Zuwendungsempfänger: Universität Bremen
Institut für Telekommunikation und Hochfrequenztechnik

Verbundprojekt: 6G-Industrieprojekte zur Erforschung von ganzheitlichen Systemen und Teiltechnologien für den Mobilfunk der 6. Generation (6G-ANNA)

Teilvorhaben: Relevanz-basiertes cell-free massive MIMO und in-X-Interferenzschätzung (ReMI)

Die Motivation von 6G-ANNA war es, einen sichtbaren Beitrag Deutschlands zur internationalen 6G Entwicklung zu leisten. Dieser Beitrag sollte von übergeordneten gesellschaftlichen Prioritäten, wie Nachhaltigkeit und klimapolitischen Zielen, der Sicherstellung der digitalen Souveränität deutscher Bürgerinnen und Bürger und der Förderung technologischer Souveränität Deutschlands und Europas geprägt sein. Hierbei sollte sichergestellt werden, dass die deutschen Nutzer, vor allem durch die frühe Integration der für Deutschland wichtigen Industrien, im Mittelpunkt stehen und die Wettbewerbsfähigkeit durch Kompetenzaufbau und Netzwerkbildung gestärkt wird. 6G-ANNA hatte den Anspruch, einen ganzheitlichen Systemansatz für 6G Mobilfunksysteme zu liefern. Es sollte ein übergreifendes Verständnis entwickelt werden, wie zukünftige Anforderungen auf der Anwenderseite mit technologischen Neuerungen adressiert werden können. Dieses tiefe Verständnis der Anforderungen, die Erforschung der technologischen Konzepte sowie die vorwettbewerbliche Konsensbildung bilden die Grundlage dafür, dass deutsche und europäische Unternehmen in der späteren Standardisierung und Markteinführung eine führende Rolle einnehmen können

Dieses Teilvorhaben widmete sich zwei Themen: verteile Interferenzschätzung in 6G Wide Area Networks mit in-X-Netzwerken und der Relevanz-basierten Signalverarbeitung für cell-free massive MIMO. Bei der zentralen Schätzung von Interferenz ist die Herausforderung, dass 6G Netzwerke aus komplexen Subnetzwerken bestehen können, die nicht vollständig zentral gesteuert werden. Die Erfassung der Interferenzsituation über eine Mobilfunkzelle muss verteilt durch Endgeräte erfolgen und an die Basisstation gemeldet werden. Fokus dieses Teilvorhabens waren daher KI- bzw. Machine Learning-Methoden für den Entwurf, das Erlernen und die Anwendung eines verteilten Interferenzschätzverfahrens. Für die Uplink-Signalverarbeitung von cell-free massive MIMO ist die Herausforderung, die Übertragung und Verarbeitung der Rohdaten mehrerer Radio Units, die dynamisch die Zelle für einen Nutzer bilden und zentral verarbeitet werden. Im Fokus standen daher Relevanz-basierte Informationsverarbeitungsalgorithmen, um eine optimale Ende-zu-Ende Übertragung für variable und verschiedene Fronthauls (z.B. wireless Fronhaul, bewegliche Radio Unit) zu erzielen. Dabei kamen hybride Verfahren zum Einsatz, die klassische, nachrichten-technische Verfahren mit Verfahren des maschinellen Lernens kombinieren, um so flexible Trade-offs zwischen Komplexität und Leistungsfähigkeit zu erreichen.

Das Konzept von Massive Multiple-Input Multiple-Output (mMIMO) hat sich als eines der wesentlichen Verfahren zur Befriedigung der ständig wachsenden Nachfrage nach höheren Datenraten und verbesserter Spektraleffizienz in modernen drahtlosen Kommunikationssystemen erwiesen. mMIMO nutzt räumliche Multiplexing- und Beamforming-Techniken mit einer großen Anzahl von (verteilten) Antennen, wodurch die gleichzeitige Versorgung mehrerer Nutzer über dieselben Zeit-Frequenz-Ressourcen ermöglicht wird. Darauf aufbauend ist Cell-Free Massive MIMO (CF-mMIMO) eine innovative und skalierbare Netzwerkarchitektur, bei der nur Empfangssignale der Antennen an einer zentralen Einheit verarbeitet werden, welche für die Datenschätzung der spezifischen Nutzer relevant sind. Zunächst wurden typische Szenarien und Parametrisierungen analysiert und entsprechende Uplink-Szenarien definiert. Ausgehend von einem einzigen Nutzer wurde die Komplexität der betrachteten Szenarien sukzessiv zu mehreren Nutzern und komplexeren Annahmen wie fehlerhaften Fronthaul-Verbindungen zu den verteilen Antennen erhöht. Um die gemeinsame Verarbeitung von Nutzerdaten über mehrere, verteilte Antennen zu optimieren, wurden verschiedene Verfahren zur

gleichzeitigen Kompression und Fehlerschutzkorrektur untersucht. Auf Basis der Information Bottleneck-Methode konnten optimale Kompressoren entworfen werden, wobei ein Teil der mMIMO-Verarbeitung bereits an den verteilten Antennen durchgeführt wird. Das resultierende komplexe Designproblem für mehrere Nutzer konnte effizient durch das GEMIB-Verfahren und dessen Erweiterung FAGEMIB auch für fehlerhafte Übertragungen erfolgreich gelöst und theoretische nachgewiesen werden. In der zentralen Verarbeitungseinheit des CF-mMIMO Systems ist die aufwandsreduzierte Realisierung der Empfängerstrukturen von hoher Bedeutung. Ein vielversprechender Ansatz zur Aufwandsreduktion besteht in der Realisierung von Verarbeitungsschritten mit geringer Bitauflösung. Hierfür wurde ein angepasster Decoder sowohl für die Verarbeitung einzelner Nutzer als auch mehrere Nutzer erarbeitet und die gesamte Verarbeitungskette inklusive der Kompression gemeinsam optimiert. Eine Erweiterung für dynamische Systeme, z.B. nicht-terrestrische Kommunikationsnetze, wurde zum einem modellbasiert mit dem Multi-FAVIB-Design und zum anderen ML-basiert mit dem Deep Multi-FAVIB vorgestellt. Dabei bietet das ML-basierte Design den Vorteil, weniger Annahmen und Vorwissen zu benötigen. Umfassende numerische Studien bestätigten die Leistungsfähigkeit der erarbeiteten Verfahren und unterstreichen die Machbarkeit Relevanz-basierter Designs für eine skalierbare und adaptive Implementierung in CF-mMIMO-Systemen.

Die sechste Mobilfunkgeneration wird verschiedene Dienste und Kommunikationsnetzwerke in Form eines Netzwerks der Netzwerke integrieren. Innerhalb dieses Paradigmas bezeichnen in-X-Netzwerke verschiedene Anwendungen wie etwa in-Robot oder in-Factory. Speziell für industrielle Kommunikation sind in-Factory-Netzwerke von hohem Interesse, um sehr hohe Zuverlässigkeit und sehr geringe Latenzen zu erreichen. Allerdings werden viele solcher Kleinstnetzwerke auf begrenzter Fläche um Funkressourcen konkurrieren und somit gegenseitig Interferenz erzeugen. Weiterhin sollen auch solche Subnetzwerke integriert werden, die nicht vollständig zentral gesteuert werden können. Um diese Interferenz zu beherrschen und ggf. für ein proaktives Ressourcenmanagement zu nutzen, werden Schätzverfahren benötigt, welche hoch-komplexe und dynamische Szenarien zuverlässig vorhersagen können. Dabei muss die Erfassung der Interferenzsituation über eine Mobilfunkzelle verteilt durch Endgeräte erfolgen und an die Basisstation gemeldet werden. Hierfür wurden zunächst verschiedene Szenarien definiert, um Kanal-, Mobilitäts und Datenverkehrsmodelle möglichst genau darstellen zu können und damit eine Datenbasis für das Training von datengestützten Verfahren zu erzeugen. Mit Hilfe verschiedener Ansätze aus dem Maschinellen Lernen (ML) wie LSTM- und Transformer-Netzwerken wurden aus den erzeugten Daten KI-Verfahren angelernt, welche die zu erwartende Interferenz in komplexen Szenarien erfolgreich prädizieren können. Allerdings stellte sich im Verlauf der Forschungsarbeiten heraus, dass für die nachgelagerte Planung von Ressourcen eine Vorhersage des Mittelwerts ungenügend ist. Daher wurden Erweiterungen durch Anpassung der Kostenfunktionen der ML-Netzwerke zu einer Quantilsvorhersage (z.B. 95% Quantil) erarbeitet, wodurch für die erzielten Vorhersagen eine definierte Zuverlässigkeit erreicht werden kann. Mit Hilfe von Methoden der konformalen Regression, kalibrierten Statistiken und der Extreme Value Theorie konnten diese Vorhersagen weiter verfeinert und statistisch abgesichert werden, so dass eine darauf basierendem zuverlässige Ressourcenvergabe auch mit geringen Datenmengen erfolgen kann. Schließlich wurde das Verfahren noch darauf erweitert, die vollständige Interferenzverteilung zu schätzen. Über diese Arbeiten hinaus wurden verteilte Architekturen für die ML-Netzwerke erarbeitet, welche auf Basis von Federated Learning und Split-Learning sowohl die Komplexität der Vorhersage aus Nutzer und Basisstationen passgenau verteilen als auch eine Verbesserung der Vorhersage durch den Austausch von Informationen zwischen vielen Subnetzwerken erlauben. Schließlich konnten die erarbeiteten Verfahren in einem 5G-Testbed in Zusammenarbeit mit der TU Dortmund mit echten Messdaten getestet werden. Außerdem wurden im letzten Projektjahr erste in 6G-ANNA entwickelte Interferenzschätzverfahren mittels eines vorhandenen AGV-Testbeds untersucht und verifiziert. Der

Aufbau eines entsprechenden Demonstrators ist für die nahe Zukunft geplant. Diese Interferenz-Testbed wird an der Universität Bremen weiterentwickelt und im Digital Hub Industrie (DHI, https://digitalhubindustry.de/) der Öffentlichkeit zugänglich gemacht.

Das Projekt hat einerseits neue Erkenntnisse im Bereich der Grundlagenforschung hervorgebracht, andererseits aber auch praxisnahe Verfahren und Methoden in Form von konkreten Lösungsansätzen für cell-free massive MIMO und die Interferenzprädiktion in Zusammenarbeit mit NOKIA Stuttgart und der TU Dortmund. Im Zusammenhang mit der oben beschriebenen Verwertung hat das Vorhaben somit Impulse für die weitere Forschung auf dem Gebiet der 6G-Forschung generiert. Die Ergebnisse des Vorhabens dienen als Basis für weitergehende nationale und internationale Forschungsvorhaben, die sowohl die Grundlagenforschung als auch die angewandte Forschung adressieren. Konkret werden die in 6G-ANNA erarbeiteten Ideen und Technologien bereits in dem BMFTR-geförderten Forschungsprojekt Open6GHub+ und zukünftig in 6G-Coverage weitergeführt.

Schlussbericht

Relevanz-basiertes cell-free massive MIMO und in-X-Interferenzschätzung (ReMI)

Im Verbundprojekt

Berichtszeitraum:	**01.01.2022 - 30.06.2025**
Laufzeit des Vorhabens:	**01.07.2022 - 30.06.2025**
Förderkennzeichen:	**16KISK109**

Projektpartner

Universität Bremen
Bibliothekstraße 1
D-28359 Bremen

Institut für Telekommunikation und Hochfrequenztechnik (ITH)
Arbeitsbereich Nachrichtentechnik
Otto-Hahn-Allee NW 1
28359 Bremen

Ansprechpartner

Prof. Dr.-Ing. Armin Dekorsy	dekorsy@ant.uni-bremen.de	0421 218 62400
Dr.-Ing. Dirk Wübben	wuebben@ant.uni-bremen.de	0421 218 62385
Dr.-Ing. Carsten Bockelmann	bockelmann@ant.uni-bremen.de	0421 218 62386

Gefördert durch:
Bundesministerium für Forschung, Technologi und Raumfahrt

Autoren

Prof. Dr.-Ing. Armin Dekorsy	dekorsy@ant.uni-bremen.de
Dr.-Ing. Dirk Wübben	wuebben@ant.uni-bremen.de
Dr.-Ing. Carsten Bockelmann	bockelmann@ant.uni-bremen.de
Dr. Alizrea Danaee	danaee@ant.uni-bremen.de
MSc. Pramesh Gautam	gautam@ant.uni-bremen.de

Inhaltsverzeichnis

1 Zielsetzung des Vorhabens 3
2 Voraussetzung der Durchführung 3
3 Planung und Ablauf des Vorhabens 4
4 Wissenschaftlich-technische Ausgangslage 4
5 Zusammenarbeit mit anderen Stellen 6
6 Wissenschaftlich-technische Ergebnisse und andere wesentliche Ereignisse 7
6.1 AP0: Projektmanagement & Scientific Exchange 7
6.2 AP1: Sustainable and Secure Architecture & System Design 7
6.3 AP2: 6G Access 7
6.4 AP3: Network of Networks 16
6.5 AP4: Automation & Simplification 16
6.6 AP5: Testbeds & Proof of Concepts 16
7 Erläuterungen zu den wichtigsten Positionen des zahlenmäßigen Nachweises 16
8 Notwendigkeit und Angemessenheit der geleisteten Arbeit 17
9 Voraussichtlicher Nutzen und Verwertbarkeit 17
10 Berücksichtigung anderer FE-Ergebnisse 18
11 Erfolgte und geplante Verwertung 18
11.1 Veröffentlichungen 18
11.2 Studentische Arbeiten 20
11.3 Fachvorträge 20
11.4 Wissenschaftskommunikation 20

1 Zielsetzung des Vorhabens

Die Motivation von 6G-ANNA war es, einen sichtbaren Beitrag Deutschlands zur internationalen 6G Entwicklung zu leisten. Dieser Beitrag sollte von übergeordneten gesellschaftlichen Prioritäten, wie Nachhaltigkeit und klimapolitischen Zielen, der Sicherstellung der digitalen Souveränität deutscher Bürgerinnen und Bürger und der Förderung technologischer Souveränität Deutschlands und Europas geprägt sein. Hierbei sollte sichergestellt werden, dass die deutschen Nutzer, vor allem durch die frühe Integration der für Deutschland wichtigen Industrien, im Mittelpunkt stehen und die Wettbewerbsfähigkeit durch Kompetenzaufbau und Netzwerkbildung gestärkt wird. 6G-ANNA hatte den Anspruch, einen ganzheitlichen Systemansatz für 6G Mobilfunksysteme zu liefern. Es sollte ein übergreifendes Verständnis entwickelt werden, wie zukünftige Anforderungen auf der Anwenderseite mit technologischen Neuerungen adressiert werden können. Dieses tiefe Verständnis der Anforderungen, die Erforschung der technologischen Konzepte sowie die vorwettbewerbliche Konsensbildung bilden die Grundlage dafür, dass deutsche und europäische Unternehmen in der späteren Standardisierung und Markteinführung eine führende Rolle einnehmen können.

Dieses Teilvorhaben widmete sich zwei Themen: verteile Interferenzschätzung in 6G Wide Area Networks mit in-X-Netzwerken und der Relevanz-basierten Signalverarbeitung für cell-free massive MIMO. Bei der zentralen Schätzung von Interferenz ist die Herausforderung, dass 6G Netzwerke aus komplexen Subnetzwerken bestehen können, die nicht vollständig zentral gesteuert werden. Die Erfassung der Interferenzsituation über eine Mobilfunkzelle muss verteilt durch Endgeräte erfolgen und an die Basisstation gemeldet werden. Fokus dieses Teilvorhabens waren daher KI- bzw. Machine Learning-Methoden für den Entwurf, das Erlernen und die Anwendung eines verteilten Interferenzschätzverfahrens. Für die Uplink-Signalverarbeitung von cell-free massive MIMO ist die Herausforderung, die Übertragung und Verarbeitung der Rohdaten mehrere Radio Units, die dynamisch die Zelle für einen Nutzer bilden und zentral verarbeitet werden. Im Fokus standen daher Relevanz-basierte Informationsverarbeitungsalgorithmen, um eine optimale Ende-zu-Ende Übertragung für variable und verschiedene Fronthauls (z.B. wireless Fronhaul, bewegliche Radio Unit) zu erzielen. Dabei kamen hybride Verfahren zum Einsatz, die klassische, nachrichten-technische Verfahren mit Verfahren des maschinellen Lernens kombinieren, um so flexible Trade-offs zwischen Komplexität und Leistungsfähigkeit zu erreichen.

2 Voraussetzung der Durchführung

Der Arbeitsbereich Nachrichtentechnik (ANT) der Universität Bremen besaß bereits vor Beginn von 6G-ANNA langjährige Expertise in der Erforschung von nachrichtentechnischen Systemen, Algorithmen und Signalverarbeitungsverfahren. Durch die Teilnahme an DFG-Programmen, einschlägigen EU-Projekten (iJOIN, METIS, Fantastic-5G), ZDKI (HiFlecs, BZKI, TACNET 4.0) und Industrieprojekten fanden die Ergebnisse der Grundlagenforschung eine direkte Verwertung in der Entwicklung und Standardisierung von Mobilfunksystemen (5G) und Sensornetzen. Das Institut bzw. Personen des Instituts sind Mitglieder in der VDE/ITG-Fachgruppe 5, dem Fachausschuss KT1, der EU-Plattform Net!Works, NGMN Alicance, 6G-IA, 5G-ACIA, ZVEI und VDMA.

Die langjährige Expertise in der Erforschung nachrichtentechnischer Systeme brachte das ANT in seinen Aktivitäten in zwei Themenbereiche ein: die verteilte Interferenzschätzung in 6GWANs mit in-X-Subnetzwerken und eine Relevant-basierte Signalverarbeitung für cell-free massive MIMO. In vorangegangenen Forschungsaktivitäten wurden hierzu bereits in den Projekten FunKI und TACNET 4.0 Vorarbeiten zum ML-basierten Entwurf von Informationsverarbeitungskonzepten bei geringer Bitbreite unter Anwendung des Information Bottleneck Frameworks, dem zugehörigen ML-basierten Entwurf von Clusteringverfahren (z.B. zum Entwurf von Quantisierern) und die Erforschung von Neuronalen Netzen für die Entzerrung von massive MIMO und Federated Learning erarbeitet.

Eine ausführliche Liste aller Publikationen und Projekte, die am ANT entstanden sind und weiterer Vorarbeiten können auf der Webseite des Instituts www.ant.uni-bremen.de eingesehen werden.

3 Planung und Ablauf des Vorhabens

Methodisch wurde 6G-ANNA als eine Kombination aus Bottom-Up- und Top-Down-Prozessen organisiert. Im Bottom-Up-Prozess wurden die Anforderungen, Konzepte und Teillösungen der 6G-ANNA-Partner, der 6G-Forschungshubs und weiterer 6G-Industrieprojekte unter Nutzung der 6G-Plattform sowie unter Berücksichtigung der zeitlichen Umsetzbarkeit und Kompatibilität integriert. Im Top-Down-Prozess wurden übergeordnete Randbedingungen, die sich aus dem internationalen Kontext und der Standardisierung ergaben, berücksichtigt und dabei auch die Durchsetzbarkeit bewertet. Im AP1 bildete ein Team erfahrener Experten ein „System Engineering"-Gremium, das eine 6G-Systemlösung entwickelte, welche die Breite der verfügbaren Lösungsansätze mit den Randbedingungen der Durchsetzbarkeit in einem Ansatz kombinierte.

Die technisch-wissenschaftlichen Schwerpunkte von 6G-ANNA wurden so gewählt, dass die besonderen Anforderungen der deutschen Gesellschaft und Industrie ein wesentliches Auswahlkriterium darstellten – insbesondere, wenn diese in anderen internationalen 6G-Aktivitäten bisher nur unzureichend berücksichtigt worden waren. Zusätzlich wurden typische übergreifende Systemdesignaufgaben auf Basis der verfügbaren Lösungsansätze bearbeitet (Disaggregation / Open RAN, Schnittstellen, Gesamtdesign RAN, RAN Numerologie, C/U Split, Konvergenz von Funkzugangs- und Kernnetz, Spektrumsnutzung etc.). Das Projekt erarbeitete neue Ansätze und Methoden in den Bereichen Security und Privacy, Netz von Netzen, D2D, Joint Communication & Sensing (JCAS) als Schlüsseltechnologie für digitale Zwillinge, KI/ML-Nutzung und Netzautomatisierung. Die Detailthemen wurden nach fachlichen Kriterien in technische Arbeitspakete (AP2, 3 und 4) sowie das Gesamtdesign-Arbeitspaket mit der System Engineering-Verantwortung (AP1) und den dazugehörigen Proof of Concepts (AP5) aufgeteilt.

Die Arbeiten der Universität Bremen konzentrierten sich auf das Arbeitspaket AP2 mit folgendem Fokus:

Für die Interferenzschätzung in AP 2.3 wurde ausgehend von synthetischen Szenarien wachsender Komplexität zunächst die Datengenerierung erarbeitet. Darauf aufbauend konnte das Trainingsproblem sowie die Architektur des verwendeten neuronalen Netzes erarbeitet und gelöst werden. Diese initiale Lösung musste dann auf viele Nutzer verteilt werden, wobei die Aufteilung des neuronalen Netzes auf Basisstation und Nutzer die Komplexität der Verarbeitung beim Nutzer bestimmte. Für ein Federated-Learning-Problem nehmen die Nutzer lokale Anpassungen am erlernten, neuronalen Netz vor und übertragen diese zur Basisstation. Abschließend wurde untersucht, wie ausgehend von einem vorgelernten Modell eine kontinuierliche Verbesserung durch Transferlernen erzielt werden kann.

Für die relevant-basierte Signalverarbeitung in AP2.2 wurde zunächst ein geeignetes cell-free massive MIMO (CF-mMIMO) Modell entwickelt und theoretisch analysiert. Dabei wurden verschiedene Vorverarbeitungsalgorithmen an der Radio Unit untersucht, um eine erste Datenreduktion zu erzielen. Auf dieser Basis konnten dann Machine-Learning-gestützte Vorverarbeitungs- beziehungsweise Kompressionsverfahren erarbeitet werden, welche die RU-Daten an den RU-spezifischen Fronthaul anpassten, wobei die Generalisierbarkeit und das Transfer-Learning für veränderliche Links untersucht wurden. Abschließend wurde die zentrale Verarbeitung an diese relevant-basierte Kompression angepasst und gemeinsam mit allen Schritten optimiert.

4 Wissenschaftlich-technische Ausgangslage

In dem Projekt 6G-ANNA hat die Universität Bremen zwei Themenbereiche bearbeiten: die verteilte Interferenzschätzung in 6GWANs mit in-X-Subnetzwerken und eine Relevant-basierte Signalverarbeitung für cell-free massive MIMO. Mit der Information Bottleneck Methode (IBM) hat der Antragsteller bereits umfassend an einer Methode für die Relevanz-basierte Signalverarbeitung in nachrichtentechnischen Systemen gearbeitet. Wesentliches Ziel war dabei umsetzbare Verfahren zu entwickeln, welche langfristig in Mobilfunksystemen eingesetzt werden können. In [Wüb17, HWD17, HWD18a] wurde das allgemeine Framework analysiert und verschiedene algorithmische Ansätze evaluiert und analysiert. Interessanterweise lassen sich bei dem IBM-basierten Quantisiererentwurf auch statistische Eigenschaften fehlerbehafteter Fronthauls mitberücksichtigen. Für das zugrundeliegende Problem der gemeinsamen Quellen- und Kanalcodierung konnten sowohl Trade-Offs zwischen der Fronthaul-Rate und der Ende-zu-Ende Datenrate numerisch evaluiert und Kri-

terien zur Parameterauswahl abgeleitet [HWD18b]. Eine vollständige Verallgemeinerung des IBM-Framework auf das Problem der gemeinsamen Quellen- und Kanalcodierung wurde in [HMW20] präsentiert. Zur Vermeidung von Initialisierungsproblemen bekannter Ansätze wurde auf Basis des „Affinity Propagation Clusterings“ ein Graphen-basierter Message-Passing Algorithmus für die Quellencodierung (fehler-freier Fronthaul) und für die gemeinsame Quellen- und Kanalcodierung (fehlerbehafteter Fronthaul) [HWD18d] entwickelt. Anhand von numerischen Evaluierungen konnte die sehr gute Leistungsfähigkeit bezüglich des Trade-Off zwischen FH-Datenrate und Ende-zu-Ende Datenrate gezeigt werden.

Der gemeinsame Entwurf von Quantisierern in einem verteilten Kommunikationsszenario wurde sowohl für fehlerfrei wie auch für fehlerbehaftete Fronthauls eingehend in [HWD19, HWD20a] analysiert. Mit dem M-FAVIB Algorithmus liegt schließlich die Verallgemeinerung des IBM-Frameworks für verteilte Systeme mit fehlerbehafteten FHs vor [HWD21]. Auch zur theoretischen Beschreibung des Problems der semantischen Kommunikation wurde IBM mit kontinuierlichen Variablen vorgeschlagen, wobei sich eine drastische Bandbreiteeinsparung mit einem DNN-basierten gegenüber einem klassischen Kommunikationsdesign zeigt [BBD22].

Für die effiziente Entzerrung von MIMO-Systemen und insbesondere massive MIMO-Systeme wurde zwei verschiedene Hybrid-Ansätze erarbeitet. In [HWD20b] wurde basierend auf einem modellbasierten Ansatz Update-Beziehungen für einen iterativen Entzerrer abgeleitet, welche dann mittels ML trainiert wurden. In [BBD21] wurde das MAP-Problem durch kontinuierliche stochastische Relaxation der diskreten Sendesymbole in ein differenzier-bares Problem überführt, welches sich dann effizient mit Stochastic Gradient Descent trainieren lässt. Beide Ansätze zeigen insbesondere bei hoher Antennenanzahl eine sehr gute Leistungsfähigkeit. Darüber hat der Arbeitsbereich Reinforcement Learning-Ansätze auf Scheduling-Probleme für kritische Kommunikation angepasst. Bereits die dynamische Auswahl von bekannten, einfachen Scheduler kann hierbei beliebige Trade-Offs zwischen Anforderungen wie Durchsatz, Latenz und Paketverlusten erzielen [GBB22a]. Fasst man hingegen die Schedulingaufgabe vollständig in der Form eine Actor-Critic-Modells und lernt die Entscheidungen für ein vorgegebenes Szenario an-hand des User-Feedbacks, so kann eine weitere Verbesserung der Gesamterfahrung aller Nutzer erreicht werden [GBB22b].

Im Bereich der verteilten Signalverarbeitung wurden im Arbeitsbereich vielfältige Problemstellungen mit klassischen Verfahren der verteilten Signalverarbeitung erforscht [WSS20]. So wurden unter anderem Diffusionsfeldschätzungen mit Kalmanfiltern in einem Hierarchischen Sensornetzwerk erarbeitet und Erweiterungen verteilter Kalmanfilter über Faktorgraphen vorgeschlagen [WD20] Für das verteilte Erlernen Modellen des maschinellen Lernens wurde im Themegebiet des Federated Learning für Satellitennetzwerke geforscht. So wurden Modelle, beispielweise zur Bildverarbeitung, in Satellitenkonstellationen mittels Federated Learning „onboard“ auf kleinen Satelliten eingesetzt, um den Datenaustausch mit den Bodenstationen zu reduzieren. Darüber hinaus wurden Standardverfahren von Federated Learning im Hinblick auf ein asynchrones Lernen erweitert [RMA22a, RMA22b].

Referenzen	
[BBD21]	E. Beck, C. Bockelmann, A. Dekorsy: CMDNet: Learning a Probabilistic Relaxation of Discrete Variables for Soft Detection with Low Complexity, IEEE Transactions on Communications, September 2021.
[BBD22]	E. Beck, C. Bockelmann, A. Dekorsy: Semantic Communication: An Information Bottleneck View. In preparation for journal submission, 2022.
[HMW20]	S. Hassanpour, T. Monsees, D. Wübben, A. Dekorsy: Forward-Aware Information Bottleneck-Based Vector Quantization for Noisy Channels, IEEE Transactions on Communications, Vol. 68, No. 7911-7926, Dezember 2020.
[HWD17]	S. Hassanpour, D. Wübben, A. Dekorsy: Overview and Investigation of Algorithms for the Information Bottleneck Method, 11th Int. ITG Conference on Systems, Communications and Coding (SCC), 2017.
[HWD18a]	S. Hassanpour, D. Wübben und A. Dekorsy: On the Equivalence of Two Information Bottleneck-Based Routines Devised for Joint Source-Channel Coding, 25th International Conference on Telecommunication (ICT 2018), Saint-Malo, Frankreich, Juni 2018.

[HWD18b]	S. Hassanpour, D. Wübben, A. Dekorsy: A Graph-Based Message Passing Approach for Joint Source-Channel Coding via IB Principle, 10th Int. Symposium on Turbo Codes & Iterative Information Processing (ISTC), 2018.
[HWD19]	S. Hassanpour, D. Wübben, A. Dekorsy: A Novel Approach to Distributed Quantization via Multivariate Information Bottleneck Method, IEEE Global Communications Conference (GLOBECOM), 2019.
[HWD20a]	S. Hassanpour, D. Wübben, A. Dekorsy: Generalized Distributed Information Bottleneck for Fronthaul Rate Reduction at the Cloud-RANs Uplink, IEEE Global Communications Conference (GLOBECOM 2020), Dezember 2020.
[HWD20b]	M. Hummert, D. Wübben, A. Dekorsy: DeEQ: Deep Equalization for Large MIMO Systems, ITG Workshop on Smart Antennas (WSA), 2020.
[HWD21]	S. Hassanpour, D. Wübben und A. Dekorsy: Forward-Aware Information Bottleneck-Based Vector Quantization: Multiterminal Extensions for Parallel and Successive Retrieval, IEEE Transactions on Communications, Vol. 69, No. 10, pp. 6633-6646, Oktober 2021.
[GBB22a]	S. Gracla, E. Beck, C. Bockelmann, A. Dekorsy: Deep Reinforcement Model Selection for Communications Resource Allocation in On-Site Medical Care, IEEE WCNC 2022, 13. April 2022
[GBB22b]	S. Gracla, E. Beck, C. Bockelmann, A. Dekorsy: Learning Resource Scheduling with High Priority Users using Deep Deterministic Policy Gradients, IEEE ICC 2022, 16. Mai 2022
[RMA22a]	N. Razmi, B. Matthiesen, A. Dekorsy, P. Popovski: On-Board Federated Learning for Dense LEO Constellations, IEEE International Conference on Communications (ICC), Mai 2022.
[RMA22b]	N. Razmi, B. Matthiesen, A. Dekorsy, P. Popovski: Ground-Assisted Federated Learning in LEO Satellite Constellations, IEEE Wireless Communications Letters, Vol. 11, No. 4, pp. 717-721, April 2022.
[WD20]	S. Wang, A. Dekorsy: A Factor Graph-Based Distributed Consensus Kalman Filter, IEEE Signal Processing Letters, November 2020.
[WSS20]	S. Wang, B.-S. Shin, D. Shutin, A. Dekorsy: Diffusion Field Estimation Using Decentralized Kernel Kalman Filter with Parameter Learning over Hierarchical Sensor Networks, IEEE International Workshop on Machine Learning for Signal Processing, Espoo, Finnland, September 2020.
[Wüb17]	D. Wübben: The Information Bottleneck Method: Fundamental Idea and Algorithmic Implementations, 10th Asia-Europe Workshop on Concepts in Information Theory (AEW), 2017.

5 Zusammenarbeit mit anderen Stellen

Das Verbundvorhaben 6G-ANNA verfolgte ein sehr breites Ziel der Erforschung von Technologien und Architekturen für die 6. Mobilfunkgenerationen. Die Universität Bremen hat hierzu konzentriert im Arbeitspaket AP2 beigetragen und in diesem einen intensiven Austausch mit dem beteiligten Projektpartner gepflegt.

Besonders hervorzuheben ist die Zusammenarbeit mit NOKIA Stuttgart zum Thema der Interferenzschätzung in in-X-Subnetzwerken. Im Rahmen eines regelmäßigen Austausches sind mehrere gemeinsame Publikationen [GSB25a, GSB25b] entstanden. Darüber hinaus wurde dieses Thema auch mit der Gruppe von Prof. Wietfeld der TU Dortmund im Hinblick auf eine praktische Verifikation weiterfolgt und in einer gemeinsamen Publikation [GBD25c] gezeigt, dass die entwickelten Verfahren mit verfügbaren 5G Equipment vielversprechende Ergebnisse erzielen.

6 Wissenschaftlich-technische Ergebnisse und andere wesentliche Ereignisse

6.1 AP0: Projektmanagement & Scientific Exchange

Die Universität Bremen war nicht an den Arbeiten in AP0 beteiligt.

6.2 AP1: Sustainable and Secure Architecture & System Design

Die Universität Bremen war nicht an den Arbeiten in AP1 beteiligt.

6.3 AP2: 6G Access

Die Universität Bremen hat an regelmäßigen Arbeitstreffen und Telefonkonferenzen des Arbeitspakets 2, seiner Fokusgruppen im Bereich AI, physikalische Schicht und Architektur teilgenommen. Die Forschungsergebnisse wurden in den Arbeitstreffen präsentiert und in den Meilensteinberichten [6G-ANNA-D2a, 6G-ANNA-D2b, 6G-ANNA-D2c] ausführlich beleuchtet. Darüber hinaus wurde ein regelmäßiger Austausch mit Nokia durchgeführt, um die Kollaboration im Themenbereich der Interferenzschätzung und -prädiktion zu intensivieren.

Basierend auf diesen Diskussionen wurde zu drei Übersichtsarbeiten von 6G-ANNA beigetragen, in denen ein Überblick über die Arbeiten der Universität in den entsprechenden Abschnitten hervorgehoben wurde. In [GBD25a] wurde ein Überblick über verteiltes maschinelles Lernen gegeben, mit besonderem Fokus auf Link-Adaptation und Aspekte des Interferenzmanagements. In [GDB25a, GDB25b] wurde ein Überblick über Signalverarbeitungsansätze für die Relevanz-basierte Informationsverarbeitung für CF-mMIMO-Systeme (Task 2.2 dieses Berichts) und das Interferenzmanagement (Task 2.3 dieses Berichts) gegeben und dabei betont, wie explizites Wissen über vorhergesagte Interferenzen und die Integration von Techniken des maschinellen Lernens die Leistung verbessern können.

Außerdem hat die Universität Bremen an den Diskussionen der Fachgruppe 4 „Roadmap“ der 6G-Platform teilgenommen und zu dem White Paper über die deutsche Perspektive zu 6G mit Anwendungsfällen, technischen Bausteine und Anforderungen aus 6G-ANNA beigetragen [WRG24].

Task 2.1 Access Overview, Interworking & Spectrum Usage

T2.1-UB-1 Eingaben aus dem AP4 von Open 6G Hub

Universität Bremen hat im Projektverlauf ausgewählte Ergebnisse des Open6GHub AP4 „Beyond Cellular“ in die Definition von Anforderungen und Basisannahmen mit Fokus auf organische, infrastrukturlose 3D-Netzwerke in die Diskussion von 6G-ANNA eingebracht.

Task 2.2 Flexible, Secure & Harmonized PHY

T2.2-UB-1 Definition und Analyse von cell-free massive MIMO Szenarien

Das Konzept von Massive Multiple-Input Multiple-Output (mMIMO) hat sich als eines der wesentlichen Verfahren zur Befriedigung der ständig wachsenden Nachfrage nach höheren Datenraten und verbesserter Spektraleffizienz in modernen drahtlosen Kommunikationssystemen erwiesen. mMIMO nutzt räumliche Multiplexing- und Beamforming-Techniken mit einer großen Anzahl von Antennen an den Basisstationen, wodurch die gleichzeitige Versorgung mehrerer Nutzer über dieselben Zeit-Frequenz-Ressourcen ermöglicht wird. Dieser Ansatz verbessert sowohl die Spektral- als auch die Energieeffizienz erheblich und positioniert mMIMO als eine Schlüsseltechnologie für drahtlose Netzwerke der nächsten Generation. Darauf aufbauend ist cell-free Massive MIMO (CF-mMIMO) in letzter Zeit als innovative Netzwerkarchitektur in den Fokus gerückt. Seine Fähigkeit, Zellgrenzen zu beseitigen und eine flächendeckende Versorgung zu bieten, unterstützt die Vision von 6G-Netzen bezüglich nahtloser Konnektivität, ultra-zuverlässige Kommunikation mit geringer Latenz und maschineller Kommunikation in verschiedenen Umgebungen. Im Zuge der Weiterentwicklung von 6G kann CF-mMIMO eine Schlüsselrolle bei der Erfüllung der Anforderungen zukünftiger drahtloser Systeme an eine hochdichte, energieeffiziente Kommunikation mit geringer Latenz spielen.

a) CF-mMIMO Szenario b) Austausch von Beobachtungen zur CPU

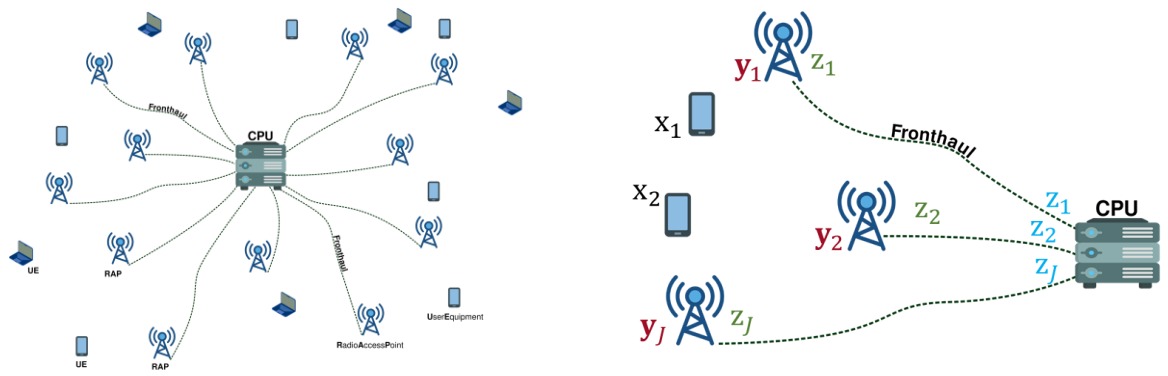

Abbildung 1: a) CF-mMIMO Szenario mit einer Vielzahl verteilter RAPs zur Versorgung der Nutzer und b) Übertragung der komprimierten Beobachtungen der RAPs zur zentralen CPUS über ratenbeschränkte Fronthauls

Im Gegensatz zu herkömmlichen Mobilfunksystemen verzichtet CF-mMIMO auf das Konzept von Zellen und Zellgrenzen indem die Nutzer gleichzeitig von allen verfügbaren Funkzugangspunkten (Radio Access Points, RAPs) bedient werden können (siehe Abbildung 1.a). Diese RAPs sind über ein Fronthaul-Netzwerk mit einer Zentraleinheit (Central Processing Unit, CPU) verbunden in dem z.B. im Uplink die Schätzung der Nutzersignal erfolgt. Das herkömmliche CF-mMIMO-Modell, bei dem jedes RAP Signale für alle Nutzer verarbeitet und überträgt, stößt jedoch aufgrund des Anstiegs der Rechenkomplexität und der Fronthaul-Raten mit zunehmender Nutzerzahl an Skalierbarkeitsgrenzen. Im Gegensatz verfolgt das benutzerzentrierte CF-mMIMO eine skalierbarere Lösung, bei der nur Empfangssignale der RAPs an die CPU weitergeleitet werden, welche für die Datenschätzung der spezifischen Nutzer relevant sind. Nachfolgend bezeichnet CF-mMIMO immer das benutzerzentrierte System.

Auf Basis einer ausgiebigen Literaturrecherche wurden typische Szenarien und Parametrisierungen zur Analyse CF-mMIMO Systemen im Uplink analysiert und es wurden Szenarien für die weitere Analyse detaillierter definiert. Zunächst wurde ein Single-User-Szenario mit nur einem aktiven Nutzer betrachtet, der von mehreren RAPs bedient wird. Jeder RAP empfängt eine verrauschte Beobachtung des gesendeten Signals und übertragt anschließend über eine ratenbegrenzte Fronthaul-Verbindung die lokale Beobachtung in quantisierter Form an die CPU wie in Abbildung 1.b verdeutlicht [DHW24]. Weiterhin wurde ein Multi-User-Szenario mit mehreren Nutzern definiert, bei dem einige Nutzer von nur einem RAP und andere Nutzer von mehreren RAPs bedient werden. Die RAPs erhalten daher mehrere verrauschte Beobachtungen von ihren Nutzern und leiten erneute quantisierte Beobachtungen über die ratenbegrenzten Fronthaul-Links an die CPU. Zur Analyse beider Szenarien wurden zunächst einfache statistische Modelle zur Abbildung der Übertragungskanäle verwendet. Anschließend wurde das Systemmodell hinsichtlich der Berücksichtigung realer Annahmen im Mehrbenutzerfall erweitert. Vorab wurde angenommen, dass die Fronthaul-Verbindungen ratenbegrenzt, aber fehlerfrei sind. Das Modell wurde nun so erweitert, dass auch Übertragungsfehler auf den Fronthaul-Verbindungen berücksichtigt werden. Dadurch verlagert sich das Problem von der verteilten Quellencodierung zu einer verteilten gemeinsamen Quellen-Kanal-Codierung (Joint Source-Channel Coding, JSCC), bei der Komprimierung und Fehlerschutz gemeinsam behandelt werden [DHW25].

Für die simulative Analyse wurde eine Simulationsumgebung für CF-mMIMO Systeme mit verteilten Nutzern, mehreren RAPs und zentraler CPU erstellt. Für die Access-Kanäle wurden unterschiedliche Kanalmodell berücksichtigt und sowohl fehlerfreie als auch fehlerbhaftete Fronthaul-Verbindungen mit Ratenbeschränkungen implementiert. Gerade die Berücksichtigung von verschiedenen Fronthaul-Modellen ermöglichte es, ein tieferes Verständnis bei dem Entwurf von Komprimierungsstrategien an den RAPs zur Extraktion der relevantesten Informationen in praktischen CF-mMIMO-Systemen in den nachfolgenden Tasks zu gewinnen.

T2.2-UB-2 Identifikation und Charakterisierung der Relevanten Information

Im Fokus der Arbeiten stand die Verarbeitung der relevanten Information zur Bestimmung der Nutzerdaten in der CPU. Hierzu wurde die Information Bottleneck (IB) Methode zum Entwurf der Quantisierer in den RAPs zu Grund gelegt, die den Kompromiss zwischen zwei Transinformationen formuliert. Der erste Term ist die relevante Information, welche den Informationsgehalt quantifiziert, während der zweite Term die Komprimierungsrate ist und somit die Kompaktheit quantifiziert. Die Nutzung der IB-Methode zum Entwurf der Quantisierer verfolgt dabei das Ziel, ein ausgewogenes Verhältnis zwischen der Beibehaltung möglichst vieler nützlicher Informationen und der Einhaltung der individuellen Kapazitätsgrenzen der Fronthaul-Verbindungen zu finden. Dies geschieht durch die Maximierung der Transinformation zwischen dem Quellsignal und dem komprimierten Empfangssignal an der CPU bei gleichzeitiger Begrenzung der Transinformation zwischen den ursprünglichen RAP-Beobachtungen und ihren komprimierten Versionen (d.h. der Kompressionsrate).

Für die Erzeugung der ratenbegrenzten Sendesignal je RAP wird dabei angenommen, dass im RAP zunächst eine räumliche und zeitliche Entzerrung so vorgenommen wird, dass die gegenseitige Interferenz beseitig wird und an den RAPs für die relevanten Nutzer unabhängig angenommene Beobachtungen vorliegen. Diese RAP-Beobachtungen werden anschließend quantisiert und dann als ratenbegrenzte lokale Beobachtungen an die CPU übermittelt. Es zeigt sich, dass der äquivalente Übertragungskanal zwischen Quellsignal und dem zugehörigen Ausgang des MIMO-Entzerrers als AWGN-Kanal modelliert werden kann.

Für das Single-User-Szenario, in dem nur ein gemeinsamer Nutzer von mehreren RAPs bedient wird, entsprechen die relevanten Informationen den Transinformationen zwischen den an der CPU ankommenden Nachrichten der RAPs und dem Quellsignal. Für den allgemeinen Multi-User-Fall mit mehreren Nutzern, wurde zunächst die Transinformation zwischen dem Quellsignal jedes Nutzers und den empfangenen Signalen an der CPU berechnet. Anschließend wurde die relevanten Informationen als Summe aller im vorherigen Schritt berechneten Terme der Transinformation je Nutzer berechnet [HDW24].

Im allgemeinen Szenario mit fehleranfälligen Fronthaul-Verbindungen besteht ein signifikanter Unterschied zum fehlerfreien Szenario: Die an der CPU ankommenden Signale entsprechen nicht mehr den exakten Ausgangssignalen der Kompressoren. Stattdessen können sie durch Übertragungsfehler verfälscht werden. Trotzdem lassen sich IB-basierte Methoden weiterhin verwendet, um die Kompressoren so zu gestalten, dass die gesamte gegenseitige Information zwischen den ursprünglichen Quellsignalen und den möglicherweise fehlerhaften Signalen an der CPU maximiert wird [HDW25]. Diese Anpassung macht das Designproblem anspruchsvoller, aber auch realistischer. Sie ermöglicht es, zu untersuchen, wie viele nützliche Informationen unter praktischen Bedingungen an der CPU extrahiert werden kann.

T2.2-UB-3 Erforschung hybrider Kompressionsverfahren für statische Fronthauls

Für das Single-User- und das Multi-User-Szenario wurden die Quantisier für die verteilten RAPs entwickelt. Hierbei wurde die Übertragung von den RAPs zur CPU sowohl über fehlerfreie als auch fehlerbehaftete Fronthaul-Links mit Ratenbeschränkung im Entwurf der Quantisierer berücksichtigt. Dabei wurde ein gemeinsamer Entwurf der lokalen Quantisierer vorgenommen, bei dem aber jeder RAP jeweils nur seine lokalen Beobachtungen ohne Kenntnis der anderen Empfangssignale komprimiert. Das Ziel besteht darin, die relevanten Informationen zu maximieren, also die gesamte Transinformation zwischen den an der CPU empfangenen Nachrichten und den Quellsignalen der Nutzer. Wie in T2.2-UB-1 erläutert wurde für den lokalen Entwurf der IB-basierten Kompressoren angenommen, dass in den RAPs bereits ein Teil der Empfangssignalverarbeitung durchgeführt wurde (z.B. räumliche und zeitliche Entzerrung) [DHW25]. Dementsprechend lassen sich die äquivalenten Kanäle von den Sendern bis zu den Quantisiereingängen durch vergleichsweise einfache statistische Modelle beschreiben.

In [DHW24] wurde zunächst ein Single-User-Szenario mit mehreren RAPs betrachtet. Unter der Annahme von fehlerfreien Fronthauls (FHs) steht mit dem Multi-IB Verfahren [HWD20a] ein Ansatz zum gemeinsamen Entwurf der lokalen Quantisierer zur Verfügung. Der M-FAVIB Algorithmus liefert die Verallgemeinerung des IBM-Frameworks für verteilte Systeme mit fehlerbehafteten FHs [HWD21]. Die vergleichende Analyse für das CF-mMIMO System demonstriert die Vorteile des gemeinsamen Entwurfs der Quantisierer gegenüber einem unabhängigen Entwurf je RAP mit dem FAVIB [HMW20] oder dem Standardverfahren

Lloyd-Max. So wird bei gleicher FH-Datenrate eine deutliche Steigerung der e2e-Datenrate von Quelle zu CPU erreicht.

Für das allgemeine CF-mMIMO-Setup mit mehreren Nutzern und fehlerfreien Fronthaul-Verbindungen wurde zudem ein neuartiges, verteiltes (Remote-)Quellencodierungsschemata entwickelt. Hierbei wurde angenommen, dass einige Benutzer von mehreren RAPs und andere Benutzer von jeweils nur einem RAP angebunden sind. Die stationären Lösungen des resultierenden, komplexen Designproblems wurden mithilfe der Variationsrechnung hergeleitet und der iterative Algorithmus „GEneralized Multivariate IB (GEMIB)“ zur effizienten Lösung des Problems formuliert [HDW24]. Zudem konnte der entsprechende Konvergenzbeweise für einen stationären Punkt der Zielfunktion durch eine detaillierte Analyse gezeigt werden. Dabei wurde zur Formulierung des Entwurfsproblems für die gemeinsame Komprimierung die Gesamtübertragungsrate als Kriterium für den Informationsgehalt verwendet. Für die Kompaktheit wurden zudem zwei separate Konstruktionsvorschriften genutzt, die parallele und die sukzessive Rekonstruktion. Bei der parallelen Rekonstruktion der Quellsignale in der CPU werden keine Nebeninformationen angenommen. Hingegen werden bei der sukzessiven Verarbeitung die Nebeninformationen von bereits geschätzten Quellsignale in der CPU genutzt. Dies führt zu einem besseren Kompromiss zwischen Informationsgehalt und Kompaktheit, bedingt allerding eine höhere Rechenkomplexität in der CPU. Im Vergleich zu herkömmlichen Quantisierungsansätzen basierend auf der Vektorquantisierung mit K-Means lassen sich deutliche Leistungsgewinne in Bezug auf die erzielte Genauigkeit als auch in der erforderlichen Datenraten auf den FH-Links erzielen [HDW24]. In [DHW25] wurde der GEMIB-Ansatz für CF-mMIMO System mit drei Nutzern und zwei RAPs analysiert. Zwei Nutzer werden dabei von jeweils nur einem RAP und der dritte Nutzer von beiden RAPs bedient. Der entwickelte GEMIB-Ansatz erzielt erneut bei gleicher Datenrate auf den Fronthauls eine höhere Transinformation zwischen den Quellen und der CPU als die K-Means Vektorquantisierung. In dem 6G-ANNA Übersichtsbeitrag zum PHY-Layer [GDB25a] wurde gemeinsam mit den Projektpartnern das CF-mMIMO Konzept erläutert und neue entwickelte Ansatz zur verteilten Quantisierung präsentiert. Der Aspekt der daten-getriebenen Entwurfs wurde zudem in dem 6G-ANNA Übersichtsbeitrag zu AI/ML-Verfahren [GDB25b] erläutert.

Um fehleranfällige Fronthaul-Verbindungen in dem Multi-User-Szenario zu berücksichtigen wurde weiterhin ein verteiltes gemeinsames Quellen-Kanal-Codierungsschema für CF-mMIMO-Systeme entwickelt. Dieses neue Schema zielt darauf ab, die relevanten Informationen zu maximieren, die trotz FH-Übertragungsfehlern weiterhin zuverlässig von der CPU wiederhergestellt werden können. Um dieses Ziel zu unterstützen, wurde erneut eine lineare Entzerrung an den RAPs angenommen, so dass an den RAPs getrennte Benutzersignale vor der Komprimierung vorliegen. Aufbauend auf der GEMIB-Methode [HDW24] wurde mit dem Forward-Aware GEMIB (FAGEMIB) eine Verallgemeinerung entwickelt. Diese neue Methode bezieht erneut die Fronthaul-Fehlereigenschaften in den Designprozess ein und findet effizient stationäre Lösungen mithilfe der Variationsrechnung. Es wurde der mathematische Beweis erbracht, dass FAGEMIB gegen einen stationären Punkt des Designziels konvergiert [HDW25]. Wie zuvor bestand das Designziel darin, einen guten Kompromiss zwischen der Menge an relevanten Informationen und der Fronthaul-Datenrate zu finden. Für die Rekonstruktion an der CPU werden dabei erneut zwei verschiedene Strategien verwendet: die parallele Rekonstruktion und die sukzessive Rekonstruktion. Beide Strategien wurden an das fehleranfällige Setup angepasst und die Simulationsergebnisse zeigen, dass der entwickelte Ansatz im Vergleich zu Standard-Vektorquantisierungsmethoden wie K-Means signifikante Verbesserungen für die gesamten relevanten Informationen und die Fronthaul-Effizienz bietet [HDW25].

T2.2-UB-4 Design der zentralen Verarbeitungsalgorithmen für Relevanz-basierte Verarbeitung

Die in T2.2-UB-3 entwickelten Konzepte zum gemeinsamen Entwurf der Quantisierer in den verteilten RAPs bezog bereits die Verarbeitung der komprimierten Information an der CPU in Form einer parallelen bzw. sukzessiven Rekonstruktion mit ein. Somit liegen pro Nutzer bereits getrennte Informationen zur finalen Schätzung der gesendeten Daten vor [HDW24].

Da nun in der zentralen Verarbeitungseinheit die Daten einer Vielzahl von Nutzern zu schätzen ist, sind aufwandsreduzierte Realisierungen der Empfängerstrukturen von hoher Bedeutung. Ein vielversprechender Ansatz zur Aufwandsreduktion besteht in der Realisierung von Verarbeitungsschritten mit geringer

Bitauflösung. In früheren Arbeiten konnte gezeigt werden, dass bei entsprechendem Design unter Minimierung des Informationsverlusts die Zuverlässigkeit diskreter Empfänger sehr nahe an der entsprechender Referenzempfänger mit Floating-Point Implementierung bleibt. Der Entwurf der diskreten Verarbeitungsschritte erfolgt dabei erneute durch Anwendung der Information Bottleneck Methode.

Als erster Schritt wurden Relevanz-basierte Decodierungsstrategien für Single-User CF-mMIMO-Systeme untersucht. Ziel war es, eine effiziente Decodierung mit geringer Komplexität durch diskrete Knotenoperationen zu ermöglichen. Hierzu wurde der im Open6GHub entwickelte Minimum Integer Computation (MIC)-Decoder als Kernkomponente verwendet. In [Buh24] wurde eine Python-Implementierung dieses diskreten LDPC-Decoders entwickelt und Konzepte zur Berücksichtigung irregulärer Knotengrade und höherstufiger Modulationen anhand des Message Alignments analysiert. Für Link-Level-Analysen des CF-mMIMO Systems wurde der entwickelte Decoder in das Sionna-Framework integriert.

Zudem wurde ein sehr effizienter Ansatz zur gemeinsamen Verarbeitung von quantisierten Kanalbeobachtungen mit unterschiedlicher Zuverlässigkeit entwickelt. Im Gegensatz zu dem gemeinsamen Entwurf der lokalen Kompressoren zur Maximierung der e2e-Datenrate werden dabei unabhängig voneinander entworfene Quantisierer an den RAPs angenommen. Dabei weisen in einem CF-mMIMO-System die Empfangssignale an verschiedenen RAPs im Allgemeinen eine unterschiedliche Zuverlässigkeit auf. Werden diese Signale mit SNR-optimierten Quantisierern diskretisiert, so haben gleiche Quantisiererausgangswerte eine unterschiedliche Bedeutung bezüglich der gesendeten Informationen. Daher ist ein Alignment der Nachrichten erforderlich. Die Grundidee besteht nun darin, den diskreten Decoder als letzten Verarbeitungsschritt auf dem PHY-Layer als gegeben anzunehmen. Entsprechend ist das Design-SNR am Decoder-Eingang festgelegt und alle vorherigen Verarbeitungsschritte sind bezüglich dieses Design-SNRs anzupassen. Dementsprechend erfolgt bereits in den RAPs nach erfolgter Entzerrung ein herkömmliches Soft-Demapping auf Log-Likelihood-Ratios (LLRs). Dabei enthalten die LLR-Werte impliziert die Zuverlässigkeit der Beobachtungen. In einem zweiten Schritt werden dann die LLR-Werte quantisiert (entsprechend dem Design-SNR des Decoders) und über die FH-Links von den RAPs an die CPU übertragen.

T2.2-UB-5 Erforschung hybrider Kompressionsverfahren und zentraler Verarbeitung für dynamische Fronthauls

In CF-mMIMO-Systemen mit dynamischen Fronthauls kann die Qualität und damit die verfügbare Kapazität der Fronthaul-Verbindungen im Laufe der Zeit erheblich schwanken. Daher sind flexible Komprimierungsverfahren erforderlich, um sich an diese veränderten Bedingungen anzupassen. Die in T2.2-UB-3 Entwickelten Konzepte berücksichtigen dabei automatisch die unterschiedlichen FH-Kanäle der einzelnen RAPs. Nach der Betrachtung von statischen Szenarien wurden in diesem Task zudem hybride Kompressionsverfahren auf dynamische Umgebungen wie nicht-terrestrische Netzwerke (Non-Terrestrial Networks, NTNs) und CF-mMIMO-Systeme ausgeweitet. Die vorab diskutierten Konzepte erfordern Kenntnisse über die Statistiken der Zugangs- und Fronthaulkanäle zur Bestimmung der Quantisierer, so dass das Design modell-basiert erfolgt. Zum Design in dynamischen Umgebungen wurden in diesem Task hingegen daten-getrieben Entwürfe entwickelt, die eine Adaption an die aktuelle Systemeigenschaften ermöglichen.

Im Rahmen des Open6GHub wurde von der Universität Bremen das Deep-FAVIB Verfahren zum ML-basierten Entwurf von Kompressionsverfahren für ein Two-Hop-Szenario (Single-User, ein RAP) mit ratenbeschränkten, fehleranfälligen Fronthaul entwickelt. Das Verfahren ist datengesteuert und benötigt keine Modellkenntnis über den Übertragungskanal. Es wurde gezeigt, dass das entwickelte Deep FAVIB-Verfahren den (modellbasierten) SotA-Methoden [HMW20] ebenbürtig ist und somit gerade für dynamische Szenarien Vorteile bieten kann.

Für das Single-User Szenario mit fehlerfreien Fronthauls wurde in [HHW25a] ein ML-basierter Ansatz zum Entwurf der Quantisier unter Anwendung des parallelen Rekonstruktionskonzeptes präsentiert. Im Vergleich zum modell-basierten Multi-IB Verfahren [HWD20a] besteht der Vorteil, dass die Lösung mit nur einem Training erlernt wird und nicht die mehrfache Ausführung des modell-basierten Algorithmus zur Vermeidung von lokalen Optima erforderlich ist. Auch hier zeigt sich, dass der daten-getriebene Entwurf nahezu die Leistungsfähigkeit des Multi-IB Verfahren [HWD20a] erzielt.

Zudem wurde der ML-basierter Ansatz Deep M-FAVIB [HHW25b] entwickelt, um das Komprimierungsschema bei fehlerbehafteten Fronthauls direkt aus endlichen Datenproben zu erlernen, ohne dass statistische Kenntnisse der Zugangskanäle erforderlich sind. Numerische Studien bestätigten, dass Deep Multi-FAVIB der Leistung der modellbasierten M-FAVIB-Methode [HWD21] sehr nahekommt und gleichzeitig eine verbesserte Anpassungsfähigkeit an dynamische Umgebungen bietet. Dies macht es besonders attraktiv für den praktischen Einsatz in CF-mMIMO-Systemen.

T2.2-UB-6 Evaluierung und Optimierung der gesamtem Relevanz-basierten Verarbeitungskette

Die gesamte Relevanz-basierte Verarbeitungskette wurde für CF-mMIMO-Systeme evaluiert, wobei der Schwerpunkt auf der Robustheit unter dynamischen und fehleranfälligen Fronthaul-Bedingungen lag. Der Aufbau umfasste ein Deep M-FAVIB-basiertes Kompressionsverfahren an den RAPs bei dem ein Nutzer von mehreren Zwischenknoten bedient wird [HHW25a, HHW25b].

Umfassende numerische Studien bestätigten, dass die Relevanz-basierte Verarbeitung bei Training mit hohem Signal-Rausch-Verhältnis eine starke Generalisierung ermöglicht und selbst bei abnehmender Übertragungsqualität eine robuste Leistung hinsichtlich der insgesamt relevanten Informationen und der Symbolfehlerrate aufrechterhält. Dies unterstützt den Einsatz Relevanz-basierter Ansätze in Szenarien mit eingeschränkter Kontrolle über die Kanalbedingungen.

Bedeutend ist, dass das End-to-End-System eine vergleichbare Leistung wie modellbasierte Designs aufweist und dabei weniger statistische Kenntnisse erfordert. Diese Ergebnisse unterstreichen die Machbarkeit Relevanz-basierter Designs für eine skalierbare und adaptive Implementierung in CF-mMIMO-Systemen und wurden durch weitere Ergebnisse aus infrastrukturfreien Netzwerkbewertungen unterstützt.

Task 2.3 RAN Protocols & Cloud-Based Architecture

T2.3-UB-1 Szenariodefinition und Datengenerierung

Die sechste Mobilfunkgeneration wird verschiedene Dienste und Kommunikationsnetzwerke in Form eines Netzwerks der Netzwerke integrieren. Innerhalb dieses Paradigmas bezeichnen in-X-Netzwerke verschiedene Anwendungen wie etwa in-Robot oder in-Factory. Speziell für industrielle Kommunikation sind in-Factory-Netzwerke von hohem Interesse, um sehr hohe Zuverlässigkeit und sehr geringe Latenzen zu erreichen. Allerdings werden viele solcher Kleinstnetzwerke auf begrenzter Fläche um Funkressourcen konkurrieren und somit gegenseitig Interferenz erzeugen. Weiterhin sollen auch solche Subnetzwerke integriert werden, die nicht vollständig zentral gesteuert werden können. Um diese Interferenz zu beherrschen und ggf. für ein proaktives Ressourcenmanagement zu nutzen, werden Schätzverfahren benötigt, welche hoch-komplexe und dynamische Szenarien zuverlässig vorhersagen können. Dabei muss die Erfassung der Interferenzsituation über eine Mobilfunkzelle verteilt durch Endgeräte erfolgen und an die Basisstation gemeldet werden.

Zunächst wurde der mögliche Einsatz der In-X Subnetze „in-factory / in-Robot“ untersucht und ein Mobilitätsmodell für ein fabrikinternes Subnetz definiert und in Python implementiert. Die Ausbreitung von Funksignalen wurde realistisch modelliert, indem Pfadverluste, small scale Fading und Gaußsches Shadowing berücksichtigt wurden. Die zu erwartende Herausforderung einer exponentiell wachsenden Anzahl drahtlos verbundener Geräte wurde ebenfalls berücksichtigt, indem ein Szenario mit einer extrem dichten Verteilung von Subnetzen analysiert wurde. Darüber hinaus wurde das Kanalmodell zu einem vollständigen 3GPP-Modell mit räumlicher Konsistenz erweitert. Das vorgeschlagene Modell berücksichtigt die Dynamik von Line-Of-Sight (LOS)- und Non-Line-Of-Sight (NLOS)-Ausbreitungswegen bei Interferenzen, um eine realistischere Modellierung industrieller Umgebungen zu erzielen.

Darüber hinaus wurden zwei Mobilitätsmodelle integriert. Das erste Modell basiert auf zufälligen Richtungsbewegungen mit konstanter Geschwindigkeit, welches weiter verfeinert wurde, um Fabrikumgebungen besser abzubilden. Diese Verbesserungen umfassen unter anderem Kollisionsvermeidung – eine gängige Fähigkeit autonomer Roboter – sowie die Einschränkung der Bewegung auf einen abgegrenzten Bereich, wobei die Bewertung in einem realistischen Fabriklayout gemäß [GBD24b] erfolgte. Um die Anwendbarkeit des in T2.3 erarbeiteten Algorithmus in einem realistischeren Szenario weiter zu bewerten, wurde zusätzlich ein Indoor-Fabriklayout der 5G Alliance for Connected Industries and Automation (5G-

ACIA) betrachtet. Der zu Grunde liegende Fabrikgrundrisses der 5G-ACIA umfasst eine 180 m × 80 m große Anlage mit ausgewiesenen Zonen für Produktion, Montage und Lagerung sowie 5 m breite Gänge für die Bewegung von eingebetteten Roboter-SNs [GBD24b].

a) Beispiel für interferierende Subnetzwerk b) 5G ACIA Modell

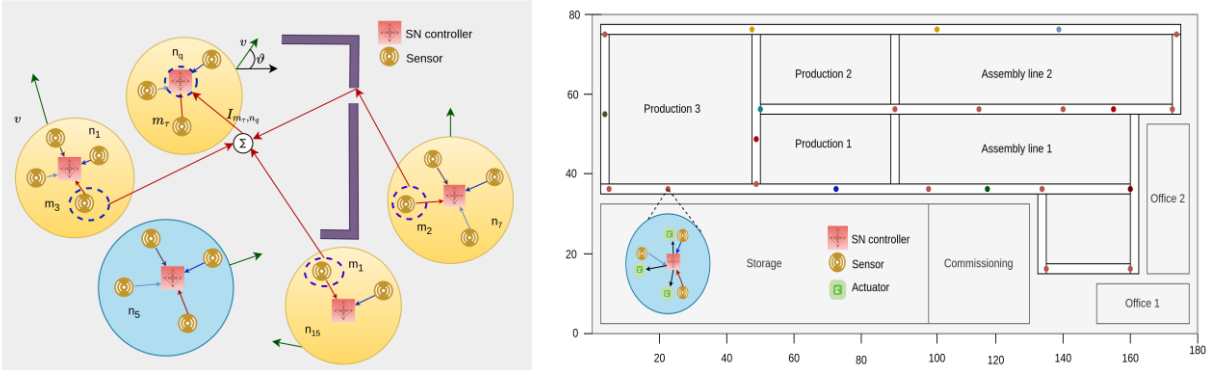

Abbildung 2 – a) Beispielszenario für Subnetzwerke bestehend aus einem Controller und Aktoren / Sensoren, die lokal kommunizieren. Die Subnetzwerke bewegen sich zufällig, so dass abrupte Wechsel zwischen Line-Of-Sight und Non-Line-Of-Sight Verbindungen durch Durchgänge oder Wände auftreten können. b) 5G-ACIA Modell mit verschiedenen Arealen und Bewegungspfaden auf denen sich Subnetzwerke durch eine Produktionshalle bewegen.

Außerdem wurden drei verschiedene Verkehrsmodelle für den angenommenen Datenverkehr betrachtet. Das erste ist ein isochrones Verkehrsmodell, bei dem Pakete in vordefinierten Intervallen in bestimmten Zeitschlitzen für bestimmte Sensor-Aktor-Paare übertragen werden. Zweitens wurde ein Bernoulli-Zufallsverkehrsmodell – als Bernoulli-isochroner Verkehr bezeichnet – verwendet, dass die Auswirkungen von Fehlanpassungen und sporadischen Ereignissen berücksichtigt. Darüber hinaus wurde das Framework um ein Push-Pull-basiertes Verkehrsmodell erweitert, das die Datenübertragung durch regelmäßige Berichte und ereignisbasierte, kontextabhängige Kommunikation abbildet. Der Umgang mit einem derartigen solchem Kreuzverkehrsverhalten stellt eine Herausforderung für die Entwicklung von Algorithmen dar.

Schließlich wurde in [Pav25] ein auf Raytracing basierendes Framework zur Interferenzgenerierung in industriellen Innenraumszenarien unter Verwendung der Sionna-Python-Bibliothek entwickelt. Das Framework beinhaltet ein detailliertes Fabriklayout, bei dem jedes Subnetzwerk aus einem einzelnen Sensor-Aktor-Paar besteht, das sich entlang einer vordefinierten Trajektorie bewegt. Dieses Setup ermöglicht eine realistische Modellierung von Funkinterferenzmustern in industriellen Umgebungen, da die räumlichen Beziehungen und dynamischen Bewegungen der Geräte innerhalb der Fabrikumgebung berücksichtigt werden.

T2.3-UB-2 Design eines tiefen neuronalen Netzes für die Interferenzschätzung

Auf Basis der Modellierungsarbeiten in Task T2.3-UB-1 wurde festgestellt, dass das Interferenzmuster der Subnetze aufgrund verschiedener Faktoren wie der zufälligen Mobilität, der Kanaleigenschaften und des Verkehrs komplex und potenziell nichtlinear ist. Da diese Subnetze niedrige Latenzzeiten und hohe Zuverlässigkeit erfordern, waren robuste und proaktive Verfahren zur Interferenzabschätzung erforderlich. Zunächst wurden Baseline-Methoden wie der Moving-Average-Ansatz und der Wiener-Prädiktor implementiert und bewertet. Dabei zeigte sich, dass diese linearen Verfahren die nichtlinearen Dynamiken der Interferenz, die durch Kanaleigenschaften, Mobilität und Verkehrsverhalten entstehen, nicht adäquat erfassen können. Daher wurde der Fokus anschließend auf Machine-Learning-gestützte Verfahren gelegt, die in der Lage sind, diese komplexen Zusammenhänge zu modellieren und somit eine robustere Interferenzabschätzung zu ermöglichen. Um dies zu erreichen, wurden Techniken des maschinellen Lernens wie rekurrente neuronale Netze, insbesondere Long Short Term Memory (LSTM), aus der vorhandenen Literatur untersucht. Ein einfaches LSTM-Modell wurde in Python implementiert und mit einem Interferenzleistungsvektor trainiert [GVB23, GBD24a]. Das implementierte Modell verwendete die Mean Squared Error (MSE)-Verlustfunktion und konnte die Interferenzleistung zuverlässig vorhersagen, solange eine deutliche Korrelation bestand und die Dynamik sowie das Verkehrsverhalten nicht zu komplex waren. Für

industrielle Subnetzwerke mit hyperzuverlässiger Kommunikation ist ein Ansatz, der nur auf dem bedingten Mittelwert basiert, jedoch unzureichend, da er Unsicherheiten und Variabilitäten in hochdynamischen Umgebungen nicht abbilden kann.

Im weiteren Verlauf wurde der Fokus auf schwer verteilte (heavy-tailed) Interferenz in industriellen Subnetzwerken gelegt, deren Verhalten stark durch sporadische Ereignisse und nicht-gaußsche Dynamiken beeinflusst wird. Zur Erfassung dieser Effekte und zur Quantifizierung der Unsicherheit wurde die Gaußsche Prozessregression (GPR) eingesetzt. Dabei zeigte sich, dass eine Student-t-Prozessregression eine höhere Robustheit gegenüber heavy-tailed-Verteilungen bietet. Allerdings erfordert die GPR eine präzise Kenntnis des Priors, was ihre Anwendbarkeit in dynamischen Szenarien einschränkt. Um diese Limitation zu überwinden, wurde der LSTM-Prädiktor zu einem Quantil-LSTM erweitert, das mithilfe einer Pinball-Loss-Funktion das 95 %-Quantil der Interferenz vorhersagt und so eine modellfreie, adaptive und robuste Interferenzabschätzung unter unsicheren Bedingungen ermöglicht.

Aufbauend auf diesen Erkenntnissen wurde der Ansatz auf 6G-Subnetzwerke (SNs) übertragen, um die entwickelten Methoden unter den Anforderungen sehr dichter Deployments und hochzuverlässiger Kommunikation mit niedriger Latenz (HRLLC) weiterzuentwickeln [GBD25b, GSB25b]. [GBD25b] stellt dabei ein risikobewusstes Vorhersagemodell für Interferenzen vor, das diesen Herausforderungen gerecht wird. In [Sap24] wurde die Vorhersage der Interferenz an einem bestimmten Quantil unter Verwendung der Pinball-Verlustfunktion evaluiert. Aufbauend darauf wurde in [GSB25b] eine vollständig neuartige Methodik entwickelt, die Extreme Value Theory (EVT), die Peak-over-Threshold (POT)-Methode, Tail-Statistiken, konformale Quantilsregression und Deep-Learning-Modelle kombiniert, um sporadische Interferenzspitzen, verursacht durch Asynchronitäten der Subnetze, Mobilität und tiefes Fading, realistisch zu erfassen.

Die POT-Methode erlaubt es, gezielt die Überschreitungen dynamischer Schwellenwerte zu beschrieben. Diese Schwellenwerte werden mithilfe des iQTransformers vorhergesagt – einer hybriden Architektur, die LSTM zur zeitlichen Modellierung, Transformer-Attention zur Erfassung von Querverbindungen zwischen SA-Paaren und Quantilsregression zur risikobasierten Wahrscheinlichkeitsabschätzung kombiniert. Das Framework nutzt die vorhandenen Korrelationen in den Interferenzdaten gezielt aus, erfüllt jedoch gleichzeitig die methodischen Voraussetzungen der eingesetzten Verfahren – etwa die Unabhängigkeitsannahmen, die für die POT-Schätzung und die konformale Quantilsregression erforderlich sind. Dadurch wird sowohl die statistische Gültigkeit als auch die realistische Abbildung korrelierter Dynamiken sichergestellt. Darüber hinaus werden Tail-Abhängigkeiten zwischen extremen Interferenzereignissen modelliert, wodurch korrelierte Spitzen realistisch erfasst werden können.

Ein besonderer Schwerpunkt in [GSB25b] liegt auf der Robustheit bei begrenzten Stichprobengrößen, da in realistischen industriellen Szenarien die Zahl beobachteter Extremereignisse naturgemäß limitiert ist. Durch die Integration konformalisierter Quantilsregression entstehen statistisch kalibrierte, verteilungsfreie Vorhersageintervalle mit formalen Abdeckungszusagen (coverage guarantees), die auch unter geringen Datenmengen eine verlässliche Interferenzvorhersage ermöglichen. Diese Kombination aus EVT, POT, Tail-Statistik, Tail-Abhängigkeitsmodellierung und konformaler Quantilsregression bildet den Kernbeitrag von [GSB25b] und ermöglicht eine proaktive, fein abgestimmte und robuste Ressourcenallokation, die selbst unter extremen und seltenen Interferenzereignissen eine zuverlässige Kommunikation sicherstellt.

Darüber hinaus betont [GSB25b] die Bedeutung kalibrierter Wahrscheinlichkeitsdichtefunktionen (PDFs) der Interferenzrandverteilungen für eine zuverlässige Interferenzvorhersage in 6G-Subnetzwerken. Im Gegensatz zu herkömmlichen Prädiktoren, die sich nur auf Mittelwerte oder feste Quantilsabschätzungen stützen, modelliert der vorgeschlagene Ansatz Interferenzüberschreitungen jenseits eines dynamisch vorhergesagten Schwellenwerts und erzeugt statistisch abgesicherte, risikobewusste Vorhersageintervalle. Die so kalibrierten Interferenzverteilungen passen sich an Herausforderungen wie Asynchronität im Traffic, sporadische Ereignisse, Tail-Abhängigkeiten, begrenzte Stichprobengrößen und starkes Fading an – zentrale Faktoren für HRLLC-Anforderungen. Durch die Einbeziehung dieser kalibrierten, risikobewussten Informationen in die Ressourcenallokation konnte eine Zuverlässigkeit von über 95 % in Bezug auf die Zielfehlerrate erreicht werden – eine robuste, theoretisch fundierte Lösung für sehr dichte und latenzkritische industrielle Umgebungen.

Außerdem wurde in dem Masterprojekt [Rut25] wird an der Subband-Zuweisung zur Interferenzverwaltung für ein Indoor-Fabrik-Szenario gearbeitet, wobei die zentralisierte Sequential Iterative Sub-band Allocation (SISA) und eine Grafenfärbung-Technik verwendet werden. Ergänzend dazu wurde im Masterprojekt [Buh24] ein Ansatz zur leistungsbasierten Interferenzsteuerung entwickelt, der auf dem Weighted Minimum Mean Square Error (WMMSE)-Algorithmus basiert und eine effiziente Leistungsregelung in dicht vernetzten industriellen Subnetzwerken ermöglicht.

T2.3-UB-3 Erforschung eines verteilten Interferenzschätzers mittels Split Learning

Für einige Anwendungsfälle ist eine Aufspaltung der in T2.3-UB-2 betrachteten LSTM-Architektur in zwei Teile notwendig, um beschränkte Rechenleistungen zu berücksichtigen. Hierfür schlägt [GSB25b] eine U-förmige Split-Learning-Architektur zur Interferenzvorhersage in ressourcenbeschränkten 6G-Subnetzwerken vor, bei der der iQTransformer in drei Komponenten aufgeteilt wird: ein Head- und Tail-Modell bei jedem Sensor-Aktor-(SA)-Paar sowie ein Server-Modell im SN-Controller. Diese Architektur ermöglicht eine lokale Einbettung und Quantilprojektion auf Geräteebene, während im zentralen Server multivariate Interferenzdynamiken mithilfe von Attention- und LSTM-Schichten erfasst werden. Dadurch werden sowohl Skalierbarkeit als auch Datenschutz gewahrt und gleichzeitig die Rechen- und Kommunikationslast erheblich reduziert. Der vorgeschlagene Split-iQTransformer wahrt zudem die theoretische Konsistenz der verwendeten Verfahren, nutzt die vorhandenen Korrelationen in den Interferenzdaten gezielt aus und erfüllt zugleich die für Extreme Value Theory (EVT) und konformale Quantilsregression erforderlichen statistischen Voraussetzungen. Entscheidend ist, dass der Ansatz die in T2.3-UB-2 entwickelte kalibrierte Tail-PDF-Schätzung integriert, indem vorhergesagte dynamische Schwellenwerte mit der EVT kombiniert und Überschreitungen mithilfe der Peak-over-Threshold-(POT)-Methode sowie der verallgemeinerten Pareto-Verteilung (GPD) modelliert werden. Die resultierenden kalibrierten Interferenzverteilungen passen sich dabei adaptiv an die lokalen Bedingungen jedes SA-Paares an. Trotz der verteilten Ausführung erreicht der Split-iQTransformer eine vergleichbare Leistung zum zentralisierten Ansatz, bietet rechnerische Flexibilität und erfüllt gleichzeitig die 95-Perzentil-Zuverlässigkeit in Bezug auf die Blockfehlerrate (BLER) – ein wesentlicher Beitrag zu einer effizienten, skalierbaren und zuverlässigen Interferenzvorhersage in 6G-Industrie-Subnetzwerken.

T2.3-UB-4 Anpassung des Interferenzschätzers an veränderliche Interferenzsituationen

Die entwickelten Interferenzprädiktoren wurden umfassend evaluiert, indem die Anzahl der Interferer, das Messintervall der Interferenzsituation sowie die Gesamtzahl der Subnetze in Indoor-Fabrikszenarien unter Verwendung des 3GPP-Kanalmodells variiert wurden. Darüber hinaus erfolgte eine Bewertung unter unterschiedlichen Verkehrsbedingungen sowie in einem realistischen Laboraufbau der Technischen Universität Dortmund, um die Praxistauglichkeit der trainierten Modelle zu überprüfen. Dabei zeigte sich, dass die vorgeschlagene Methode im Vergleich zu Basismethoden wie dem gleitenden Durchschnitt eine deutlich höhere Vorhersagegenauigkeit erzielt.

Ergänzend wurde das im Rahmen von [GBD25c] entwickelte Verfahren RASP (Reliability-Aware SINR Prediction) integriert, welches ein probabilistisches, BLER-bewusstes Vorhersageframework für industrielle Subnetzwerke darstellt. In industriellen Subnetzwerken erweisen sich herkömmliche Offline-Vorhersagemethoden als unzureichend, da sich Mobilität, Ausbreitungsbedingungen und Kanaldynamik im Zeitverlauf verändern und vollständige Datensätze im Voraus oft nicht verfügbar sind. RASP begegnet diesen Herausforderungen, indem es ein Mixture Density Network (MDN) mit einer Conditional Value-at-Risk (CVaR)-basierten Optimierung kombiniert, um die SINR-Verteilung unter Zuverlässigkeitsbedingungen realistisch zu modellieren und die Hyper-Reliability-Anforderungen industrieller Kommunikation zu erfüllen. Da sich die Interferenzcharakteristik und das Verkehrsverhalten je nach Fabrikumgebung, Netzwerktopologie und Szenariokonfiguration unterscheiden, wird das trainierte Modell jeweils an die spezifischen Bedingungen des betrachteten Szenarios angepasst. Die Methode wurde anhand realer Industrie-Messkampagnen mit koexistierenden Subband-Interferenzen verifiziert und zeigte eine Zuverlässigkeit bei einem 95-Perzentil-BLER, wodurch sie die Baseline-Modelle signifikant übertrifft. Obwohl in der aktuellen Testumgebung kein Federated Learning (FL) erforderlich war, kann dieser Ansatz – in Anlehnung an [GBD24a]

– zukünftig eingesetzt werden, um einen dezentralen Modellaustausch zwischen Subnetzwerken zu ermöglichen und die Skalierbarkeit sowie Anpassungsfähigkeit des Systems weiter zu erhöhen.

T2.3-UB-5 Untersuchung von Lernarchitekturen mit Federated Learning an den Endgeräten

Über die zentrale Lösung des Interferenzvorhersageproblems hinaus wurde das Konzept der kooperativen Interferenzschätzung und -vorhersage für Teilnetze entwickelt, die auf demselben Subband arbeiten und sich somit gegenseitig stören. Es konnte gezeigt werden, dass eine Beschränkung der Analyse auf die tatsächlich interferierenden Teilnetze – anstelle einer globalen Analyse aller Teilnetze – ohne signifikanten Genauigkeitsverlust möglich ist. Aufbauend auf den Arbeiten zu FedAVG-in-X [GVB23] und [GBD24a] wurde ein föderiertes, LSTM-basiertes Lernverfahren entwickelt, das den Wissensaustausch zwischen einzelnen Sensor-Aktor-(SA-)Paaren innerhalb des Teilnetzes ermöglicht, ohne dass Rohdaten oder Interferenzmesswerte geteilt werden müssen. Dies gewährleistet sowohl den Datenschutz als auch eine Reduzierung der Rechenkomplexität. Gleichzeitig beschleunigt der föderierte Lernansatz die Modellkonvergenz deutlich und verbessert die Vorhersagegenauigkeit im Vergleich zu klassischen Verfahren wie dem Wiener-Prädiktor oder dem gleitenden Mittelwertschätzer um bis zu 4,5 dB. Dieser Ansatz stellt somit eine skalierbare und effiziente Lösung für die lokale Interferenzvorhersage in industriellen Teilnetzen dar, die flexibel an unterschiedliche Szenarien, Mobilitätsmodelle und Verkehrsmuster angepasst werden kann.

T2.3-UB-6 Erforschung von Methoden zur unüberwachten Datenannotierung

In manchen Fällen kann die Ermittlung des Labels und der Interferenzleistung am Subnetz-Controller zu höheren Kommunikationskosten führen. Um diese Herausforderung zu bewältigen, wurden zunächst Arbeiten zu einem online-unüberwachten erweiterten Kalman-Filter für die Schätzung und Vorhersage von Störungen durchgeführt [GSB24]. Aufbauend darauf stellt [GSB25c] ein robustes und effizientes Framework zur dynamischen Interferenzvorhersage vor, dass die Link Adaptation (LA) in ultradichten 6G-Industriesubnetzwerken mit strengen HRLLC-Anforderungen unterstützt. Die Interferenz wird dabei als latenter Prozess über ein neuartiges vektorbasiertes diskretes Zustandsraummodell (vDSSM) modelliert, das aus verzögerten, quantisierten und verrauschten CQI-Berichten abgeleitet wird. Um der komplexen Verteilung der Interferenz und sporadischen Störungen gerecht zu werden, kombiniert das Verfahren eine Sparse Student-t Process Regression (SPTPR) mit einem modifizierten Unscented Kalman Filter (MUKF). Diese hybride Methode ermöglicht skalierbare, rechenaufwandsarme und robuste Vorhersagen mit deutlich reduzierter Komplexität im Vergleich zu herkömmlichen Verfahren. Das Framework wurde unter reduzierter Stichprobenanzahl sowie unter Verwendung des 5G-ACIA-basierten Mobilitätsmodells aus T2.3-UB-1 in einem realistischen Fabrikszenario mit mehreren interferierenden Subnetzwerken evaluiert. Trotz weniger und sparse vorliegender Trainingsdaten konnte das Framework dank seines rekursiven Verfeinerungsverfahrens eine stabile und zuverlässige Vorhersageleistung aufrechterhalten. In dynamischen Fabrikumgebungen mit Subband-Interferenzen gewährleistete der Ansatz eine zuverlässige Link Adaptation und erfüllte die Anforderungen an Hyper-Reliability und Echtzeitkommunikation in latenzkritischen industriellen Anwendungen.

6.4 AP3: Network of Networks

Die Universität Bremen war nicht an den Arbeiten in AP3 beteiligt.

6.5 AP4: Automation & Simplification

Die Universität Bremen war nicht an den Arbeiten in AP4 beteiligt.

6.6 AP5: Testbeds & Proof of Concepts

Die Universität Bremen war nicht an den Arbeiten in AP5 beteiligt.

7 Erläuterungen zu den wichtigsten Positionen des zahlenmäßigen Nachweises

Alle geplanten Anschaffungen wurden plangemäß durchgeführt.

8 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die zur Durchführung des oben beschriebenen Vorhabens erforderlichen Mittel konnten nicht aus anderweitigen Quellen wie beispielsweise Institutsmitteln bereitgestellt werden. Förderungen seitens der Industrie schieden ebenfalls aus, denn aufgrund der Innovationshöhe der erforschten Methoden und Verfahren bestand zur Laufzeit des Projektes ein zu hohes technisches und fachliches Risiko für Industriepartner. Zudem ist eine isolierte Betrachtung von Technologien, vor allem für die hier angedachte Erarbeitung eines 6G-Gesamtkonzepts, nicht zielführend. Vielmehr müssen die Technologien in Form eines integrativen Ansatzes in Zusammenarbeit und Diskussion mit Partnern eingebettet werden. Die Komplexität der Gesamtaufgabe erforderte somit die Zusammenarbeit mit akademischen Partnern und Industriepartnern.

9 Voraussichtlicher Nutzen und Verwertbarkeit

Die wissenschaftlichen Arbeiten des Vorhabens führten zu grundlegenden neuen Methoden, Modellen und Verfahren im Bereich verteilter Interferenzschätzung in 6GWANs mit in-X-Subnetzwerken und Relevanz-basierter Signalverarbeitung für cell-free massive MIMO.

Die wissenschaftliche Verwertbarkeit und der Nutzen sind auch in unmittelbarer Zukunft als positiv anzusehen. Im Themenbereich cell-free massive MIMO wurden zahlreiche hochrelevante Ergebnisse erarbeitet (verteilten Quellencodierung und verteilte gemeinsame Quellen-Kanal-Codierung im cell-free massive MIMO, cell-free Kompressionsarchitekturen, modellbasierte und modellfreie Algorithmen). Auch der adressierte Themenbereich der Interferenzschätzung in 6GWANs mit in-X-Subnetzwerken hat bereits zu neuen Erkenntnissen (z.B. KI-gestützte Schätz- und Prediktionsverfahren, KI-Architekturen für verteilte Schätzung, Quantil-Regression und die Integration von Extreme Value Theory) geführt und bietet hervorragende Anknüpfungspunkte zu weitergehenden theoretischen, aber auch praxisrelevanten Fragestellungen wie bereits in einer ersten Zusammenarbeit mit der TU Dortmund aufgezeigt. Es kann daher davon ausgegangen werden, dass die erzielten Ergebnisse sowohl von weiteren Forschungseinrichtungen als auch von der industriellen Forschung aufgegriffen werden. Zum Teil ist dies bereits im Rahmen der fortschreitenden 6G Diskussionen erfolgt, da Interferenzschätzung, - Prädiktion und -Management auch in der 3GPP stark an Bedeutung gewonnen haben. Entsprechend werden die im Rahmen des Projekts gewonnenen Erkenntnisse und Ergebnisse auch weiterhin auf internationalen wissenschaftlichen Tagungen und Workshops präsentiert und in international renommierten Fachzeitschriften publiziert.

Das Projekt hat einerseits neue Erkenntnisse im Bereich der Grundlagenforschung hervorgebracht, andererseits aber auch praxisnahe Verfahren und Methoden in Form von konkreten Lösungsansätzen für cell-free massive MIMO und die Interferenzprädiktion in Zusammenarbeit mit NOKIA Stuttgart und der TU Dortmund. Im Zusammenhang mit der oben beschriebenen Verwertung hat das Vorhaben somit Impulse für die weitere Forschung auf dem Gebiet der 6G-Forschung generiert. Die Ergebnisse des Vorhabens dienen als Basis für weitergehende nationale und internationale Forschungsvorhaben, die sowohl die Grundlagenforschung als auch die angewandte Forschung adressieren. Konkret werden die in 6G-ANNA erarbeiteten Ideen und Technologien bereits in dem BMFTR-geförderten Forschungsprojekt Open6GHub+ und zukünftig in 6G-Coverage weitergeführt.

Das Projektspektrum der Universität Bremen umfasst auch Projekte in der industrienahen Forschung und es ist weiterhin ein Ziel, durch die in diesem Projekt gewonnen Erkenntnisse zukünftige kooperative Vorhaben zu initiieren. Die entwickelte 6G-Interferenzprädiktionsverfahren und cell-free MIMO-Technologien werden für den Anwendungsfall der 3D Netzwerke an der Universität Bremen weiterentwickelt und in Demonstratoren integriert, die als Teil der Ausstellungsflächen im Digital Hub Industrie (DHI, https://digitalhubindustry.de/) und des OpenLab „3D Netzwerke“ regelmäßig der Öffentlichkeit zugänglich gemacht. Darüber hinaus ist das hier adressierte Themengebiet auch mittelfristig noch Gegenstand weltweiter Forschungs-, Standardisierungs- und Normierungsaktivitäten, so dass grundsätzlich von einer hohen bis sehr hohen wissenschaftlichen und wirtschaftlichen Verwertbarkeit ausgegangen werden kann.

Weiterhin wurden die erarbeiteten Ergebnisse und Erkenntnisse für die Nachwuchsförderung genutzt. Das Thema 6G sowie die bearbeiteten wissenschaftlichen Themen haben in das angebotene Lehrangebot Eingang

gefunden und wurden den Studierenden durch selbständiges Arbeiten in studentischen Projekten und in einem Masterprojekten und vier Masterarbeiten nahegebracht. Zudem befindet sich eine Dissertation zum Thema Interferenzprädiktion derzeit in Arbeit.

10 Berücksichtigung anderer FE-Ergebnisse

Die Durchführung der Arbeiten in 6G-ANNA wird sowohl durch Fortschritte in der Literatur und Forschung berührt. Die Universität Bremen ist im Bereich der PHY/MAC - Forschung für 6G in verschiedenen relevanten Bereichen (Theoretische Grenzen, Prädiktionsverfahren, MIMO-Technologien) und im Bereich des maschinellen Lernens für Kommunikationsnetze international vernetzt und verfolgt aktuelle Forschungserbnisse. Im Berichtszeitrum sind keine Forschungsergebnisse bekannt geworden, die eine Änderung der geplanten Arbeiten notwendig machten.

11 Erfolgte und geplante Verwertung

11.1 Veröffentlichungen

Wissenschaftliche Publikationen	
[DHW24]	A. Danaee, S. Hassanpour, D. Wübben, A. Dekorsy *Relevance-based Information Processing for Fronthaul Rate Reduction in Cell-free Massive MIMO Systems* 19th International Symposium on Wireless Communication Systems (ISWCS), Rio de Janeiro, Brasilien, Juli 2024, doi: 10.1109/ISWCS61526.2024.10639087.
[DHW25]	A. Danaee, S. Hassanpour, D. Wübben, A. Dekorsy *Relevance-Based Multi-User Data Compression for Fronthaul Rate Reduction in Cell-Free Massive MIMO Systems* International ITG Conference on Systems, Communications and Coding (SCC 2025), Karlsruhe, März 2025, doi: 10.1109/IEEECONF62907.2025.10949116.
[GVB23]	P. Gautam, M. Vakilifard, C. Bockelmann, A. Dekorsy *Cooperative Interference Estimation using LSTM-based Federated Learning for In-X Subnetworks* IEEE Global Communication Conference (Globecom 2023), Kuala Lumpur, Malaysia, Dezember 2023, doi: 10.1109/GLOBECOM54140.2023.10437515.
[GBD24a]	P. Gautam, C. Bockelmann, A. Dekorsy *Interference Prediction in Unconnected In-X Mobile 6G Subnetworks Using a Data-Driven Approach* IEEE International Conference on Communications (ICC 2024), XURLLC6G Workshop, Denver, CO, USA, Juni 2024, doi: 10.1109/ICCWorkshops59551.2024.10615704.
[GBD24b]	P. Gautam, C. Bockelmann, A. Dekorsy *Probabilistic Interference Prediction for Dynamic 6G In-X Sub-networks* IEEE Open Journal of the Communications Society, Vol. 6, pp. 2454-2473, April 2025, doi: 10.1109/OJCOMS.2025.3554993.
[GBD25a]	P. Gautam, C. Bockelmann, A. Dekorsy, et al. *Towards AI in 6G: Concepts, Techniques, and Standards* in IEEE Access, vol. 13, pp. 143843-143874, 2025, doi: 10.1109/ACCESS.2025.3595752.
[GBD25b]	P. Gautam, C. Bockelmann, A. Dekorsy *Extreme Value Theory-based Predictive Interference Management for 6G Subnetworks with Transformer* IEEE International Conference on Communications (ICC 2025), Montreal, Canada, 2025, pp. 6964-6969, doi: 10.1109/ICC52391.2025.11161159.

[GDB25a]	P. Gautam, A. Danaee, C. Bockelmann, D. Wübben, A. Dekorsy et al. *6G PHY: Insights from 6G-ANNA Research Initiative* angenommen zur Veröffentlichung in IEEE Open Journal of the Communications Society
[GSB25a]	P. Gautam, R. Sharan, P. Baracca, C. Bockelmann, T. Wild, A. Dekorsy *Dynamic Interference Prediction for In-X 6G sub-networks* International ITG Conference on Systems, Communications and Coding (SCC 2025), Karlsruhe, März 2025, doi: 10.1109/IEEECONF62907.2025.10949124.
[GSB25b]	P. Gautam, S. Sapkota, C. Bockelmann, S. R Pandey, A. Dekorsy *Extreme Value Theory-based Distributed Interference Prediction for 6G Subnetworks* IEEE Open Journal of the Communications Society, vol. 6, pp. 9121-9143, 2025, doi: 10.1109/OJCOMS.2025.3625505.
[HDW24]	S. Hassanpour, A. Danaee, D. Wübben, A. Dekorsy *Multi-Source Distributed Data Compression Based on Information Bottleneck Principle* IEEE Open Journal of the Communications Society, Vol. 5, pp. 4171-4185, Juli 2024, doi: 10.1109/OJCOMS.2024.3426049.
[HDW25]	S. Hassanpour, A. Danaee, D. Wübben, A. Dekorsy *Forward-Aware Multi-Source Distributed Quantization for Noisy Channels via Information Bottleneck Principle* IEEE Open Journal of the Communications Society, Vol. 6, pp. 7424 - 7440, September 2025, doi: 10.1109/OJCOMS.2025.3605737.
[HHW25a]	S. Hassanpour, M. Hummert, D. Wübben, A. Dekorsy *Deep Learning-Based Distributed Remote Source Coding via Information Bottleneck Method: The Parallel Processing Scheme* IEEE International Conference on Communications Workshops (ICC Workshops), Montreal, Kanada, Juni 2025, doi: 10.1109/ICCWorkshops67674.2025.11162203.
[HHW25b]	S. Hassanpour, M. Hummert, D. Wübben, A. Dekorsy *A Deep Variational Approach to Multiterminal Joint Source-Channel Coding Based on Information Bottleneck Principle* IEEE Open Journal of the Communications Society, Vol. 6, pp. 4462 - 4475, Mai 2025, doi: 10.1109/OJCOMS.2025.3570446.
[WRG24]	D. Wübben, M. Rihan, P. Gautam, A. Dekorsy et al. *German Perspective on 6G – Use Cases, Technical Building Blocks and Requirements* FAU University Press, FAU Studien aus der Elektrotechnik, Vol. 28, Erlangen, Deutschland, 19. Dezember 2024, ISBN 978-3-96147-797-5

Wissenschaftliche Publikationen (eingereicht zur Veröffentlichung)	
[GBD25c]	P. Gautam, C. Bockelmann, A. Dekorsy, et al. *RASP: Reliability-Aware SINR Prediction for Realistic Industrial Subnetworks* eingereicht zu IEEE International Conference on Communications (ICC 2026).
[GSB25c]	P. Gautam, R. Sharan, P. Baracca, C. Bockelmann, T. Wild, A. Dekorsy *Dynamic Interference Prediction for Link Adaptation for Industrial Sub-networks* eingereicht bei IEEE Wireless Communication Letters
[GDB25b]	P. Gautam, A. Danaee, C. Bockelmann, D. Wübben, A. Dekorsy et al. *AI/ML-based solutions for enhancing 6G PHY performance and energy efficiency* eingereicht bei IEEE Access

11.2 Studentische Arbeiten

Studentische Arbeiten	
[Sap24]	Sushmita Sapkota *Interference Estimation using Machine Learning for in-X subnetworks* Masterarbeit, Universität Bremen, 2024
[Buh24]	Nicolas Buhr *Finite-Alphabet Message Passing Decoder for 5G LDPC Codes* Masterarbeit, Universität Bremen, 2024
[Rut25]	Uderagudi Shelciya Ruth *Dynamic Resource Allocation for 6G Network of Networks* Masterproject, Universität Bremen
[Mbn25]	Md Riad Hossain Bhuiyan *Dynamic Power Control for In-X subnetworks* Masterproject, Universität Bremen, 2025
[Pav25]	Chandrika Pavuluri *Ray Tracing for Channel Modeling in 6G Subnetworks* Masterproject, Universität Bremen, 2025

11.3 Fachvorträge

Fachvorträge	
[Wüb21]	D. Wübben *Artificial Intelligence for Mobile Communications* 25. VDE/ITG Fachtagung Mobilkommunikation, Osnabrück, 3.-4.11.2021
[Dek22]	A. Dekorsy *Relevance-Based Information Processing* „Future of Communications“ Workshop des Villum Investigator Grant "WATER", Aalborg, Dänemark, 12.12.2022
[Bok25]	C. Bockelmann *Role and potential of AI for the Radio Physical Layer* VDE ITG Expert Symposium “Future of Networking”, Ilmenau, 04.09.2025

11.4 Projektberichte

Berichte	
[6G-ANNA-D2a]	6G-ANNA, WP2 – 6G Access *Milestone Report D2a* *Juni 2023*
[6G-ANNA-D2b]	6G-ANNA, WP2 – 6G Access *Milestone Report D2b* *Juni 2024*
[6G-ANNA-D2c]	6G-ANNA, WP2 – 6G Access *Milestone Report D2c* *Juni 2025*

11.5 Wissenschaftskommunikation

Das ANT hat im Laufe des Projektzeitraums im Rahmen des Bremer Digital Hub Industry (DHI, https://digital-hubindustry.de/) neue Räumlichkeiten für den Aufbau und die Vorführung von Testbeds und Demonstratoren erhalten. Das ANT hat eine Ausstellungsfläche entwickelt, auf welcher verschiedene Demonstratoren zum Thema 5G, 6G und KI in der Nachrichtentechnik aufgebaut wurden.

Im letzten Projektjahr wurden erste in 6G-ANNA entwickelte Interferenzschätzverfahren mittels eines vorhandenen AGV-Testbeds untersucht und verifiziert. Der Aufbau eines entsprechenden Demonstrators ist für die nahe Zukunft geplant. Seit 2023 hat das die Universität Bremen regelmäßig über 30 Vorführungstermine pro Jahr durchgeführt, um Industrie, Schülern und Stakeholdern aus Politik und Gesellschaft die vom BMFTR-geförderten Forschungsvorhaben vorzustellen und verständlich zu machen.