. So fiel die Wahl auf die Überwachung einer Exzenterschneckenpumpe, da hier mechanische Veränderungen frühzeitig in den Messsignalen sichtbar werden und sich gut für KI-gestützte Analyseverfahren eignen.

Im Rahmen der Analyse wurden auch die physikalischen Messgrößen identifiziert, die für spätere KI-Modelle relevant sind. Körperschall- und Vibrationsdaten bilden den zentralen Informationskanal, ergänzt durch Luftschallmessungen. Prozessparameter wie Drehzahl, Druck und Durchfluss wurden ebenfalls als essenziell eingestuft, da sie wertvolle Label-Informationen liefern und eindeutige Zuordnungen zu den jeweiligen Zuständen ermöglichen. Darüber hinaus untersuchte das Konsortium den Status quo bestehender industrieller Multisensoren und stellte fest, dass viele Systeme große Rohdatenmengen übertragen und daher einen vergleichsweise hohen Energiebedarf besitzen. Genau hier setzte ThinKIsense an, indem es die Analyse in den Sensor verlagern wollte, um Kommunikationsaufwand und Energieverbrauch zu reduzieren.

T-AP 1.2: Auswahl und Beschreibung der Ziel-Anwendungsfälle

Nachdem die potenziellen Anwendungsbereiche analysiert waren, wurden in diesem Teilarbeitspaket die konkreten Ziel-Anwendungsfälle definiert, die später durch die verschiedenen KI-Modelle erkannt oder vorhergesagt werden sollen. Dazu gehörten sowohl reguläre als auch gezielt herbeigeführte Betriebszustände, die typisch für den Lebenszyklus einer industriellen Pumpe sind. Die Partner verständigten sich auf folgende Kernzustände: Normalbetrieb, Kavitation, Überdruck sowie verschiedene Last- und Drehzahlzustände. Diese Szenarien bilden ein breites Spektrum realer Betriebssituationen ab und eignen sich daher gut als Grundlage für die KI-Modellierung.

Gemeinsam wurde festgelegt, welche Messpunkte an der Pumpe relevant sind und welche Sensoren dort eingesetzt werden sollen. Drei Messpunkte wurden ausgewählt, die unterschiedliche Bereiche der Pumpe abdecken: der Einlass, der Statorbereich und der Glockenbereich. Die Kombination ermöglicht eine umfassende Erfassung der Schwingungsmuster und bietet robuste Daten für die spätere Analyse. Diese Entscheidungen wurden im Rahmen eines zentralen Workshops zur Anwendungsfall-Definition im Oktober 2022 abgestimmt. In diesem Workshop wurden außerdem die notwendigen Label-Informationen, Messraten, Datenformate und Testbedingungen festgelegt.

T-AP 1.3: Definition der Anforderungen an Sensorik, Daten, Algorithmen und Hardware

In diesem Teilarbeitspaket wurden die Anforderungen an das spätere Sensorsystem detailliert beschrieben. Dazu gehörten funktionale Anforderungen, Datenanforderungen, KI-Anforderungen sowie Anforderungen an Hardware, Firmware und Sicherheit.

Auf der Datenebene wurde definiert, dass die Vibrationsdaten mit hohen Abtastraten aufgenommen werden müssen, da nur so die relevanten Frequenzanteile industrieller Pumpensysteme abgebildet werden können. Außerdem war festgelegt, dass alle Messkanäle exakt zeitlich synchronisiert sein müssen. Um dies sicherzustellen, wurde eine NTP-basierte Zeitsynchronisation als verbindliche Anforderung definiert. Die Datenerfassung sollte kontinuierlich und ohne Ausfälle erfolgen, da vollständige Zeitreihen für das Modell-Training erforderlich sind. Auch die eindeutige Erfassung der Prozessparameter war ein zentraler Bestandteil der Anforderungen, damit die Modelle zuverlässig zwischen Normal- und Fehlerzuständen unterscheiden können.

Auf algorithmischer Ebene wurde definiert, dass das System drei verschiedene KI-Pfadvarianten unterstützen muss:

- klassische KI-Modelle auf einem energieeffizienten ARM-Mikrocontroller,
- einen RISC-V-basierten, hardwarebeschleunigten Ansatz,
- neuromorphe Modelle auf Basis von Spiking Neural Networks.

Diese Varianten sollten sich modular in das Gesamtsystem integrieren lassen und auf der gleichen Datenbasis arbeiten. Die KI-Modelle mussten besonders kompakt sein, um die begrenzten Ressourcen eines Edge-Sensors einzuhalten. Die Firmware sollte flexibel genug sein, um alle drei Modellvarianten mit angemessener Energieeffizienz auszuführen.

Da die Sensordaten potenziell sensibel sind, wurden bereits in AP1 grundlegende Sicherheitsanforderungen festgelegt. Dazu gehören: geschützt startende Firmware, verschlüsselte Kommunikationskanäle, signierte Updates und Schutzmechanismen für KI-Modelle und Schlüsselmaterial. Diese Anforderungen bildeten den Ausgangspunkt für die spätere STRIDE/DREAD-Analyse in AP2, die das Sicherheitskonzept weiter verfeinerte.

T-AP 1.4: Erstellung der Grobarchitektur und Schnittstellendefinition

Als Abschluss von AP1 wurde eine Grobarchitektur erstellt, die die grundlegenden Bausteine des Systems, die Datenflüsse sowie alle maßgeblichen Schnittstellen beschreibt. Diese Architektur wurde in einem gemeinsamen Workshop im November 2022 abgestimmt. Sie umfasst:

- die Sensorik und deren Messketten,
- die Vorverarbeitungsschritte auf der Host-MCU,
- die modularen KI-Pfade (klassisch, RISC-V-beschleunigt und neuromorph),
- die Kommunikationsmodule NB-IoT und BLE,
- das Cloud-Backend zur Aufnahme und Visualisierung der KPIs.

Für alle Komponenten wurden die Schnittstellen eindeutig beschrieben. Dazu gehören Datenformate zwischen Sensorik und Firmware, SPI-basierte Schnittstellen zu den KI-Beschleunigern, Struktur der internen Datenpipelining-Kette sowie die Definition der übertragenen Ergebnisdaten über NB-IoT. Die Grobarchitektur fungierte als verbindlicher technischer Referenzrahmen für die weiteren Arbeitspakete und stellte sicher, dass alle Partner entlang der gleichen technischen Spezifikation arbeiten.

Ergebnis von AP1

Mit Abschluss des Arbeitspakets lagen ein definierter Use Case, vollständige technische Anforderungen, eine abgestimmte Grobarchitektur sowie umfangreiche Messdaten aus zwei Datenerfassungsphasen vor. Diese Ergebnisse bilden die Grundlage für die Umsetzung der KI-Modelle, die Security-Analyse und die Systemintegration in den Arbeitspaketen 2 bis 6. AP1 wurde im Jahr 2023 abgeschlossen.

AP2: Industrial IoT Sensorik und resilientes Thin-Edge-System

Das Arbeitspaket 2 beschäftigte sich mit allen technischen Grundlagen für ein robustes, energieeffizientes und abgesichertes Thin-Edge-Sensorsystem. Während AP1 die fachlichen und anwendungsbezogenen Rahmenbedingungen definierte, konzentrierte sich AP2 darauf, diese Anforderungen in ein tatsächliches Systemdesign zu überführen. Dazu gehörten Analysen der verfügbaren IoT-Sensorik, der Sicherheitsanforderungen, der Datenverarbeitung auf dem Edge-Knoten sowie der Aufbau einer vollständigen Zielarchitektur. Die hier erarbeiteten Konzepte bildeten den Kern des späteren Demonstrators und legten fest, wie die KI-Modelle technisch ausgeführt und abgesichert werden können.

T-AP 2.1: Analyse des Status Quo aktueller IoT-Sensorik

Zu Beginn von AP2 wurde der aktuelle Stand industrieller IoT-Sensorik umfassend bewertet. Dabei zeigte sich sehr deutlich, dass viele heute eingesetzte Multisensorgeräte große Mengen an Rohdaten erfassen, diese jedoch weitgehend ungefiltert an eine Cloud senden, wo anschließend die eigentliche Auswertung stattfindet. Die dabei transmit-tierte Datenmenge ist beträchtlich und führt in der Praxis zu einem spürbar hohen Energiebedarf, was ein entscheidender Nachteil für batteriebetriebene oder schwer zugängliche Sensorknoten ist. Diese Beobachtung bestätigte die bereits in AP1 formulierte Grundannahme des Projekts: Um industrielle Predictive-Maintenance-Sensoren praxistauglicher zu machen, ist eine Verlagerung der Datenverarbeitung in den Sensor selbst notwendig.

Neben dem Energieverbrauch wurde untersucht, wie kommerzielle Sensoren mit Störungen umgehen. Viele Sensoren bieten zwar einfache Puffermechanismen, reagieren aber empfindlich auf Funkabbrüche, kurze Netzunterbrechungen oder stark schwankende Datenraten. Diese Erkenntnisse waren zentral für die spätere Entwicklung eines deutlich robusteren Thin-Edge-Systems, das durch lokales Preprocessing, Wiederanlaufmechanismen und synchronisierte Messabläufe zuverlässiger arbeiten soll.

T-AP 2.2: Analyse der resilienten Datenverarbeitung auf der Edge

Nachdem klar war, dass die bestehenden Systeme in Bezug auf Resilienz und Energieeffizienz an Grenzen stoßen, wurde untersucht, wie eine robuste Datenverarbeitung direkt im Sensor realisiert werden kann. Eine wesentliche Erkenntnis war, dass die gesamte Signalverarbeitung nur dann zuverlässig funktioniert, wenn die Sensorik hochgradig synchron arbeitet. Aus diesem Grund wurde der spätere Prüfstand mit einer NTP-basierten Zeitsynchronisation ausgestattet, wodurch Vibrations- und Prozessdaten eindeutig einem Zeitraster zugeordnet werden konnten. Dies ist für die spätere KI-Verarbeitung entscheidend, da KI-Modelle empfindlich auf zeitliche Verschiebungen reagieren.

Im Rahmen dieses Teilpakets wurden zudem grundlegende Überlegungen zu Fehlertoleranz und Wiederanlaufverhalten entwickelt. Da der Edge-Sensor im späteren Betrieb regelmäßig in tiefen Schlafmodi laufen soll, musste gewährleistet werden, dass Daten sowohl vor als auch nach einem Sleep-Zyklus korrekt erfasst und verarbeitet werden. Ebenso wurde betrachtet, wie mit kurzzeitigen Störungen in der Funkübertragung oder Sensorperipherie umzugehen ist. Diese Erkenntnisse flossen direkt in die Firmware-Architektur ein, die später so aufgebaut wurde, dass sie Messdaten puffern, erneut synchronisieren und konsistent weiterverarbeiten kann.

T-AP 2.3: Analyse von Systemangriffen und Bedrohungsszenarien

Ein zentraler Bestandteil von AP2 war die umfangreiche Sicherheitsanalyse. Die Partner führten eine vollständige STRIDE-/DREAD-Risikoanalyse durch, in der alle relevanten Komponenten des Systems betrachtet wurden. Vom physischen Vibrationssensor über die MCU-Firmware und die KI-Modelle bis hin zur Cloud-Infrastruktur und dem NB-IoT-Modem. Dabei zeigte sich, dass insbesondere drei Bereiche besonders schutzbedürftig sind: die Firmware der MCU, die KI-Modelle selbst und der Kommunikationspfad zum Server. Für alle Bereiche wurden konkrete Bedrohungsszenarien identifiziert, etwa die Manipulation von Firmware, das Abgreifen von Modellparametern oder das Einschleusen gefälschter Ergebnisdaten. Die Analyse bildete die Grundlage für ein ganzheitliches Sicherheitskonzept, das in den folgenden Teilarbeitspaketen umgesetzt wurde. Die systematische Betrachtung aller Assets und Bedrohungen war ein wichtiger Schritt, um sicherzustellen, dass das Thin-Edge-System nicht nur energieeffizient, sondern auch zuverlässig und sicher betrieben

werden kann, insbesondere im industriellen Umfeld, in dem Sensordaten geschäftskritisch sein können.

T-AP 2.4: Entwicklung der Zielarchitektur für Konnektivität und Datenverarbeitung

Auf Basis der Analysen aus den vorherigen Teilpaketen wurde die vollständige Zielarchitektur für ThinKIsense spezifiziert. Diese Architektur definiert, wie Sensorik, Edge-Verarbeitung, Beschleunigerhardware und Kommunikation zusammenwirken. Im Zentrum steht die Host-MCU, die sämtliche Schritte des Mess- und Verarbeitungsablaufs steuert: vom Auslesen der Vibrationsdaten über das Preprocessing bis hin zur Ausführung der KI-Inferenz und der Übertragung der kompakt berechneten Kennwerte. Die Architektur wurde so entworfen, dass sie drei parallele KI-Pfade unterstützt: klassische KI-Modelle, die direkt auf der MCU laufen, einen RISC-V-basierten Beschleuniger sowie einen neuromorphen FPGA-basierten Beschleuniger. Jeder dieser Pfade nutzt die gleichen Eingangsdaten, was eine vergleichende Bewertung ermöglicht.

Neben der Inferenzkette definiert die Architektur auch die Kommunikationsstrategie. Die Übertragung der Ergebniskommunikation erfolgt über ein NB-IoT-Modem, während BLE vor allem für Konfiguration und die Übertragung von Trainingsdaten vorgesehen ist. Dadurch wird die Funklast im späteren Feldbetrieb minimiert, gleichzeitig aber eine flexible Inbetriebnahme gewährleistet. Die Architektur bildet somit ein Bindeglied zwischen den theoretischen Anforderungen aus AP1 und der praktischen Umsetzung in AP5.

T-AP 2.5: Auswahl und Implementierung der Algorithmen

In diesem Teilarbeitspaket wurden die Algorithmen entwickelt, die die Grundlage für die spätere KI-Inferenz bilden. Dazu gehörten sowohl klassische Signalvorverarbeitungsmethoden als auch Algorithmen zur Normalisierung, Filterung und Ereigniskodierung. Besonders wichtig war die Entwicklung einer flexiblen Preprocessing-Pipeline, die sich auf drei KI-Varianten anwenden lässt, also auf klassische neuronale Netze, RISC-V-basierte Beschleuniger und neuromorphe Spiking-Modelle.

Ein großer Schwerpunkt lag auf der Vorbereitung der Daten für die neuromorphen Verfahren. Diese benötigen keine kontinuierlichen Signale, sondern Spike-Trains, die in T-AP 2.5 erstmals systematisch erzeugt wurden. Parallel dazu wurden auch die Datenpfade für klassische KI-Modelle implementiert und die notwendigen Schnittstellen zur RISC-V-Hardware definiert. Diese Arbeiten legten die Grundlage für die spätere Integration in AP3 und AP4, wo die konkreten KI-Engines weiterentwickelt wurden.

T-AP 2.6: Auswahl und Implementierung der Sicherheitsmechanismen

Der Abschluss von AP2 bestand in der Umsetzung der zuvor definierten Sicherheitsmaßnahmen. Basierend auf den Ergebnissen der STRIDE-DREAD-Analyse wurden konkrete Schutzmechanismen ausgewählt und in die Systemarchitektur integriert. Dazu gehören ein Secure-Boot-Mechanismus, der sicherstellt, dass nur autorisierte Firmware auf dem Gerät ausgeführt wird, sowie eine signierte Update-Kette, die Firmware-Manipulation verhindert. Für die Datenkommunikation wurde eine TLS-basierte Ende-zu-Ende-Verschlüsselung vorgesehen, ergänzt durch eine sichere Schlüsselverwaltung innerhalb des EdgeLock Secure Subsystems der MCU.

Zusätzlich wurden Mechanismen entwickelt, um sicherheitsrelevante Ereignisse wie Reboots, ungewöhnliche Datenmuster oder potenzielle Manipulationsversuche erkennen und protokollieren zu können. Die Sicherheitsarchitektur des Systems wurde damit vollständig

spezifiziert und bildet eine wichtige Grundlage für einen praxisnahen Einsatz in industriellen Umgebungen.

Ergebnis von AP2

AP2 lieferte ein vollständiges technisches Fundament:

- eine robuste Analyse der bestehenden IoT-Sensorik,
- ein Sicherheits- und Resilienzkonzept,
- eine modulare, industrietaugliche Zielarchitektur,
- implementierte Daten- und KI-Algorithmen für alle drei KI-Pfade,
- definierte Sicherheitsmechanismen für Firmware, Daten und Kommunikation.

Damit schuf AP2 die funktionale Basis für die spätere Umsetzung und Integration in AP3 bis AP6.

AP3: Ressourcensparsame klassische Predictive-Maintenance-KI-Elektronik

Arbeitspaket 3 befasst sich mit der Entwicklung, Optimierung und Evaluierung der klassischen KI-Modelle, die direkt auf der Host-MCU ausgeführt werden. Im Gegensatz zu den in AP4 betrachteten Beschleunigern (neuromorph und RISC-V) liegt der Fokus hier auf der Ausführung eines vollständigen KI-Modells auf einem energieeffizienten Mikrocontroller. Ziel des AP3 war es, ein KI-Modell zu entwickeln, das trotz begrenzter Rechenleistung präzise Ergebnisse liefert, dabei aber so wenig Energie wie möglich verbraucht.

T-AP 3.1: Aufbau der ARM-basierten Evaluationsumgebung

Zu Beginn wurde eine vollständige Evaluationsumgebung für klassische KI-Modelle aufgebaut. Dies umfasste die Bereitstellung von Entwicklungsboards, Treibern, Firmwaregrundlagen und Messumgebungen. NXP stellte hierfür u. a. die ARM-basierte Host-MCU-Plattform bereit und unterstützte die Partner beim Set-up der Entwicklungsumgebung. Das Ziel war, eine verlässliche Infrastruktur zu schaffen, in der sowohl erste Modellversionen als auch spätere optimierte Pipelines getestet werden konnten. Bereits früh im Projekt wurde deutlich, dass die Evaluationsumgebung nicht nur für Tests des KI-Modells, sondern auch für Energie- und Laufzeitmessungen benötigt wird. Diese Messungen lieferten essenzielle Kennwerte für spätere Optimierungen.

Ein besonderer Schwerpunkt lag darauf, reproduzierbare Messbedingungen zu schaffen. Dies war notwendig, um Modellversionen, Firmware-Optimierungen und Energieeinsparmaßnahmen objektiv vergleichen zu können. Die Evaluationsumgebung wurde deshalb eng an den Datenerfassungsprozess aus AP1 angebunden, sodass alle Modelltests mit realen Pumpendaten erfolgen konnten.

T-AP 3.2: Entwicklung und Optimierung klassischer KI-Modelle auf der MCU

In diesem Teilarbeitspaket wurden die von Siemens bereitgestellten klassischen KI-Modelle weiterentwickelt und für die Edge-Ausführung optimiert. Ursprünglich enthielten diese Modelle rekurrente Strukturen wie LSTM-Schichten, die für Mikrocontroller jedoch zu speicher- und rechenintensiv sind. Daher wurde die Architektur grundlegend überarbeitet: Die LSTM-Komponente wurde entfernt und durch eine effizientere 1D-CNN-basierte Feature-Extraktion ersetzt, die weniger Speicher benötigt und gleichzeitig eine robuste Merkmalsgewinnung aus Vibrationszeitreihen ermöglicht. Diese Änderung führte zu einer höheren Effizienz bei gleichzeitig stabiler Modellleistung.

Parallel dazu wurden wichtige Hyperparameter angepasst, etwa die Anzahl der Filter, Kernelgrößen und die Tiefe der Modellarchitektur. Das Ziel war, ein ausgewogenes Verhältnis

zwischen Modellqualität und Rechenaufwand zu erreichen. Das Modell sollte in der Lage sein, sowohl Regressionen (z. B. Schätzung von Druck oder Drehzahl) als auch Klassifikationen (z. B. Normalzustand, Kavitation, Überdruck) in einem einzigen Modell („Multi-Task-Ansatz“) durchzuführen. Die Modellarchitektur wurde so gestaltet, dass sie auf dem Speicherbudget des Controllers Platz findet, aber dennoch ausreichend hohe Genauigkeit bietet.

T-AP 3.3: Quantisierung und Anpassung an die Zielhardware

Ein wesentliches Element dieses Arbeitspakets war die Konvertierung des Modells auf ein quantisiertes 8-Bit-Format, um es auf der Neural-Processing-Unit (NPU) der MCX-N-Plattform ausführen zu können. Floating-Point-Modelle mit 32-Bit-Präzision sind für Mikrocontroller häufig zu speicherintensiv, weshalb die Quantisierung ein entscheidender Schritt für die spätere Einbettung war. Dabei mussten geeignete Quantisierungsstrategien gewählt werden, um Genauigkeitsverluste zu minimieren. Durch angepasste Operatoren, optimierte Pooling-Strategien und eine sorgfältige Auswahl quantisierungsfreundlicher Aktivierungsfunktionen gelang es, ein kompaktes Modell zu erzeugen, das sich auf dem Zielsystem effizient ausführen lässt.

Die erfolgreiche Quantisierung stellte sicher, dass das Modell nahtlos in die spätere Inferenzpipeline integriert werden konnte. Dies schloss sowohl die Ausführung auf der MCU selbst als auch die Nutzung der integrierten NPU mit ein.

T-AP 3.4: Energie- und Laufzeitmessungen sowie Firmware-Optimierung

Um das klassische KI-Modell praktikabel zu machen, waren umfangreiche Messungen notwendig. In T-AP 3.4 wurden Messreihen zur Bestimmung von Energieverbrauch, Latenz und Speicherbedarf durchgeführt. NXP unterstützte hierbei aktiv beim Aufbau der Energie-Messumgebung und bei der Auswertung der Messergebnisse. Die Messungen zeigten, welche Firmwarepfade die meiste Energie benötigen und wie die Einbettung des KI-Modells den Gesamtverbrauch beeinflusst. Diese Ergebnisse flossen direkt in die Optimierung der Firmware ein.

Parallel zu den Modelloptimierungen wurden Energiesparmaßnahmen auf Firmware-Ebene integriert:

- Sleep- und Deep-Sleep-Modi,
- selektives Abschalten ungenutzter Peripherie,
- Takt- und Speichergating,
- strukturierte Ablaufsteuerung, um Wartezeiten effizient zu nutzen.

Diese Maßnahmen wurden im Rahmen der späteren Arbeiten des Jahres 2025 weiter verfeinert, als die Firmware für den Demonstrator fertiggestellt wurde.

Ergebnis von AP3

Am Ende des Arbeitspakets lagen vollständig funktionsfähige, optimierte KI-Modelle vor, die robust auf dem Mikrocontroller liefen. Die erzielten Ergebnisse zeigen, dass klassische Ansätze trotz begrenzter Ressourcen eine hohe Genauigkeit erreichen können. In den Evaluierungen der Partner konnte das optimierte CNN-Modell stabile Regressions- und Klassifikationsergebnisse liefern und lag in zentralen Kennwerten im Bereich industrieller Anforderungen. Die Modelle dienten später als Baseline für den Vergleich mit den Beschleuniger-Varianten aus AP4.

Durch AP3 konnte gezeigt werden, dass klassische KI-Methoden eine praxistaugliche, energieeffiziente und technisch robuste Lösung darstellen. Sie bieten eine solide Grundlage

für Edge-KI in industriellen Anwendungen, insbesondere wenn Hardware-Kosten, Energieverbrauch und Langlebigkeit entscheidende Faktoren sind.

AP4: Ressourcensparsame neuromorphe Predictive-Maintenance-Elektronik

Arbeitspaket 4 beschäftigte sich mit der Entwicklung und Bewertung eines neuromorphen KI-Ansatzes für die Zustandsüberwachung der Pumpe. Während ein Großteil der Arbeiten (insbesondere die Implementierung der neuromorphen Modelle, die FPGA-Entwicklung und die Ausarbeitung der Spike-Architektur) von den Forschungspartnern FZI und TU Darmstadt übernommen wurde, spielte NXP eine zentrale unterstützende Rolle in der Systemarchitektur, der Integration und dem Hardware-/Firmware-Zusammenspiel. Die konkreten NXP-Beiträge fokussierten sich in diesem Arbeitspaket auf T-AP 4.4 und die Integrationsarbeiten von T-AP4.7

T-AP 4.4: Entwurf einer ressourcensparenden neuromorphen Hardware-Architektur

In diesem Teilarbeitspaket unterstützte NXP das FZI bei der Definition und Ausarbeitung der neuromorphen Gesamtarchitektur. Ein Schwerpunkt lag auf der Abstimmung zwischen den Anforderungen aus der Zielarchitektur (AP2) und den Eigenschaften der neuromorphen Hardware, die auf dem FPGA-System des FZI realisiert wurde. NXP brachte dabei die Perspektive eines industriellen Systemarchitekten ein und stellte sicher, dass die neuromorphe Verarbeitung in das Gesamtkonzept eines Edge-Sensors integriert werden kann.

Wesentlich war dabei die Berücksichtigung der Bottlenecks aus AP 2.4, insbesondere hinsichtlich Speicherbedarf, Datenfluss, Schnittstellen und Energieverbrauch. NXP half dabei, die Rollen der einzelnen Verarbeitungsschritte (Messung, Spike-Konvertierung, Event-Übertragung und Inferenz) in einer Weise zu strukturieren, die die beschränkten Ressourcen eines Edge-Geräts realistisch abbildet. Auch die Security-Architektur aus AP2.6 wurde in diesem Zusammenhang berücksichtigt, um sicherzustellen, dass auch der neuromorphe Pfad innerhalb der definierten Sicherheitsgrenzen betrieben werden kann.

Auf Basis dieser Abstimmungen wurden Anforderungen an Firmware, Datenpfade und Tooling definiert. NXP stellte sicher, dass die Host-MCU die für die neuromorphe Verarbeitung benötigten Datenformate bereitstellen kann und dass die Anbindung an den FPGA-Beschleuniger über eine performante und zuverlässige SPI-Schnittstelle möglich ist. Damit leistete NXP einen entscheidenden Beitrag dazu, dass der neuromorphe Ansatz nicht nur konzeptionell, sondern auch systemarchitektonisch konsistent in ThinKIsense eingebettet wurde.

T-AP 4.7: Evaluation des neuromorphen Systems

Obwohl der FPGA-basierte neuromorphe Beschleuniger durch das FZI implementiert wurde, spielte NXP eine bedeutende Rolle beim Aufbau der Entwicklungs- und Integrationsumgebung, in der die neuromorphe KI später getestet wurde.

NXP stellte hierfür wichtige Systemkomponenten zur Verfügung, darunter Mikrocontroller-Boards, Board-Support-Packages, Treiber und grundlegende Firmwarebausteine, die für die Einbindung des Beschleunigers in das Gesamtsystem erforderlich waren. Die neuromorphe Testumgebung sollte (wie die klassischen und RISC-V-Ansätze) in das ThinKIsense-Sensor-Testsystem integriert werden. NXP unterstützte durch die Bereitstellung der Host-MCU-Umgebung, die Implementierung der Verarbeitungssequenzen und die Anpassung der Firmware, sodass die Event-Daten des

Beschleunigers korrekt empfangen, interpretiert und in die gesamte Datenkette eingebunden werden konnten.

Damit stellte NXP sicher, dass die neuromorphe Pipeline nicht isoliert funktioniert, sondern im späteren Demonstrator zusammen mit Sensorik, Kommunikationsmodulen und den übrigen KI-Pipelines lauffähig ist.

Weitere Mitwirkung von NXP in AP4

Neben den zwei formal dokumentierten NXP-Teilpaketen hat NXP auch in den übrigen Arbeiten von AP4 unterstützend mitgewirkt. Diese Mitwirkung betraf vor allem:

- die Bereitstellung der Host-MCU-Plattform, die für alle drei KI-Pfade (klassisch, RISC-V, neuromorph) identisch ist,
- Abstimmungen zur Schnittstellendefinition, um sicherzustellen, dass die neuromorphen Modelle in das vorhandene Systemdesign passen,
- Unterstützung bei Integrationsfragen, z. B. bei der Übergabe von Preprocessing-Daten, beim Debugging von Schnittstellen oder beim Verhalten des Beschleunigers in realen Messzyklen,
- Mitwirkung an Architektur- und Tooling-Diskussionen, um sicherzustellen, dass die Entwicklungsumgebung konsistent bleibt.

Ergebnis von AP4:

AP4 stellte einen entscheidenden Baustein des Gesamtvorhabens dar. Es zeigte, dass neuromorphe KI-Verfahren in einem praxisnahen industriellen Szenario erfolgreich eingesetzt werden können. Die entwickelte Pipeline fügt sich vollständig in die ThinKIsense-Architektur ein und bildet einen der drei bewerteten KI-Pfade im späteren Demonstrator.

AP5: Integration, Verifikation und Test der Elektronik

Arbeitspaket 5 bildete eine zentrale Phase des Projekts, da hier die zuvor entwickelten Komponenten (Sensorik, Firmware, KI-Modelle, neuromorphe und RISC-V-Beschleuniger sowie Kommunikationsmodule) erstmals in ein gemeinsames Gesamtsystem integriert wurden. Während die Partner in diesem Arbeitspaket jeweils spezifische technische Teilbeiträge leisteten, lag ein wesentlicher Anteil der übergreifenden Integrationsarbeit bei NXP. Die Aufgabe bestand darin, die einzelnen Hardware- und Softwarekomponenten zu verbinden, ihre Interoperabilität sicherzustellen und die Teil- und Gesamtfunktionalität schrittweise zu verifizieren.

T-AP 5.1: Aufbau der Teilintegrationsumgebung

Im Rahmen von T-AP 5.1 entwickelte NXP eine Teilintegrationsebene, auf der die Sensorik, die ARM-basierte Host-MCU und die sicherheitsrelevanten Hardwarekomponenten gemeinsam betrieben werden konnten. Dies war notwendig, um die in AP2 definierten Algorithmen und Sicherheitsmechanismen sowie die in AP3 entwickelten KI-Modelle und die neuromorphe Pipeline aus AP4 auf einer gemeinsamen Hardwarebasis zu testen.

Die Integrationsumgebung basierte auf der zuvor spezifizierten Schnittstellenarchitektur und stellte sicher, dass Daten aus dem Vibrationssensor zuverlässig zur MCU gelangten und dort korrekt vorverarbeitet werden konnten. Zusätzlich wurde die Verbindung zu den externen Beschleunigern auf FPGA-Basis vorbereitet, indem die entsprechenden SPI-Kanäle auf der Host-Plattform implementiert und getestet wurden. NXP übernahm hierbei die Aufgabe, das

Zusammenspiel von Sensorik, Prozessor und Security-Subsystem abzustimmen, dass die einzelnen Verarbeitungsschritte robust und deterministisch abliefen.

Die Umgebung wurde modular aufgebaut, sodass sowohl klassische KI-Modelle als auch der neuromorphe und der RISC-V-Pfad integriert werden konnten. NXP stellte damit sicher, dass die spätere Systemintegration nicht auf mehrere isolierte Testplattformen angewiesen war, sondern auf einer konsistenten, zentralen Architektur aufbaute.

T-AP 5.2: Integration der Schaltungen aus AP3 und AP4

Ein großer Teil der praktischen Arbeit in AP5 bestand darin, die Kommunikation und Datenflüsse zwischen MCU, Sensorik und den FPGA-basierten Beschleunigern systematisch zu testen. NXP war dafür verantwortlich, die Verifikation der vibrosensorischen Daten auf der Host-MCU durchzuführen und sicherzustellen, dass alle Messwerte korrekt eingelesen, gepuffert und dem Preprocessing übergeben wurden.

Darüber hinaus wurden die Schnittstellen zu den Beschleunigern getestet. Dies beinhaltete sowohl die Übertragung der Rohdaten (für den RISC-V-Pfad) als auch der Spike-Daten (für die neuromorphe Verarbeitung). NXP testete die zeitliche Konsistenz der Datenübertragung, die Latenzzeiten und die korrekte Rückübermittlung der KI-Ergebnisse. Besonderes Augenmerk lag auf der Verifizierung der BLE-Datenpfade, die eine stabile drahtlose Kommunikation zwischen Sensor und Monitoring-System sicherstellen sollten.

Die Verifikationsarbeiten waren entscheidend, um sicherzustellen, dass alle Module zuverlässig mit der Host-MCU interagieren können.

Weitere Mitwirkung von NXP in AP5

- Beratung bei der Sensorhardware-Integration,
- Unterstützung der Forschungspartner bei der Nutzung der MCU-Plattform,
- gemeinsame Debugging-Sessions zur Beschleunigerkommunikation,
- Abstimmung der Firmware-Versionen und Toolchains,
- Sicherstellung der Kompatibilität zwischen MCU-Firmware und FPGA-Logik.

Ergebnis von AP5

In AP5 wurde alle zuvor entwickelten Hardware- und Softwarekomponenten in einem durchgängigen Gesamtsystem zusammengeführt. Die von NXP entwickelten und betreuten Integrationspfade stellten sicher, dass die Sensorik, die KI-Modelle, die Beschleunigerhardware und die Kommunikationsmodule zuverlässig zusammenarbeiten.

Durch die Arbeiten in AP5 wurde die technische Grundlage geschaffen, auf der der Demonstrator in AP6 erfolgreich aufgebaut und betrieben werden konnte. Die in diesem Arbeitspaket durchgeführten Integrations- und Verifikationsschritte lieferten den Nachweis, dass das ThinKIsense-Konzept praxistauglich ist und alle Systemkomponenten auf einer gemeinsamen Plattform laufen können.

AP6: Evaluation und Demonstration

Arbeitspaket 6 bildet den Abschluss des technischen Entwicklungsprozesses im Projekt ThinKIsense. In diesem Arbeitspaket wurden die zuvor entwickelten Komponenten in einen volständigen Demonstrator überführt, in Betrieb genommen und abschließend bewertet.

T-AP 6.5: Dimensionierung und Konstruktion

NXP übernahm in diesem Teilarbeitspaket die Aufgabe, die gesamte Elektronik des ThinKIsense-Sensorsystems so zu dimensionieren und zu konstruieren, dass alle zuvor entwickelten Komponenten, insbesondere die Host-MCU, die Beschleunigerpfade (klassisch, neuromorph, RISC-V), die Kommunikationsmodule und die Security-Funktionen, in einem funktionalen Gesamtdemonstrator zusammengeführt werden konnten. Dazu gehörte die Unterstützung bei der Inbetriebnahme der Elektronikkomponenten, einschließlich der Sensor-PCB, der FPGA-Beschleuniger und der drahtlosen Kommunikationsmodule.

Ein weiterer wichtiger Beitrag war die Unterstützung bei der Realisierung der Sicherheitskomponenten. Die im Projekt definierte Sicherheitsarchitektur (Secure Boot, signierte Firmware, geschützte Schlüsselablage und TLS-Kommunikation) musste im Demonstrator vollständig und zuverlässig umgesetzt werden. NXP stellte sicher, dass die Sicherheitsmechanismen in der finalen Firmware korrekt aktiviert und in die Gesamtabläufe des Sensors integriert sind.

Darüber hinaus stellte NXP angepasste BSPs (Board Support Packages), Treiber und MCU-nahe Softwarebausteine zur Verfügung, damit die Partner ihre Komponenten (z. B. FPGA-Beschleuniger) nahtlos auf der NXP-Elektronik integrieren konnten. Diese Softwareeinheiten bildeten die Grundlage für die korrekte Ansteuerung der Sensorik, der Beschleunigerhardware und der Kommunikationsmodule. Durch die Arbeiten in T-AP 6.5 konnte der Gesamtdemonstrator technisch vollständig aufgebaut werden und bildete die Basis für die nachfolgende Inbetriebnahme und Evaluierung.

T-AP 6.6: Montage und Inbetriebnahme

NXP errichtete einen Messaufbau ein, der zu internen Test-, Evaluierungs- und Integrationszwecken diente. Dieser Messaufbau replizierte Teile der elektrischen Architektur des offiziellen Projekt-Demonstrators, jedoch ohne mechanischen Pumpenaufbau, ohne Gesamtsystem-Replikation und ohne Präsentationscharakter.

Der Messaufbau umfasste:

- ausgewählte elektronische Komponenten des Sensorsystems,
- die Host-MCU-Plattform mit der finalen Firmware,
- die angeschlossenen Beschleunigerpfade (klassisch, RISC-V, neuromorph) soweit für Tests notwendig,
- die Kommunikationsmodule, inklusive NB-IoT-Integration für KPI-Tests,
- die Security-Mechanismen, die im Rahmen des Projekts entwickelt wurden.

Dieser Aufbau diente dazu, Firmware-Versionen zu verifizieren, Software-Tooling zu testen und Funktionstests unter kontrollierten Bedingungen durchzuführen.

Eine vollständige Pumpenumgebung war nicht Bestandteil dieses Messaufbaus.

T-AP 6.7: Evaluierung der entworfenen Hardware und KI

In diesem Teilarbeitspaket führte NXP die abschließende Validierung der Ergebnisse aus AP3 und AP4 im fertig aufgebauten Demonstrator durch. Dies beinhaltete die Überprüfung, ob die klassischen KI-Modelle aus AP3 sowie die mittels Beschleuniger unterstützten KI-Modelle korrekt, stabil und mit der erwarteten Performance laufen. Hierbei wurde insbesondere getestet, ob das Preprocessing, die Modellaufrufe, die Datenübertragung im System sowie die energetischen Rahmenbedingungen mit den Anforderungen aus der Zielarchitektur übereinstimmen.

Darüber hinaus validierte NXP das im Projekt entwickelte Sicherheitskonzept im kompletten Systemverbund. Dies beinhaltete Tests des Secure-Boot-Prozesses, der Integritätsprüfungen für Firmware- und Modellcode, der TLS-Kommunikation über das NB-IoT-Modem sowie der sicheren Speicherung sensibler Schlüssel im EdgeLock-Subsystem. Ziel dieser Validierung war es, sicherzustellen, dass der Demonstrator am Ende nicht nur funktional, sondern auch sicherheitsseitig vollständig den Spezifikationen entsprach.

Die Evaluierungsarbeiten schlossen mit einer Systembewertung ab, die bestätigte, dass die im Projekt entwickelten KI- und Sicherheitsmechanismen in der realen Demonstratorumgebung verlässlich funktionieren. Damit wurde der praktische Nachweis erbracht, dass das ThinKIsense-System erfolgreich integriert und betriebssicher ist.

Ergebnis von AP6

Mit AP6 konnte NXP die wichtigen finalen Schritte des Projekts erfolgreich abschließen: die technische Realisierung des Demonstrators und die vollständige Evaluierung der Hardware- und KI-Komponenten im realitätsnahen Betrieb. Die abgeschlossenen Arbeiten zeigten, dass die im Projekt entwickelten KI-Verfahren auf der energieeffizienten MCU-Plattform stabil funktionieren und in den sicheren Kommunikations- und Gesamtablauf eingebettet sind. Damit wurde der finale Funktionsnachweis des ThinKIsense-Konzepts erbracht und das Projekt erfolgreich in die Demonstrationsphase überführt.

Zusammenfassung der Ergebnisse

Vergleich der verfolgten Ansätze

Alle drei untersuchten KI-Varianten wurden funktionsfähig implementiert und im Demonstrator getestet.

Die klassischen KI-Modelle erzielten eine hohe Genauigkeit bei geringem Ressourcenbedarf. Durch die optimierte Feature-Extraktion und die kompakte Modellarchitektur ließen sie sich gut auf Mikrocontrollern ausführen.

Die RISC-V-basierten Beschleuniger ermöglichten eine flexible Erprobung spezialisierter Recheneinheiten. Diese Variante war für die Bewertung zukünftiger hardwarebasierter Optimierungen wichtig.

Die neuromorphen Ansätze zeigten, dass sich das Konzept der ereignisbasierten Datenverarbeitung gut für zeitkritische Signale eignet. Der zugehörige Beschleuniger wurde als offene Forschungsplattform umgesetzt und im Demonstrator integriert.

Energieeffizienz

Die lokale Auswertung im Sensor reduziert die Datenmenge erheblich. Deshalb müssen nur wenige Ergebniswerte übertragen werden. Die Firmware wurde darauf ausgelegt, lange Schlafphasen zu nutzen und alle nicht benötigten Komponenten abzuschalten. Die Tests im Projekt zeigen, dass sich der Energieverbrauch des Gesamtsystems spürbar senken lässt, wenn KI-Modelle direkt im Sensor ausgeführt werden.

Demonstrator

Zum Projektabschluss wurde ein funktionaler Demonstrator aufgebaut. Er zeigt die vollständige Prozesskette in einer industrierelevanten Umgebung:

- Erfassung von Vibrationsdaten einer Pumpe,

- Auswertung der Daten über klassische, RISC-V-basierte oder neuromorphe KI-Modelle,
- Übertragung der Ergebnisdaten über eine schmalbandige Mobilfunkverbindung,
- Darstellung und Analyse in einer Cloud-basierten Benutzeroberfläche,
- Regression von Drehzahl, Druck und Volumenstrom im Pumpensystem,
- Anomalieerkennung wie Überdruck und Kavitation im Pumpensystem.

Der Demonstrator wurde bei Veranstaltungen präsentiert und konnte die Funktionsfähigkeit der im Projekt entwickelten Lösungen überzeugend zeigen.

Fortschritt im Vergleich zur Planung

Das Projekt erreichte alle wesentlichen Meilensteine. Die Datenaufnahme erfolgte vollständig. Die Modelle wurden aufgebaut, trainiert und optimiert. Die Firmware, die Beschleuniger und das Sicherheitskonzept wurden integriert. Die Demonstration des Systems erfolgte wie geplant am Ende des Projekts.

Einige Arbeiten, besonders rund um die Kommunikation und die Energieoptimierung, wurden iterativ weiterentwickelt, da sich im Verlauf neue Erkenntnisse ergaben. Diese Anpassungen waren für die Qualität der Ergebnisse notwendig.

Nutzen und Ausblick

ThinKIsense hat gezeigt, dass intelligente, energieeffiziente Sensorik für industrielle Anwendungen realisierbar ist. Die im Projekt entwickelten Konzepte können in zukünftigen Geräten genutzt werden, um Wartungsprozesse zu verbessern, Anlagenzustände präziser zu erfassen und Batterielaufzeiten zu steigern.

Darüber hinaus eröffnen die Ergebnisse neue Möglichkeiten im Bereich Edge-KI, neuromorpher Verarbeitung und sicherer Sensorarchitekturen. Die beteiligten Industrie- und Forschungspartner werden die gewonnenen Erkenntnisse in ihre jeweiligen Entwicklungen, Produkte und Forschungsarbeiten einfließen lassen.