Institut für Mikroelektronische Systeme

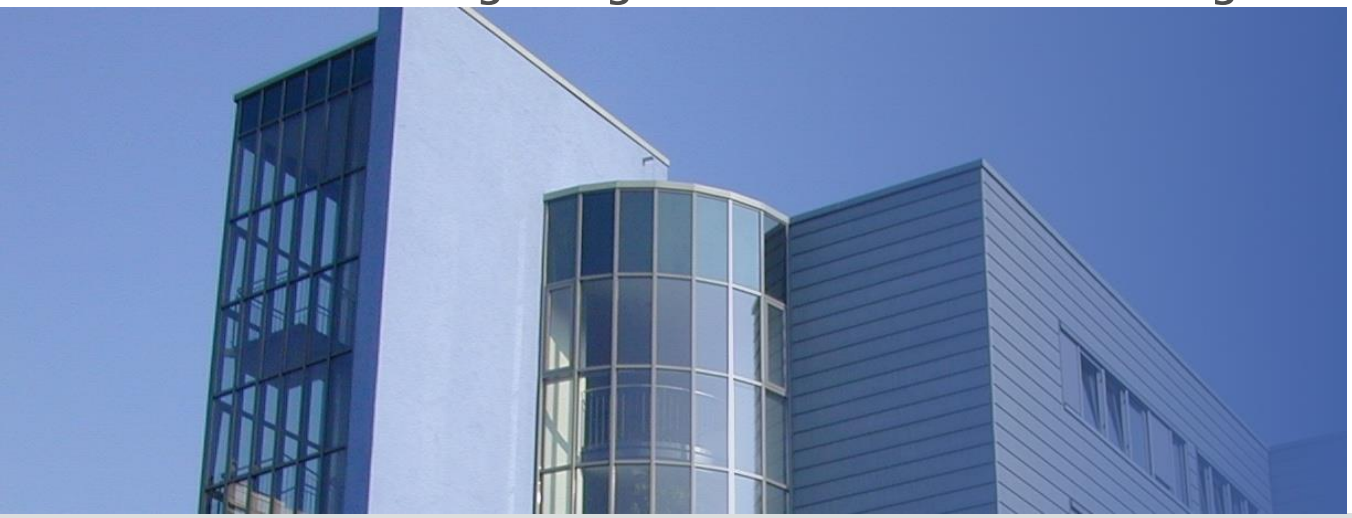
Smart HeaP

Leibniz Universität Hannover

Evaluation and optimization of a Tensilica processor for hearing aids

Tensilica Day 2019

M.Sc. Jens Karrenbauer, Dipl. Ing. Lukas Gerlach,
Prof. Dr.-Ing. Holger Blume, Jun.-Prof. Dr.-Ing. Guillermo Payá Vayá

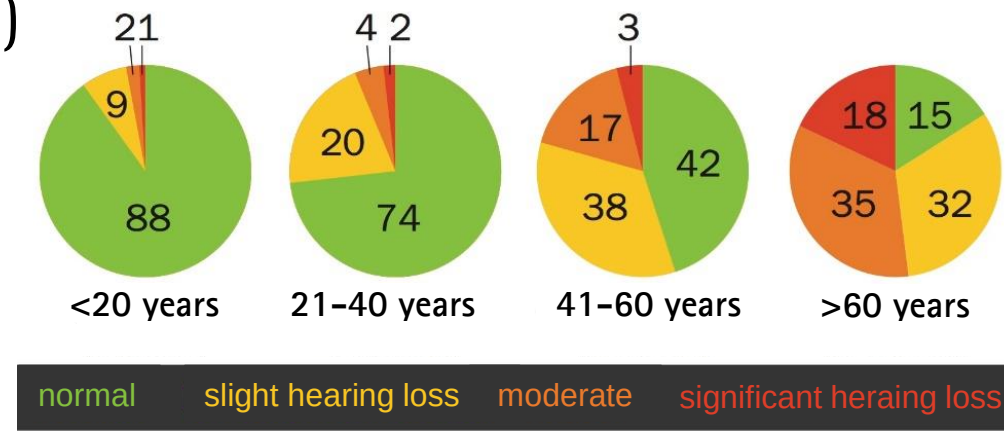


Outline

- Motivation
- Smart Hearing Aid Processor (Smart HeaP)
- Framework
- Profiling And Optimization Techniques
- Virtual Prototyp
- Outline and Future Work

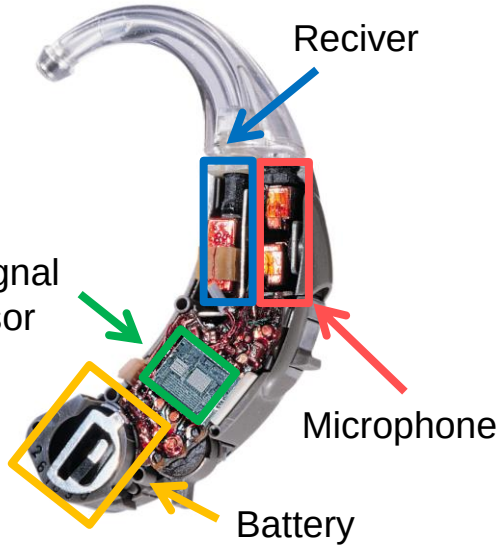
Motivation
Challenges For Hearing Aid Processors

- Around 466 million people suffer from a hearing loss (WHO)
- Less than 16% of the affected Americans use hearing aids (NIDCD)

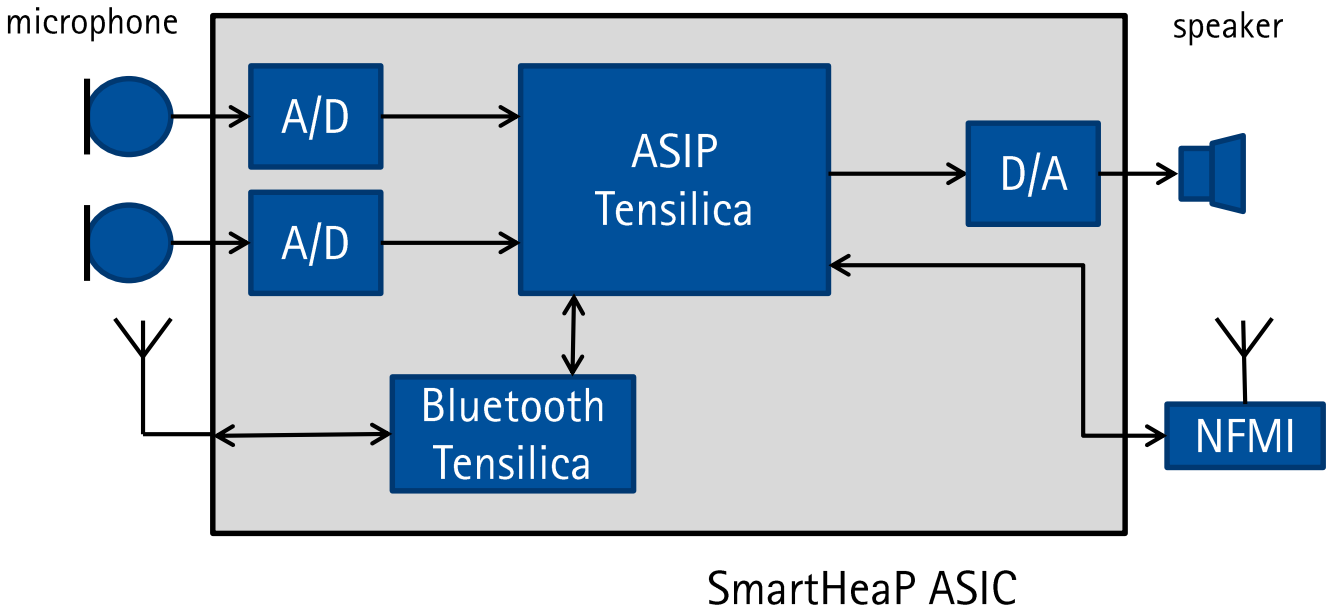


"obs Fördergemeinschaft Gutes Hören (FGH)"

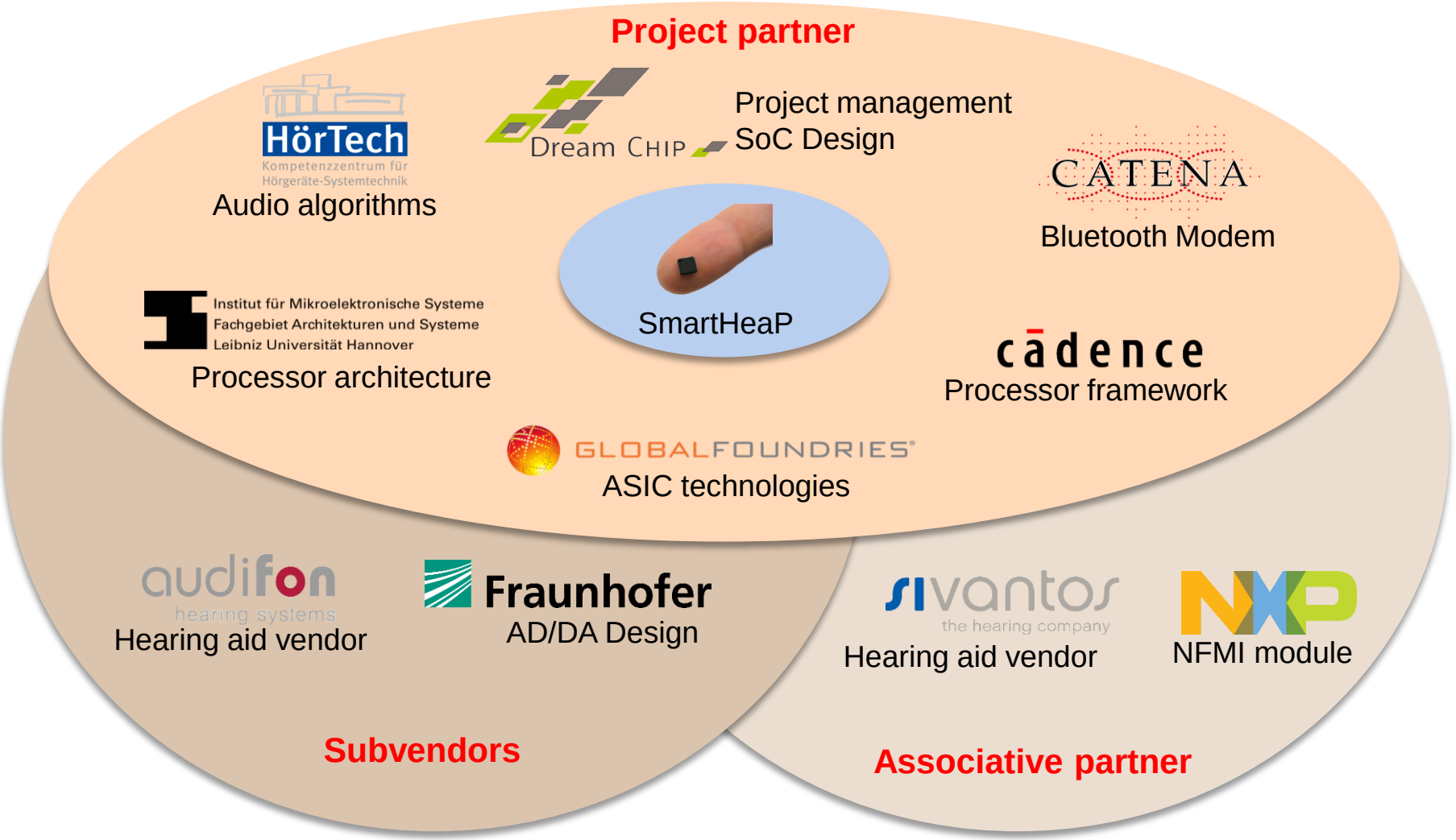
- Small form factor
- Low power: *a few mW*
- Low processing delay: *< 10 ms*
- **Programmability / flexibility**



Smart Hearing Aid Processor (SmartHeaP)
ASIC Concept

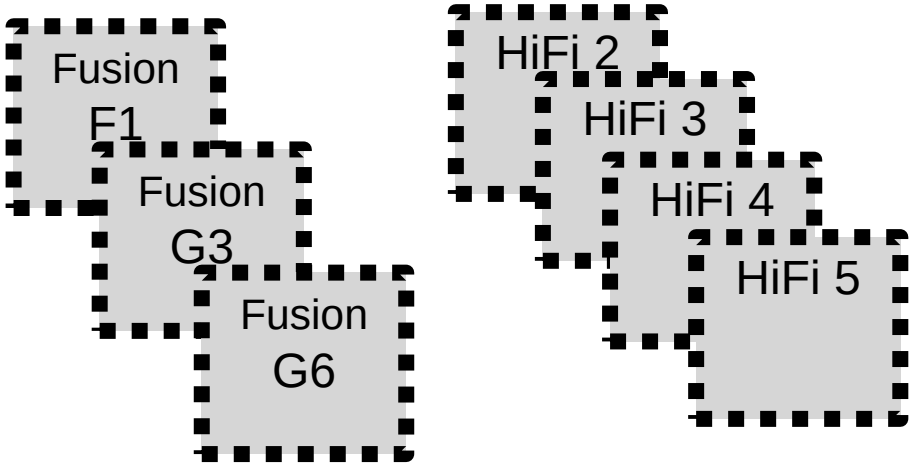


Smart Hearing Aid Processor (SmartHeaP)
Collaborated BMBF Project



Smart Hearing Aid Processor (SmartHeaP)
Steps And Goals

- Selection of a processor architecture



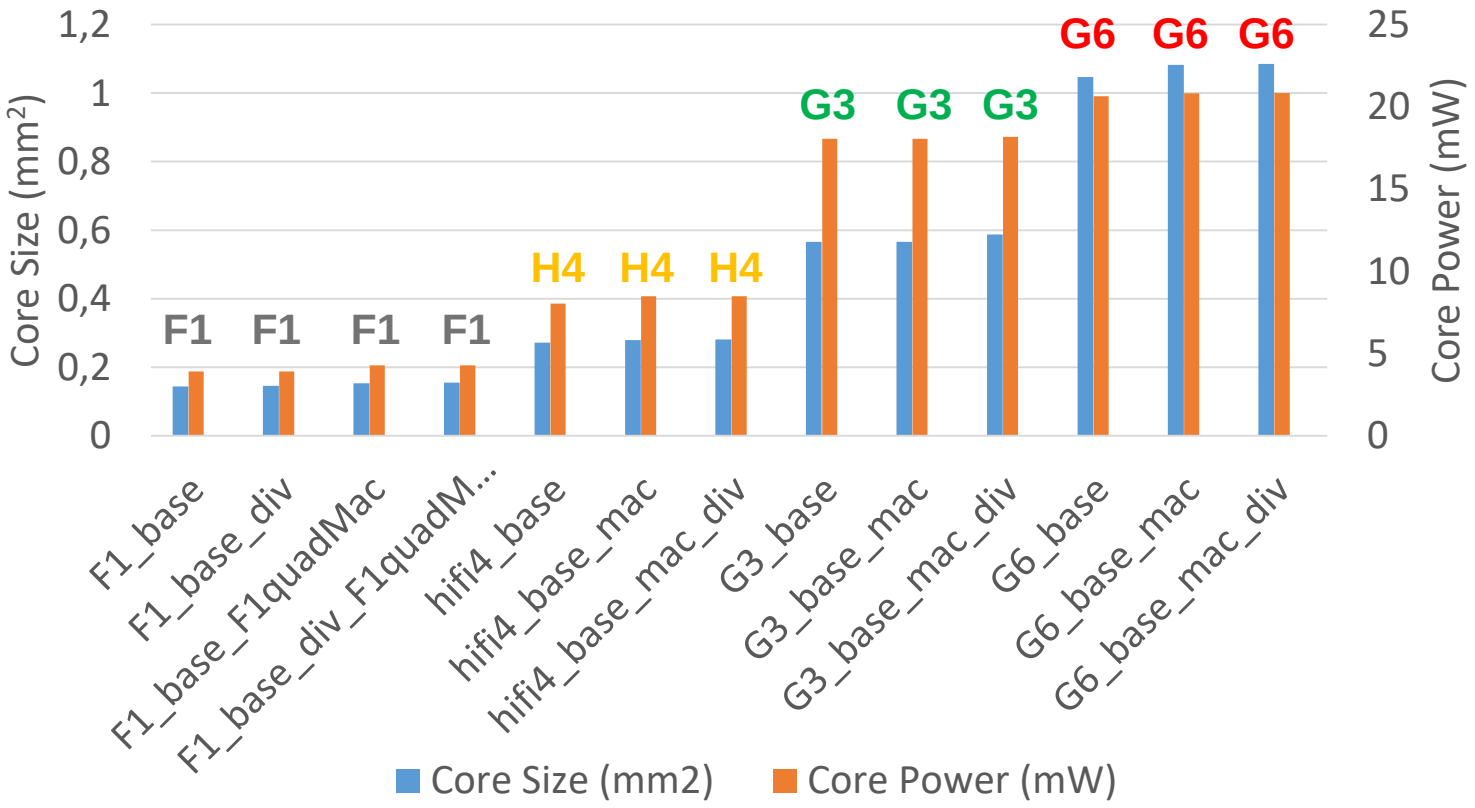
Cadence Tensilica Processors

	Hifi4	G3	G6
Load/Store (Bits)	2x32	2x128	2x256
ALU ops/cycle *	4	4	8
MACs/cycle *	4	4	8
Issue-Slots	4	4	4
SIMD *	2	4	8

*32-Bit operands

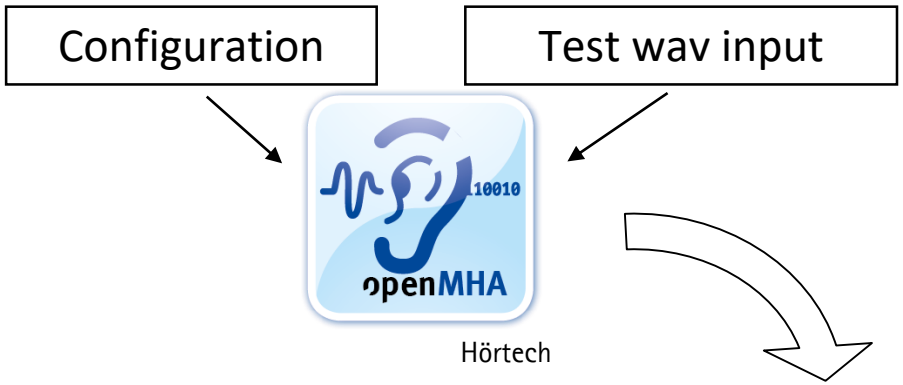
- Implementation of reference algorithms
- Profiling and design space exploration
- Processor evaluation
- Performance gain evaluation

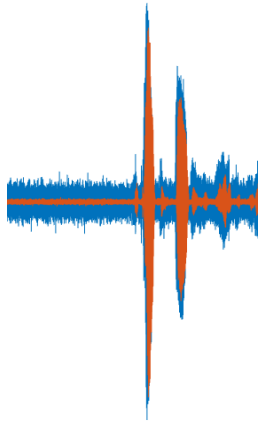
Smart Hearing Aid Processor (SmartHeaP)
Area and Power Consumption



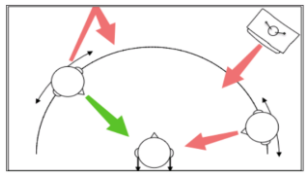
- 28 nm HPL technology at 100 MHz

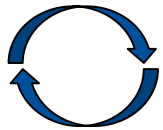
Framework
Fixed-Point Code And DSE Flow



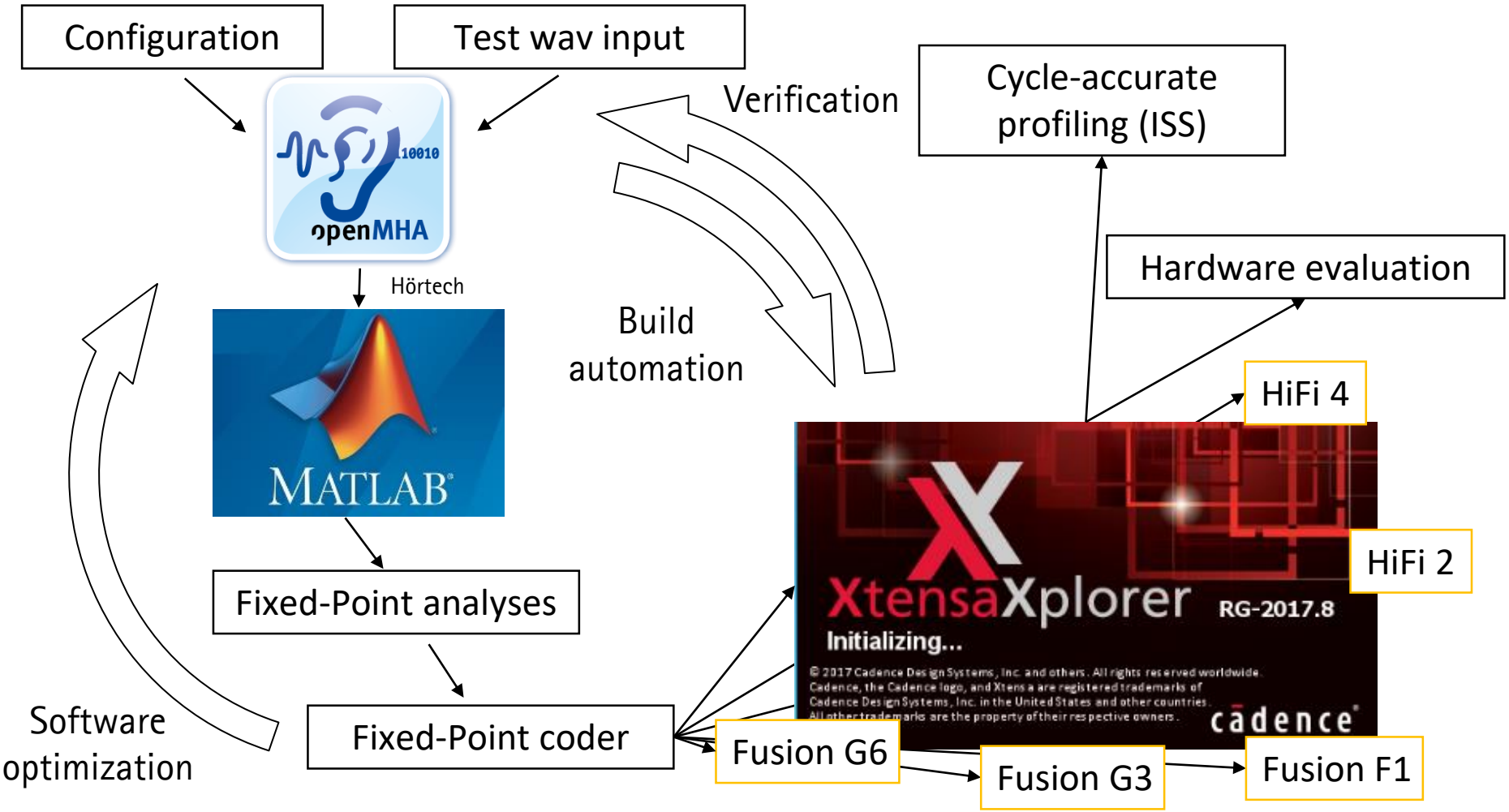


- Noise reduction
- DOA-SVM-MVDR
- Feedback cancellation
- Dynamic compressor



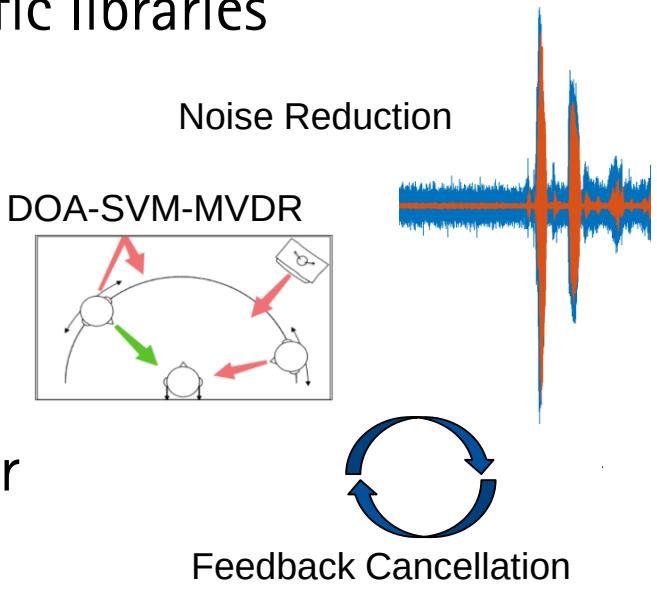


Framework
Fixed–Point Code And DSE Flow

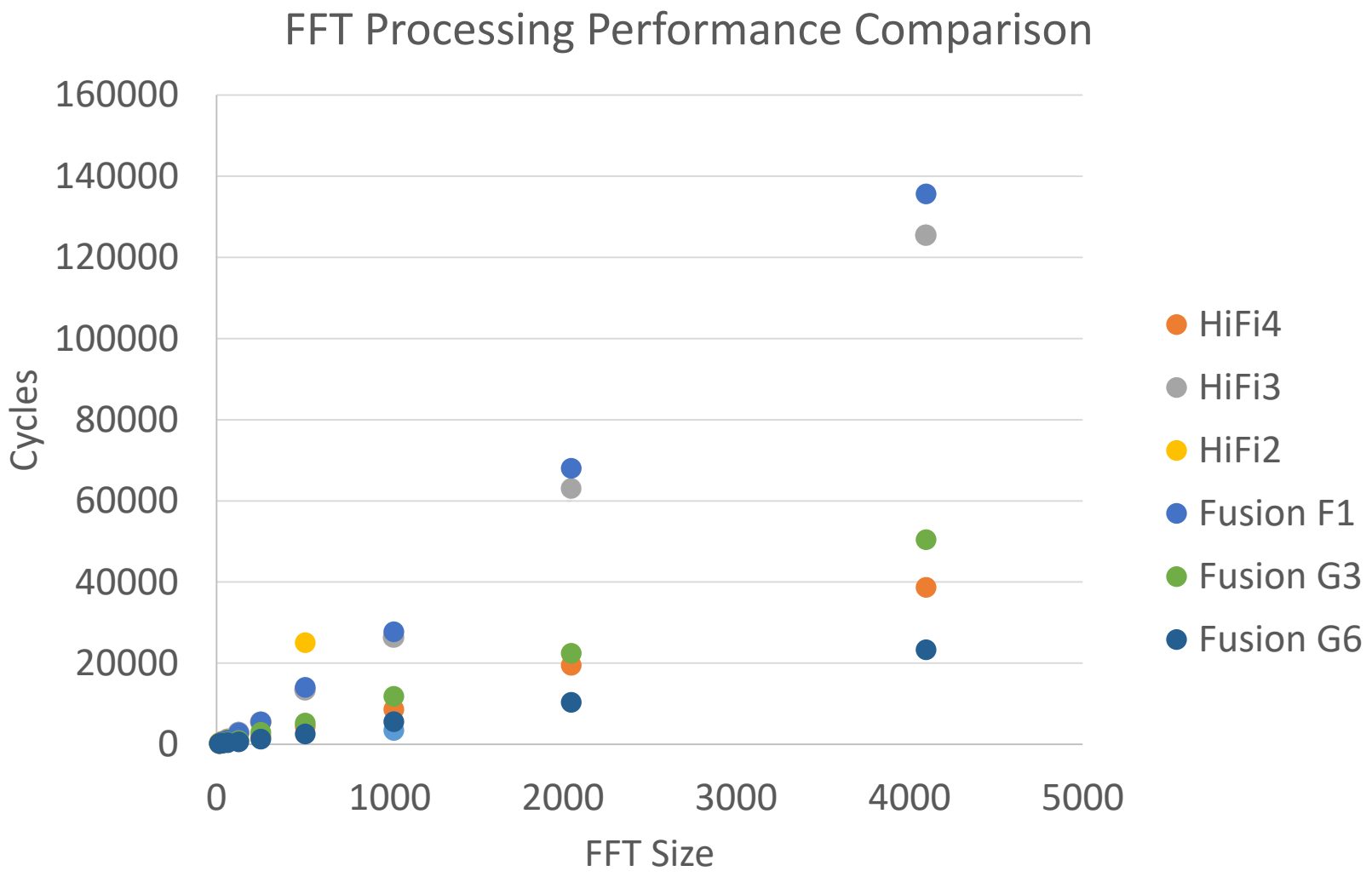


Profiling And Optimization
Techniques

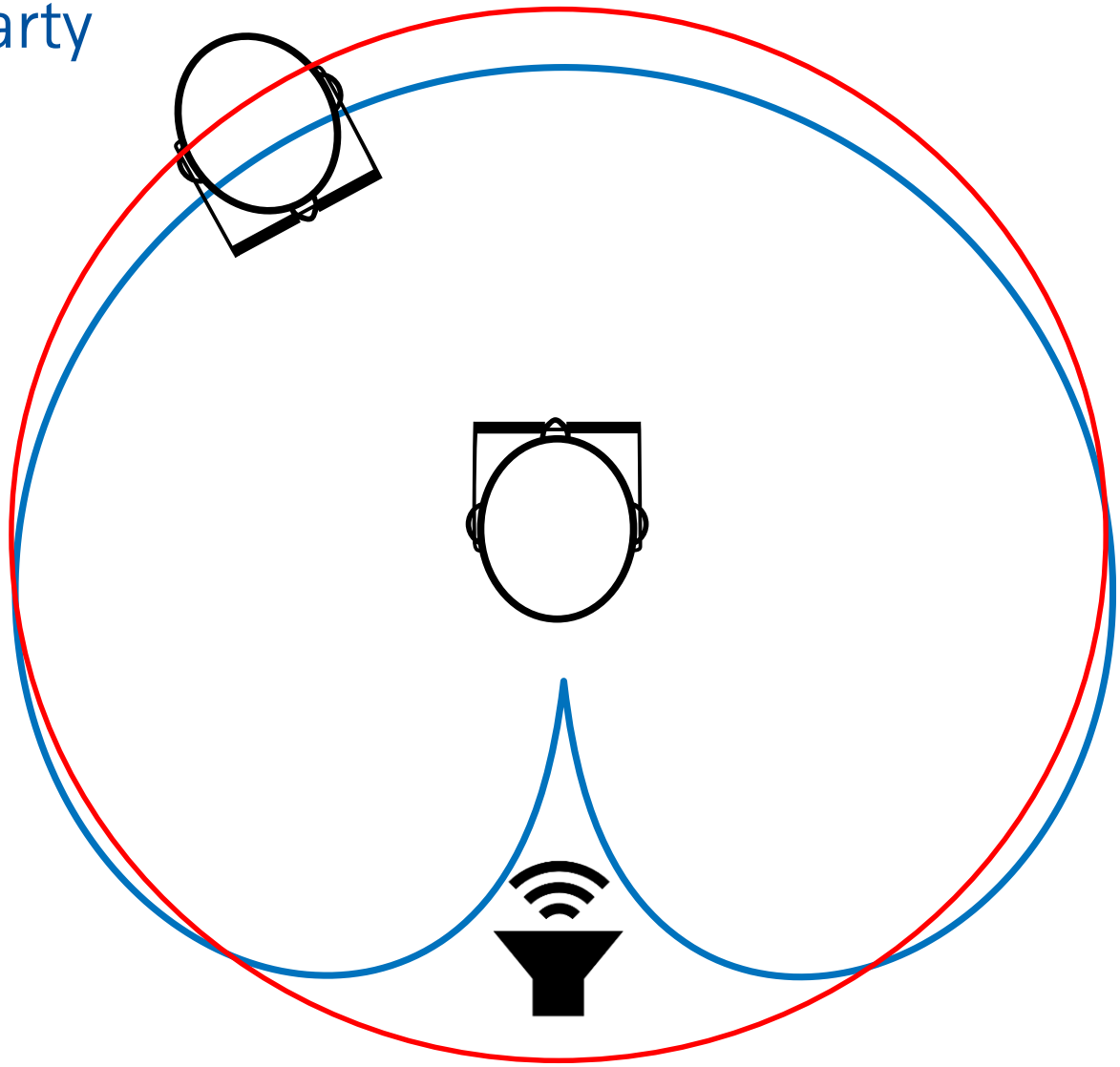
- Bit width reduction
- Optimization of the compiler options (Memory-Alignment, Auto-Vectorize, Loop-Unrolling, IPA ...)
- Use and evaluation of processor specific libraries
  - FFT
  - Division
  - Exponential functions
  - Logarithmic functions
  - Absolut value of a complex number
  - Multiplication



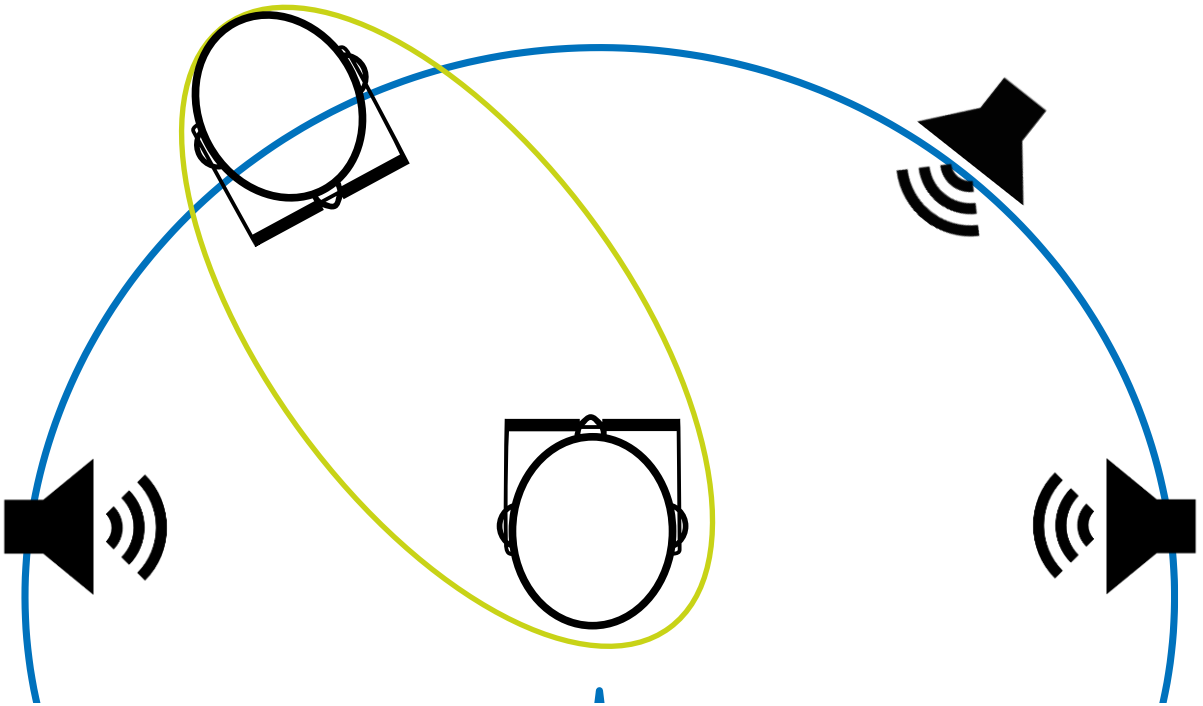
Profiling And Optimization
Processor Specific Libraries

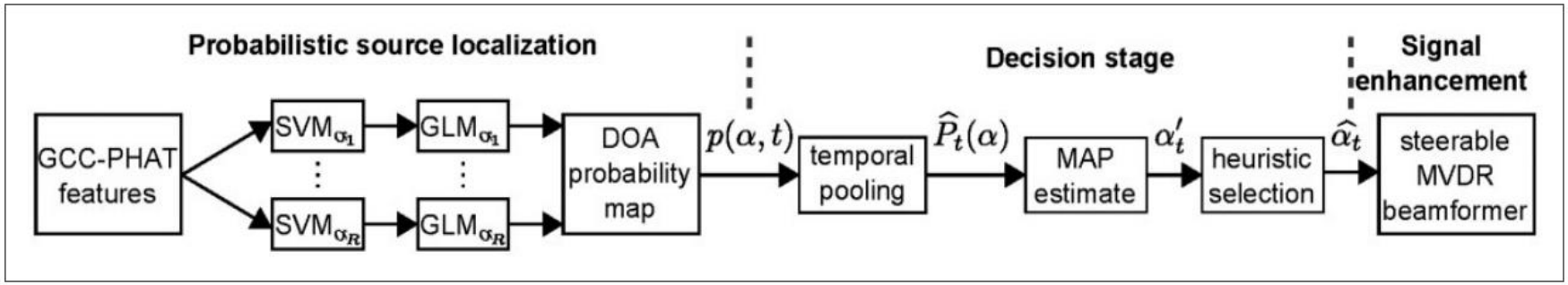


Beamforming
Cocktail Party



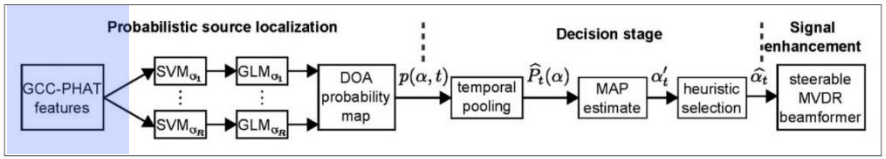
Binaural Steering Beamformer

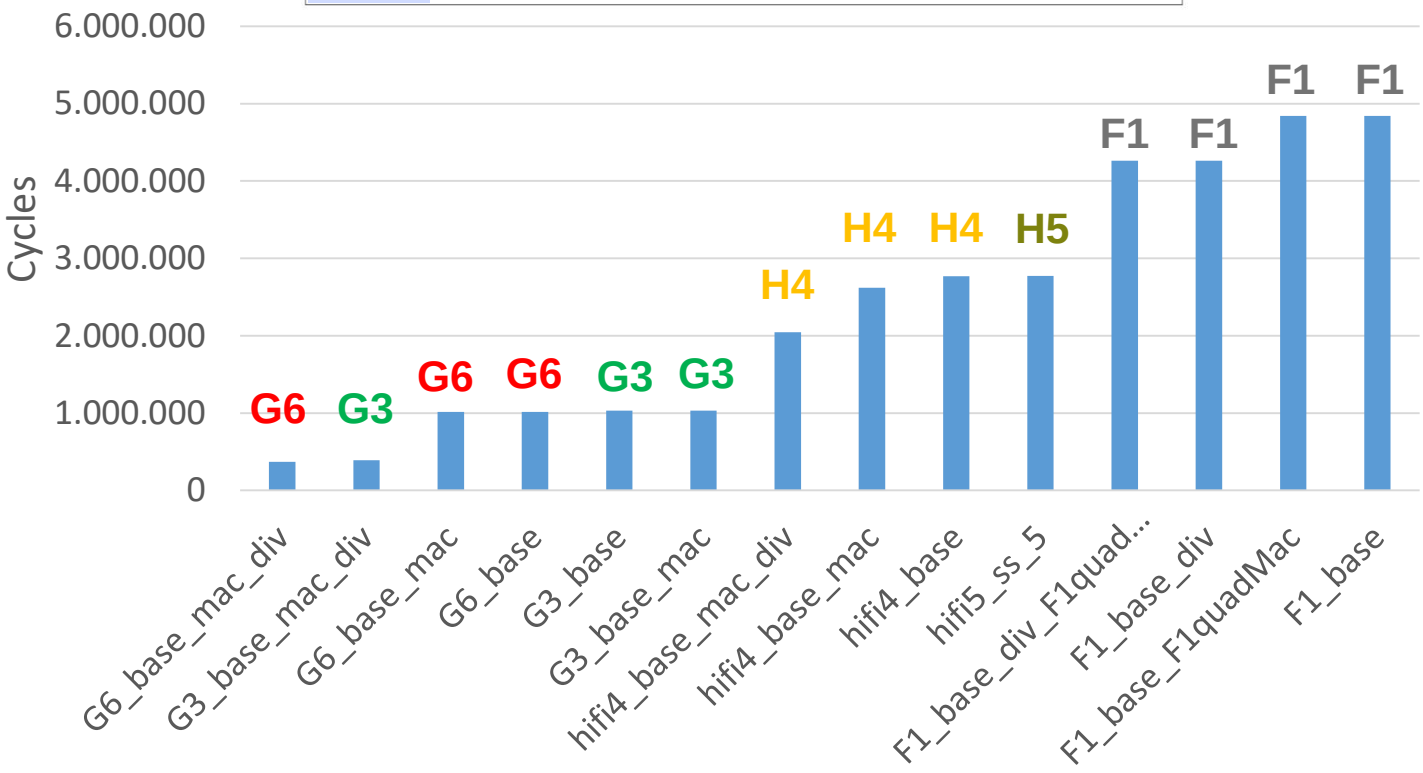




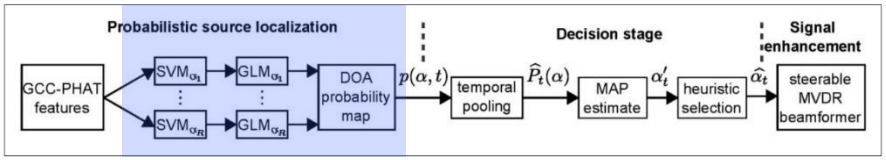
Adiloğlu, Kamil, et al. "A binaural steering beamformer system for enhancing a moving speech source." Trends in hearing (2015)

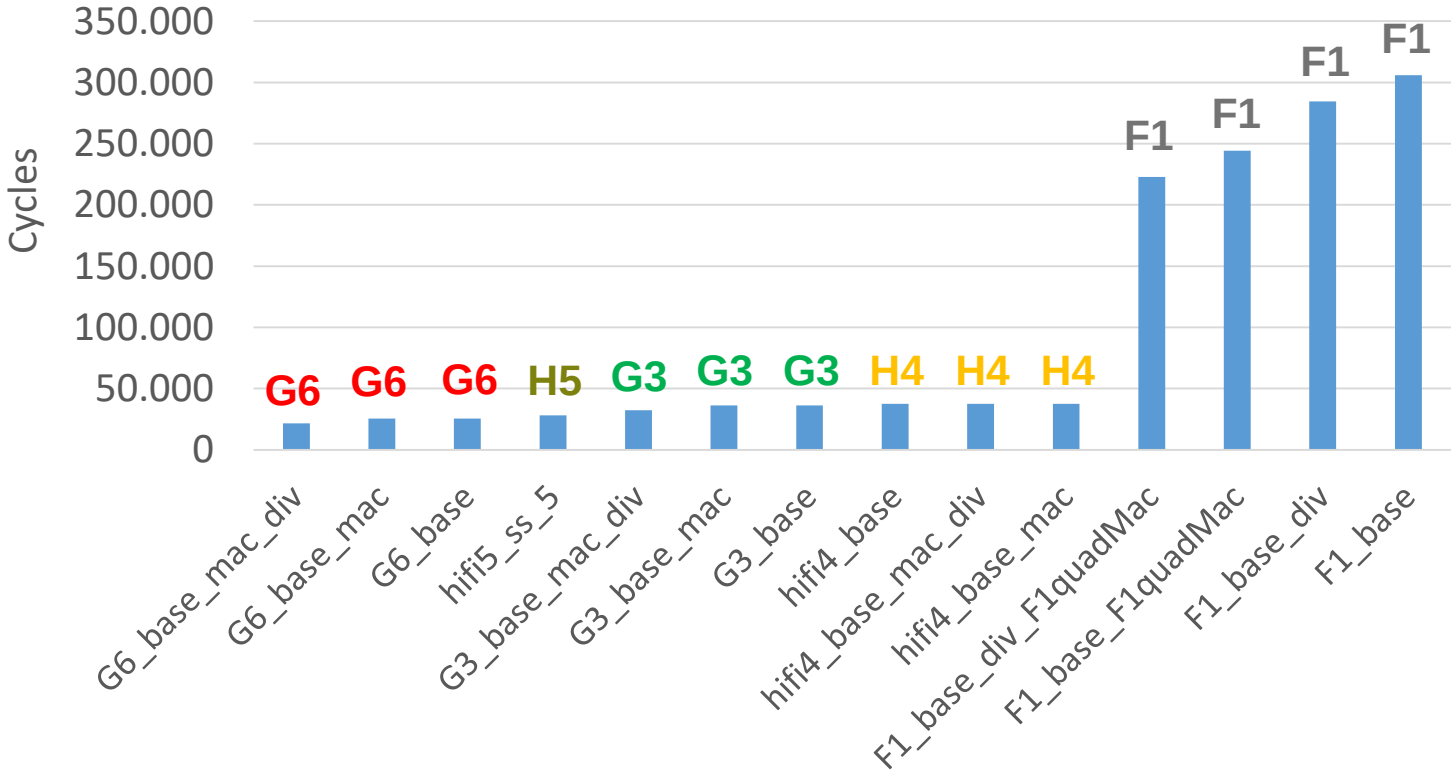
Profiling And Optimization
GCC Feature Extraction Profiling Results



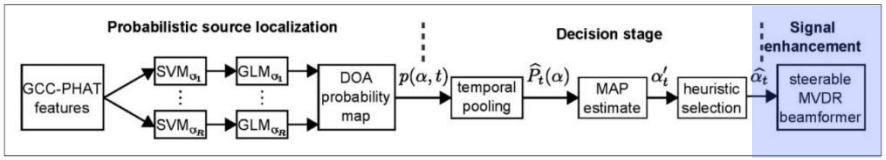


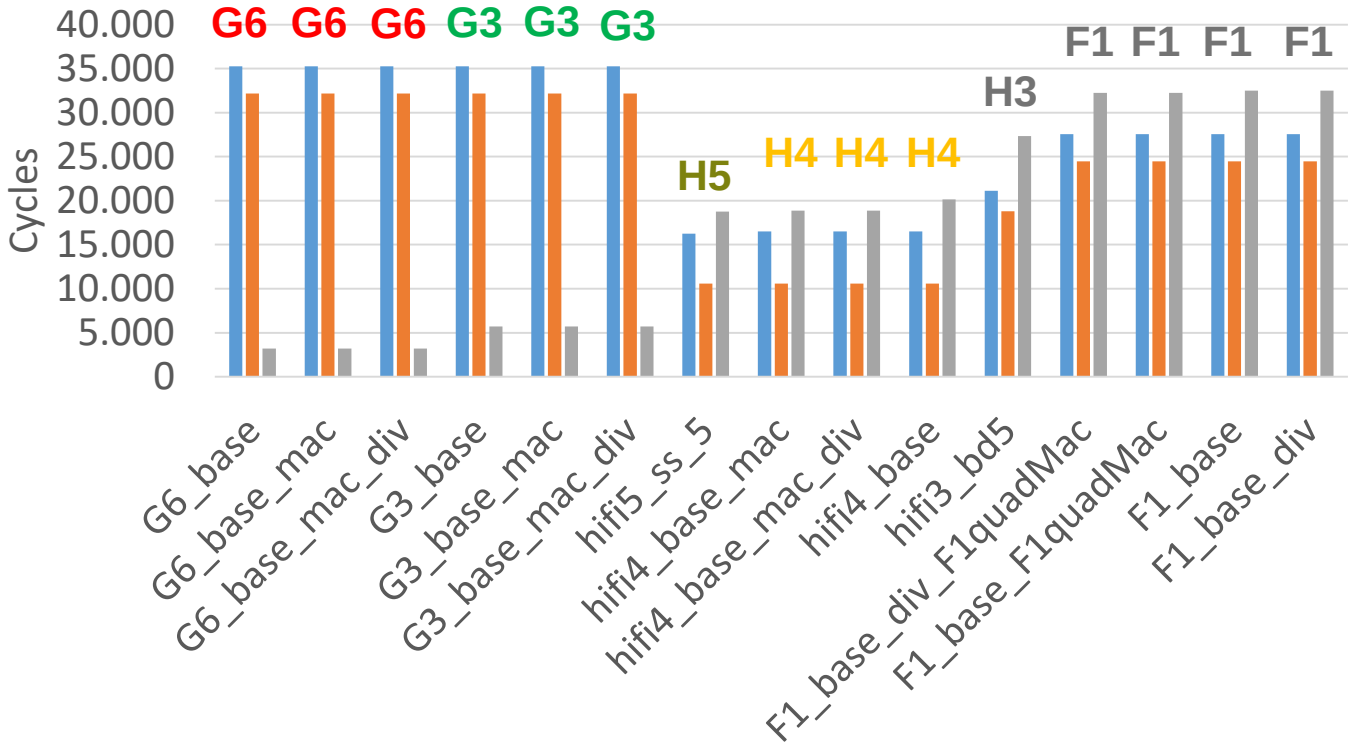
Profiling And Optimization
Support Vector Machine Profiling Results



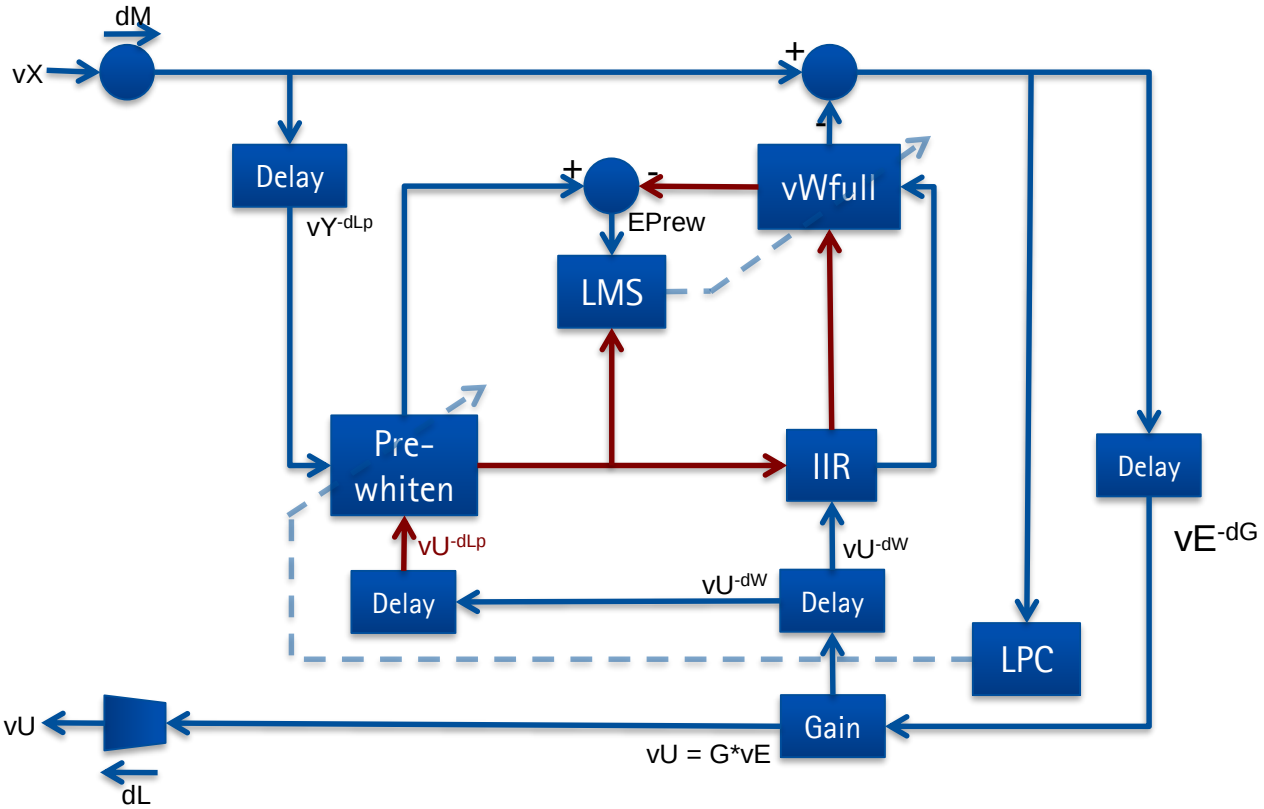


Profiling And Optimization
MVDR Beamformer



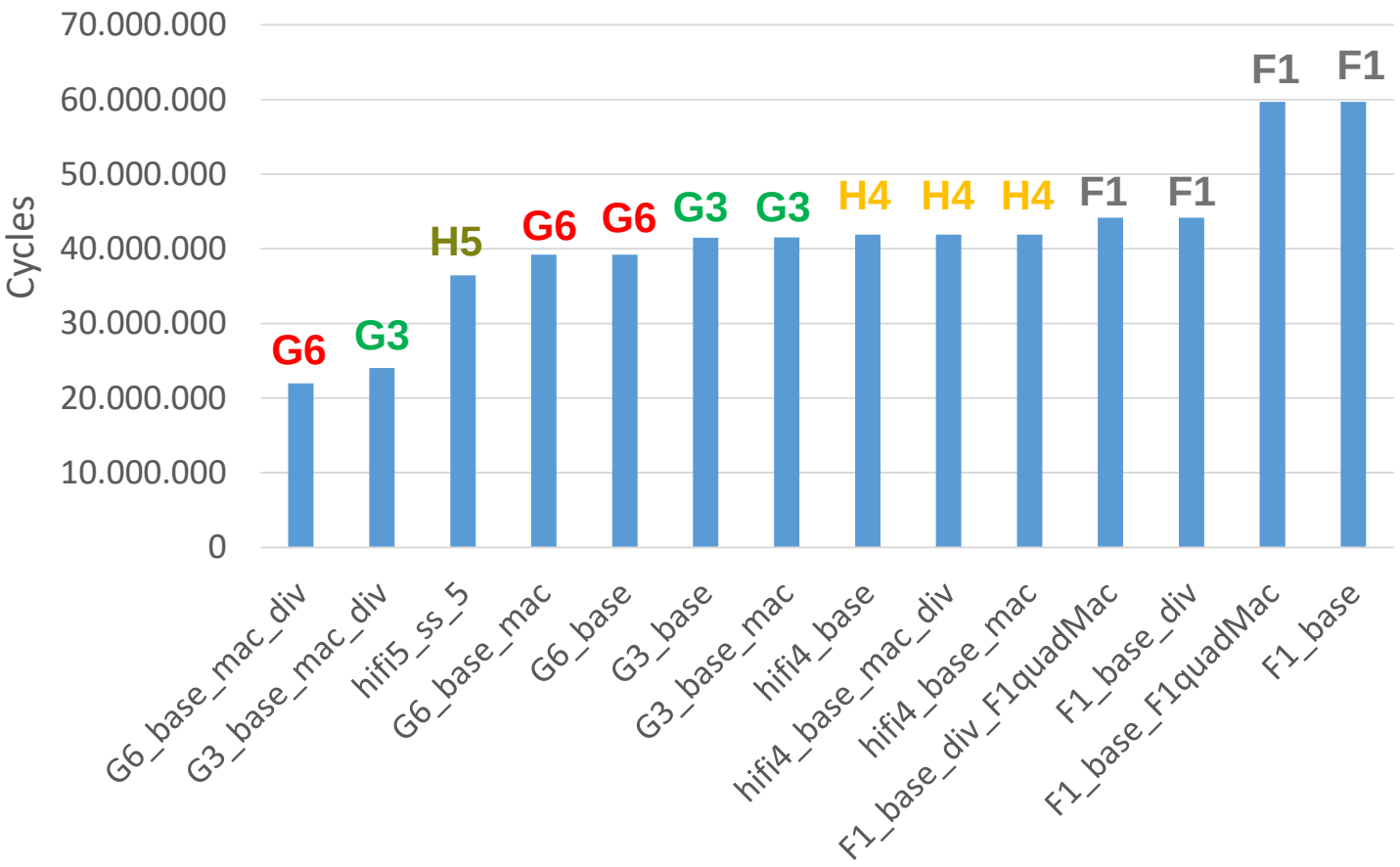


Profiling And Optimization
Adaptive Feedback Cancellation (AFC)

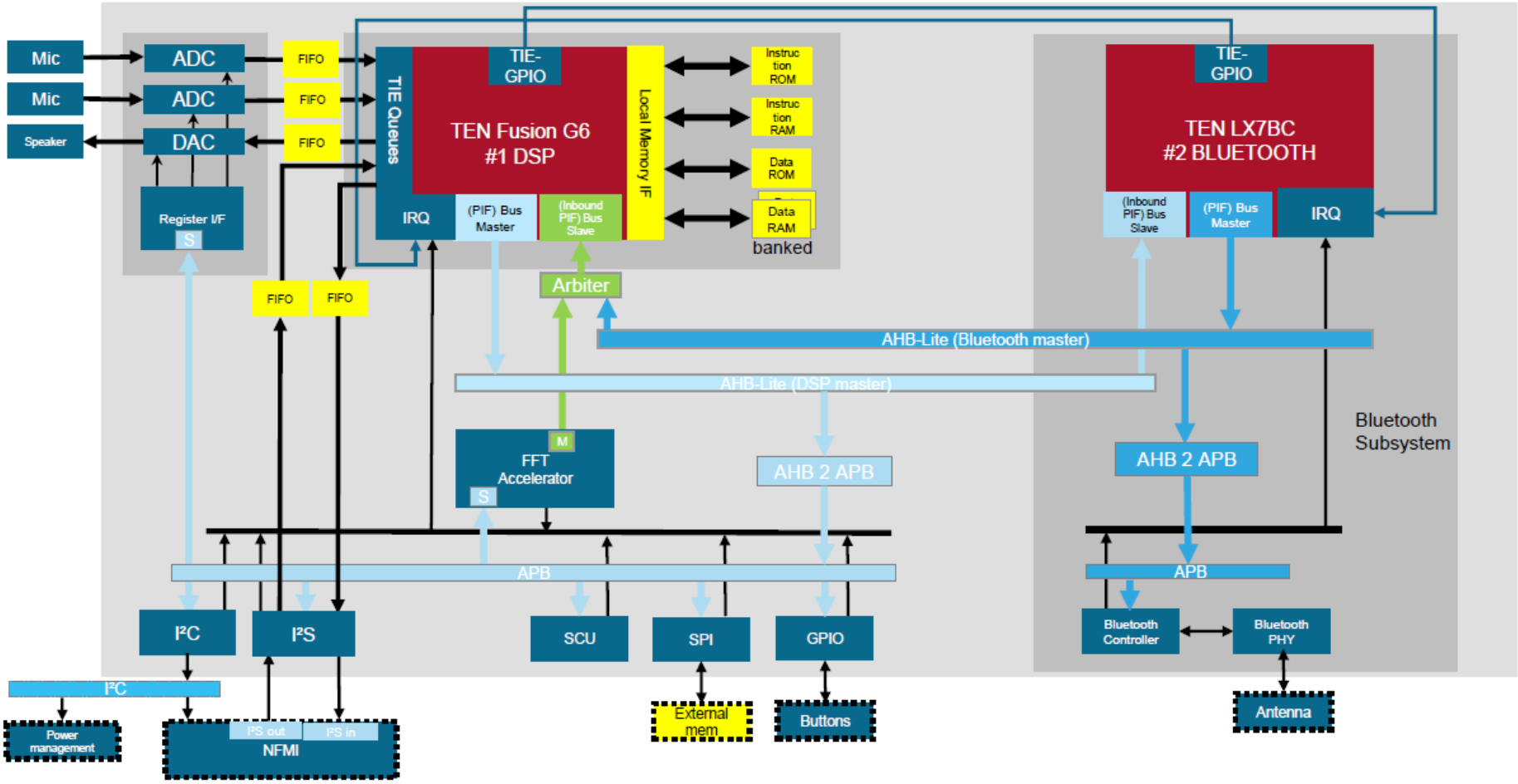


H. Schepker, S. Doclo, "A semidefinite programming approach to min-max estimation of the common part of acoustic feedback paths in hearing aids," IEEE Trans. Audio, Speech, Lang. Process., vol. 24, no. 2, Feb. 2016, pp. 366-377.

Profiling And Optimization
Adaptive Feedback Cancellation (AFC)



Verification with a virtual prototype made by Cadence



F. Kautz, H.-M. Blüthgen, D. Langen, C. Sauer: Smart Hearing Aid Architecture Overview, Cadence Design Systems, 6. SmartHeaP Projekttreffen in Erlangen, 13.08.19

Outline And Future Work

- Challenges for Hearing aid processors
- Framework for design space exploration
- Profiling and optimization results
- ➢ Decision made for Tensilica Fusion G6

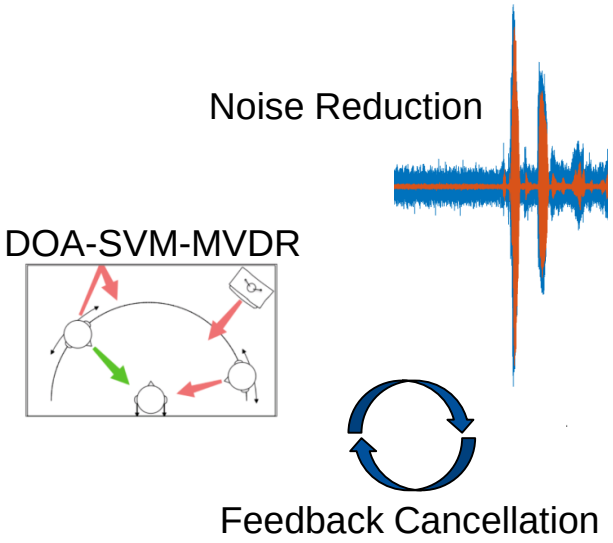
- Continue the optimizations
  - Code optimizations
  - TIEs
- Tape out in 22nm FDSOI

Thank you for your attention!

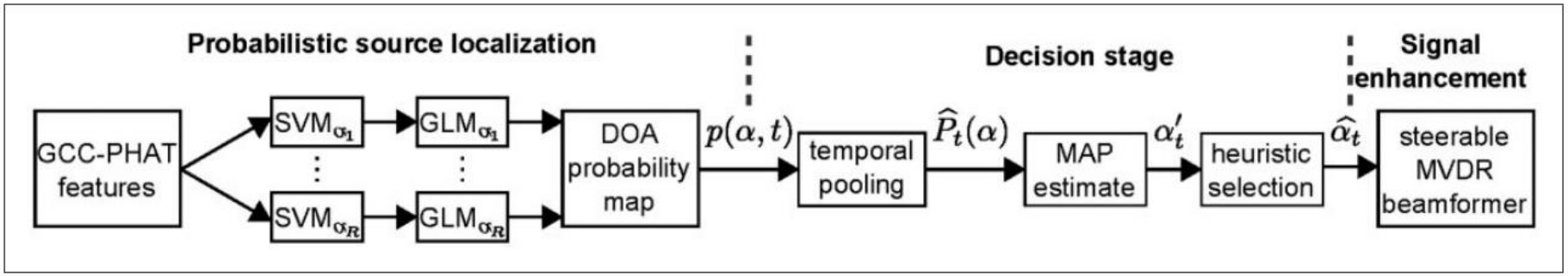
Further Questions?

Software Mapping und Optimierungen

- Erstellung und Nutzung prozessor-spezifischer Bibliotheken
  - FFT
  - Division
  - Exponentialfunktion
  - Logarithmusfunktion
  - Betrag einer komplexen Zahl
  - Multiplikationen
- Reduktion der Wortbreite
- Optimierungen für den Compiler (Memory-Alignment, Auto-Vektorisierung, Loop-Unrolling, IPA ...)



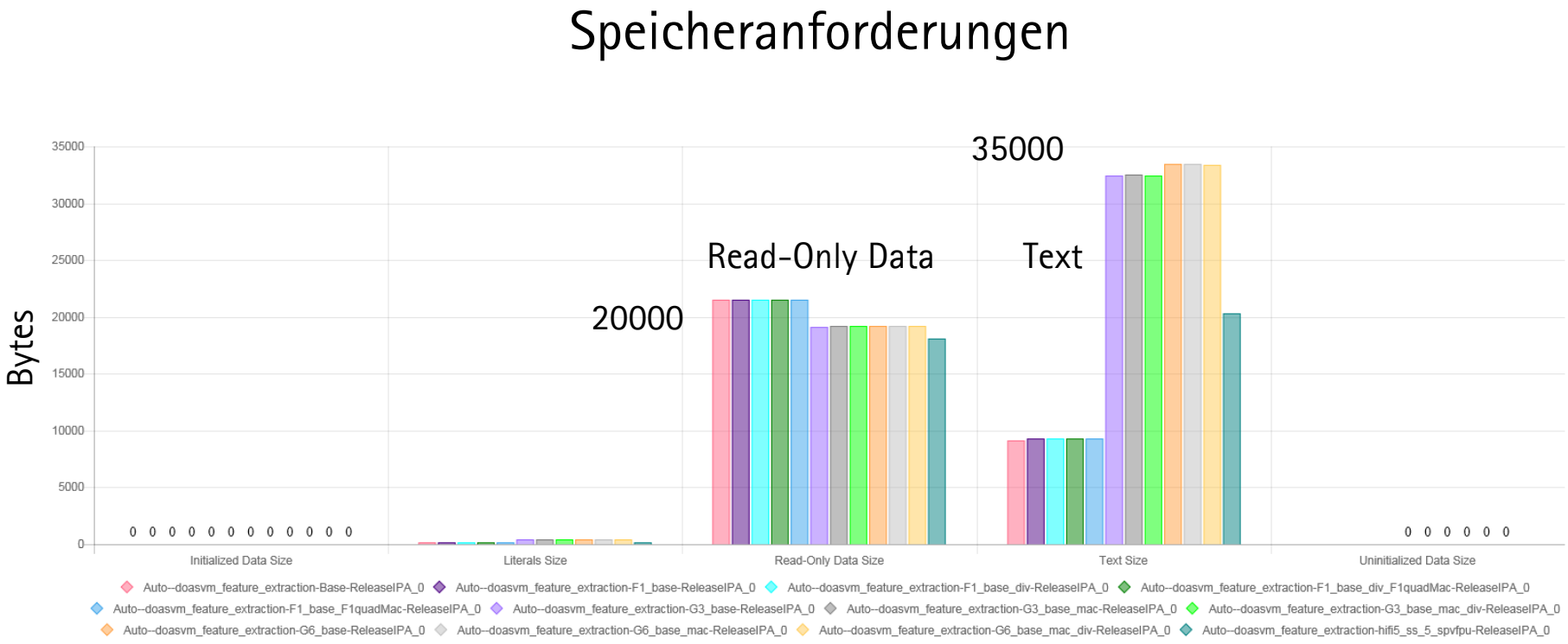
A Binaural Steering Beamformer System for Enhancing a Moving Speech Source



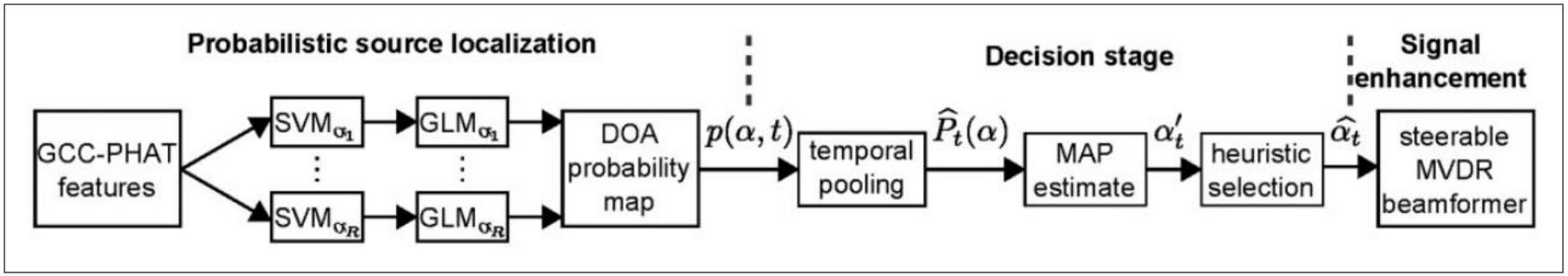
Adiloğlu, Kamil, et al. "A binaural steering beamformer system for enhancing a moving speech source." Trends in hearing 19 (2015)

A Binaural Steering Beamformer System for Enhancing a Moving Speech Source

GCC Feature Extraction Profiling Results

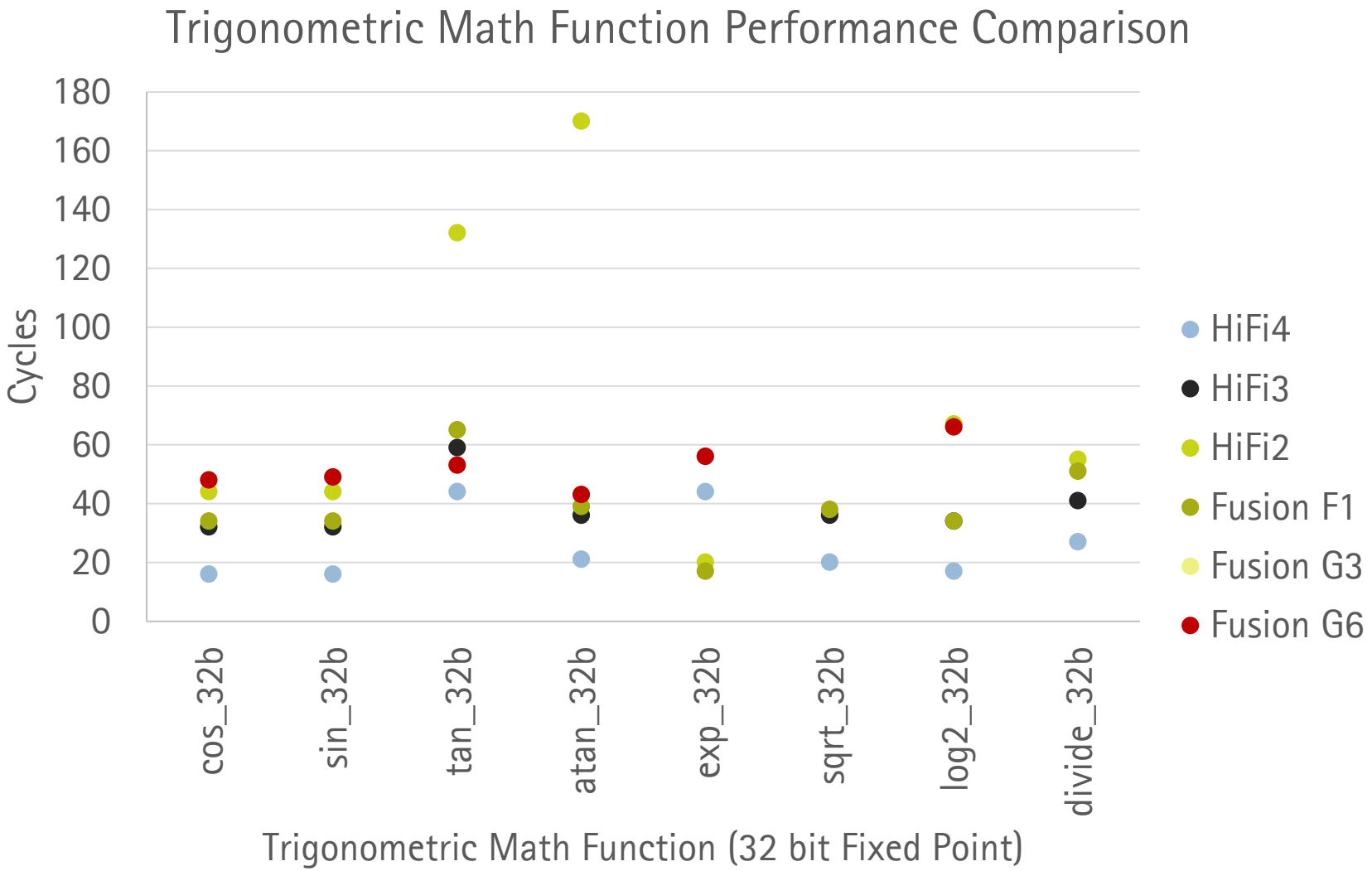


A Binaural Steering Beamformer System for Enhancing a Moving Speech Source



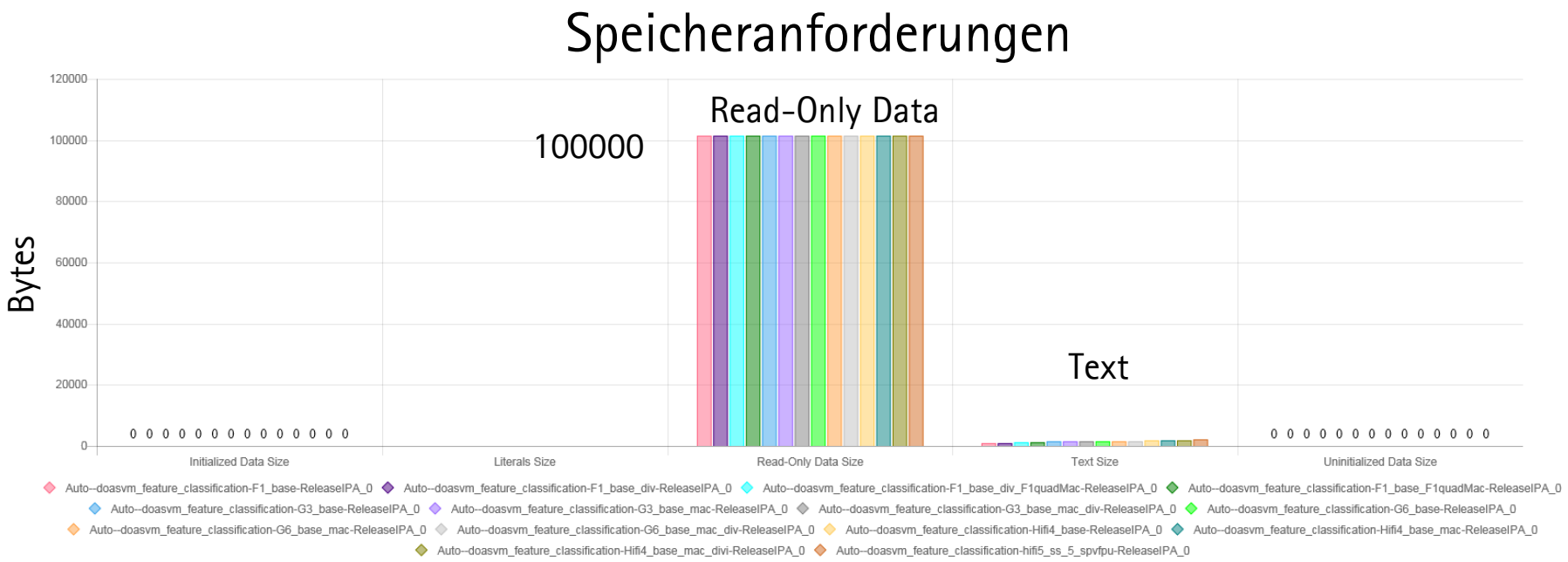
Adiloğlu, Kamil, et al. "A binaural steering beamformer system for enhancing a moving speech source." Trends in hearing 19 (2015)

Prozessor-spezifische Bibliotheken

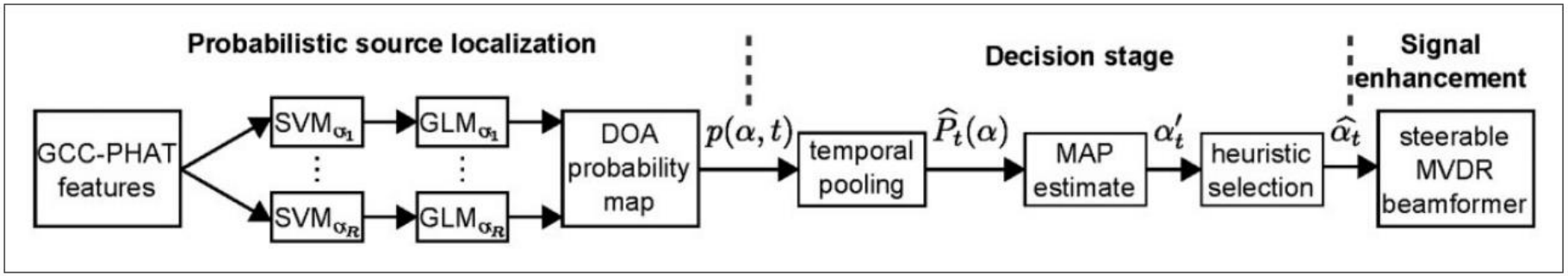


A Binaural Steering Beamformer System for Enhancing a Moving Speech Source

Classification Profiling Results



A Binaural Steering Beamformer System for Enhancing a Moving Speech Source

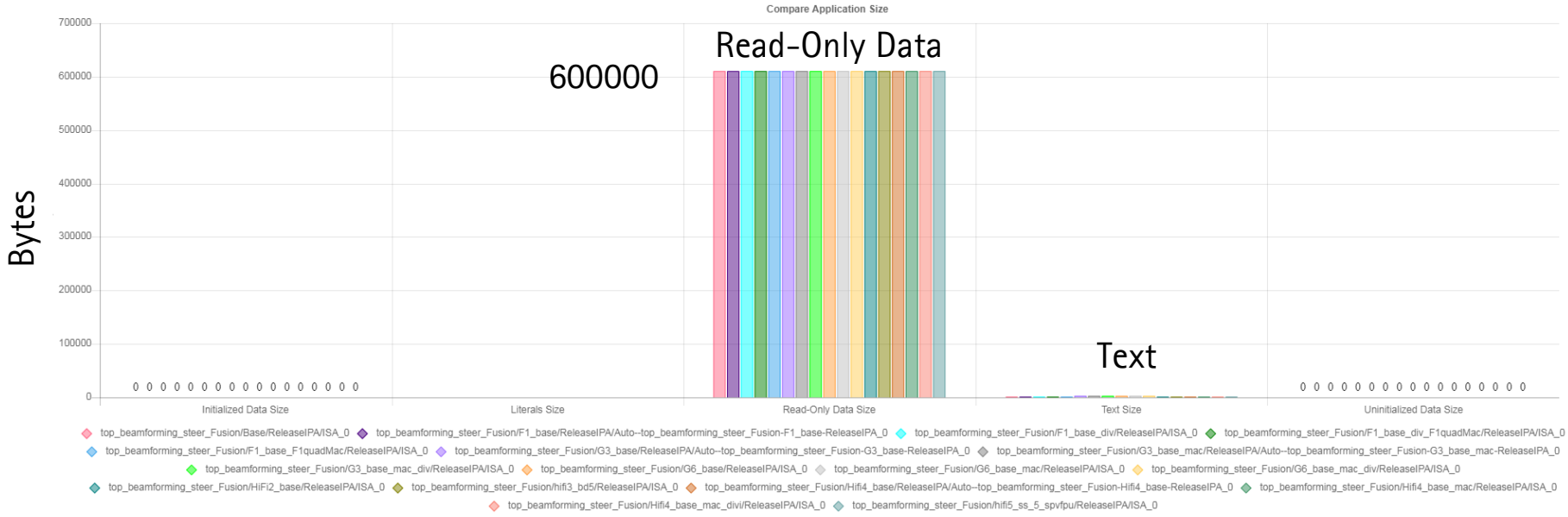


Adiloğlu, Kamil, et al. "A binaural steering beamformer system for enhancing a moving speech source." Trends in hearing 19 (2015)

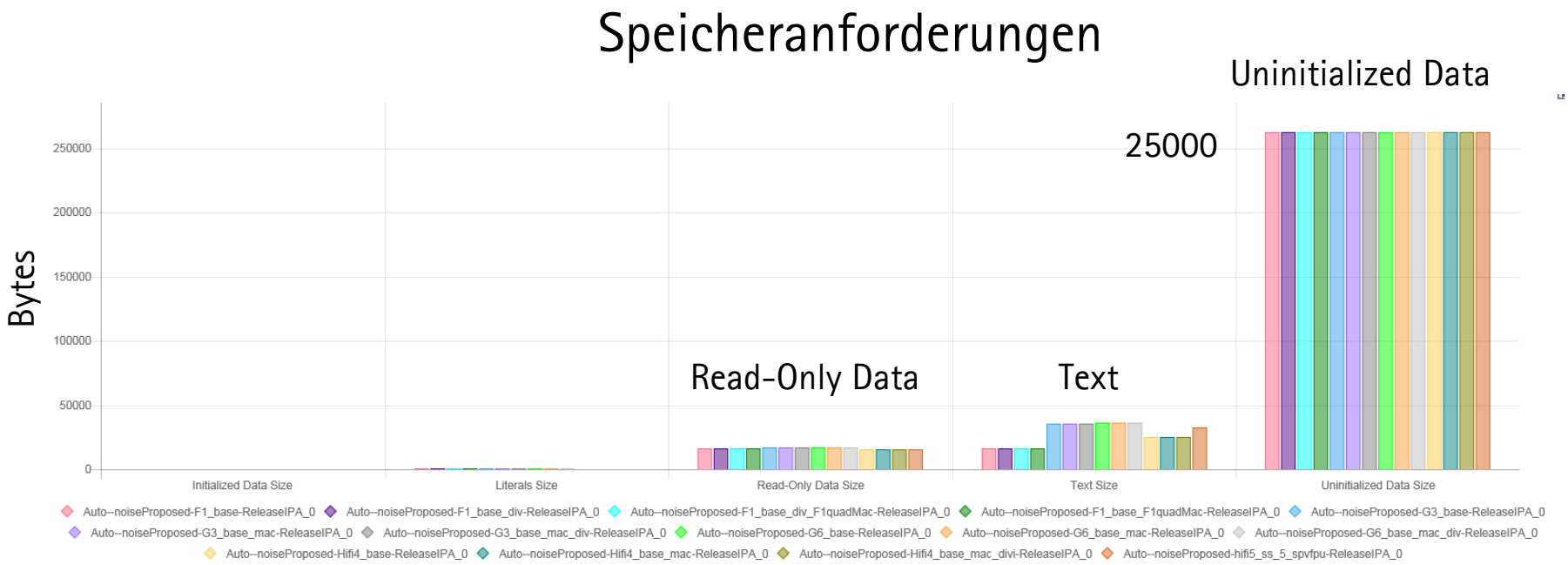
A Binaural Steering Beamformer System for Enhancing a Moving Speech Source

MVDR Beamformer

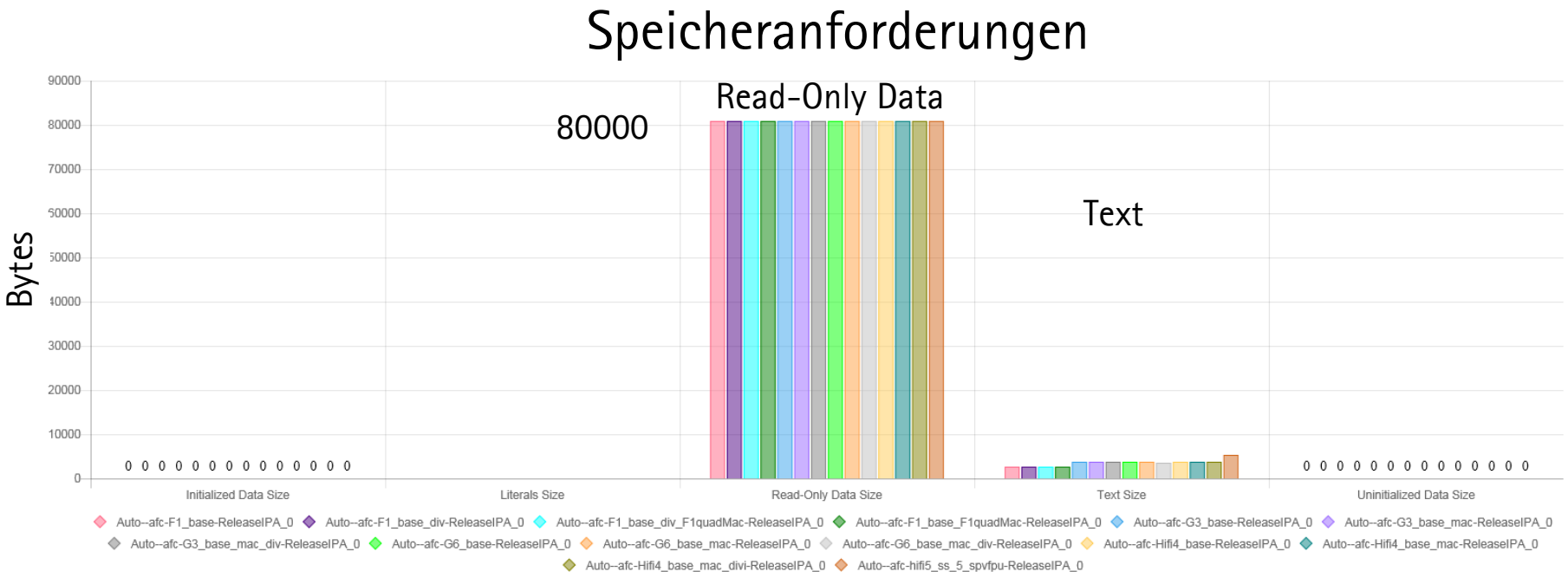
Speicheranforderungen



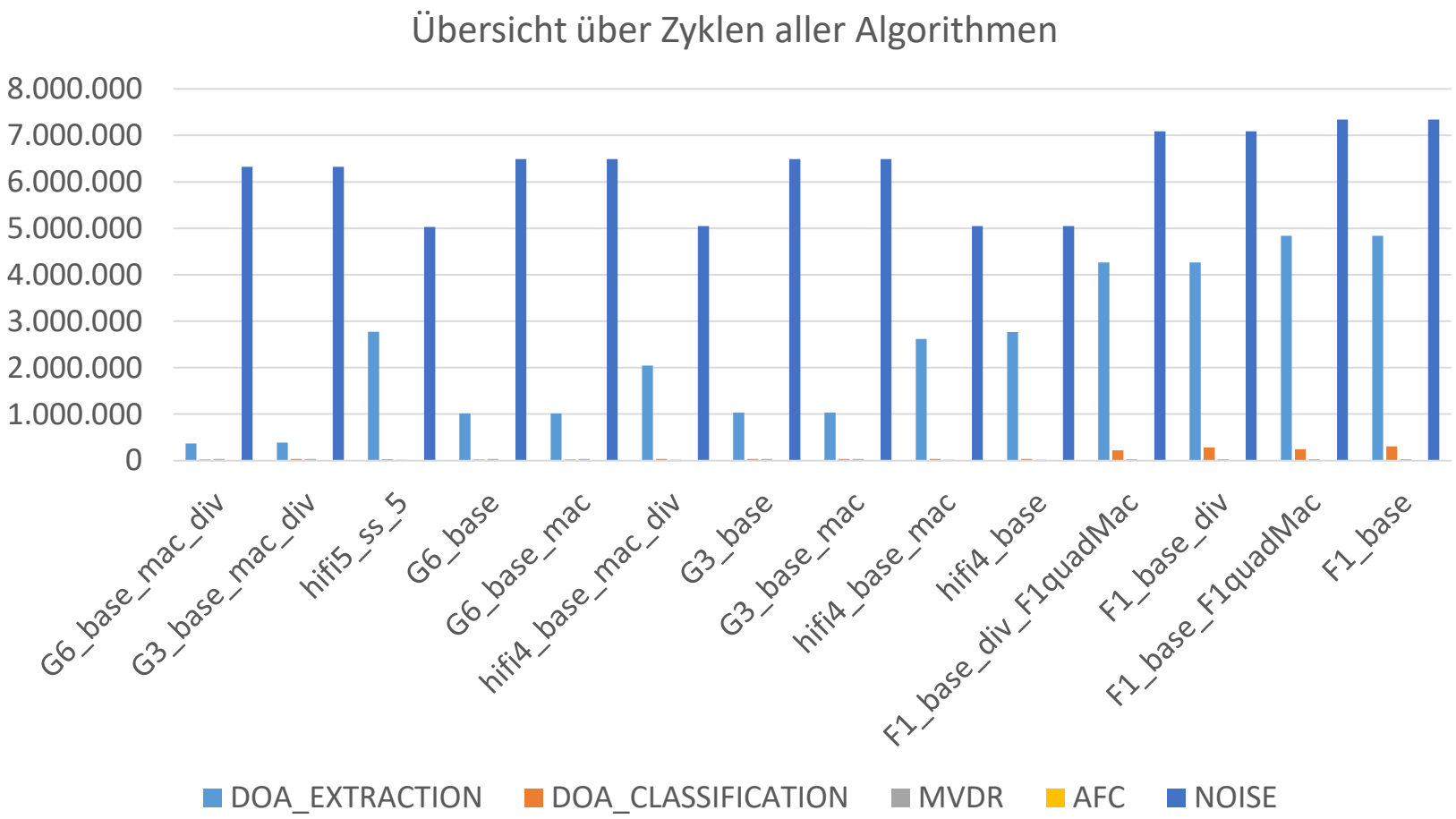
Unbiased MMSE-based Noise Power Estimation with Low Complexity and Low Tracking Delay



Feedback Algorithmus (AFC)

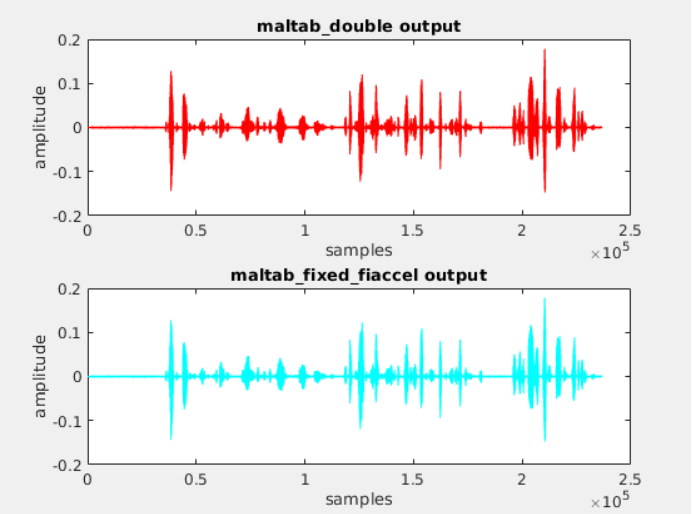


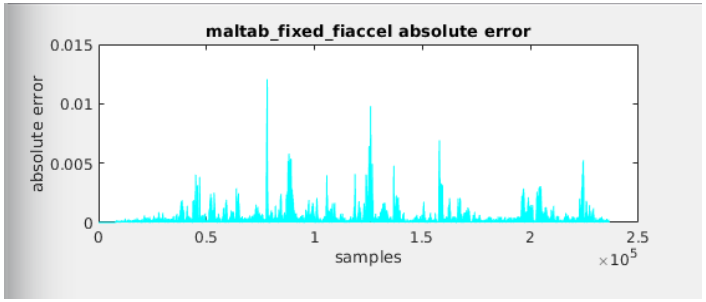
Zusammenfassung der Profiling Ergebnisse



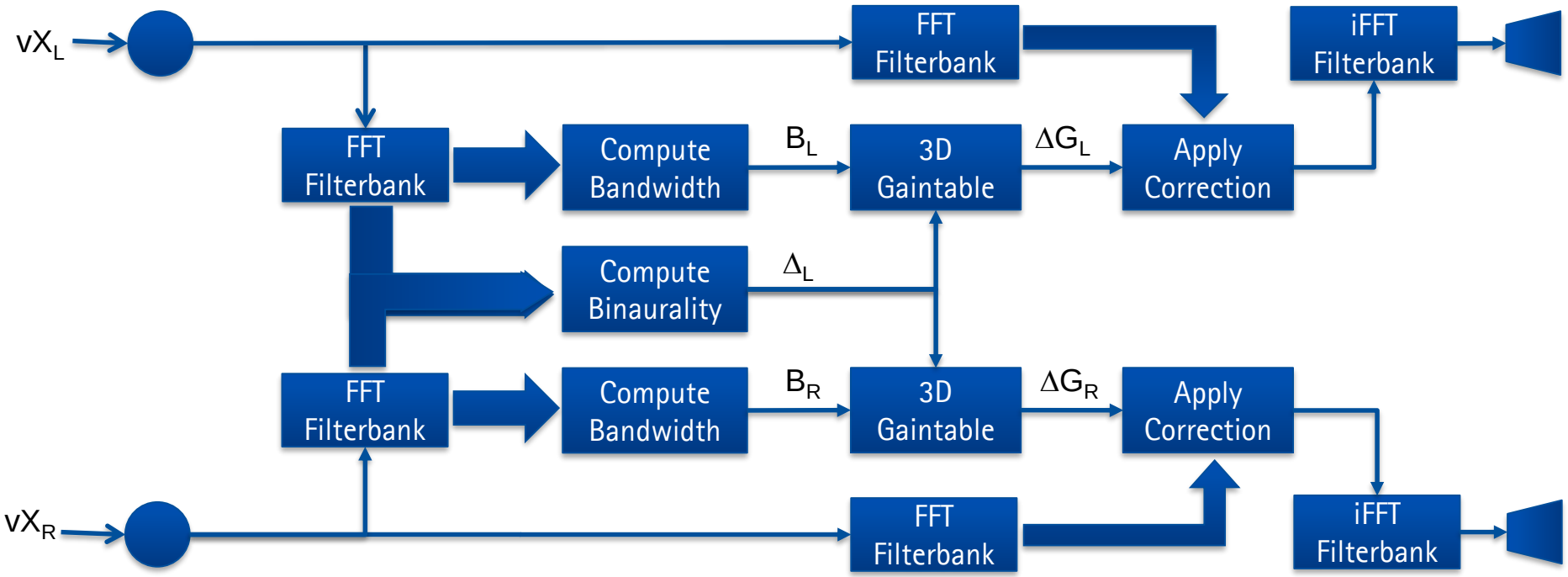
Verifikation of Soft- and Hardware

- Verifikation ist Teil der Algorithmen-Implementierung
- Verifikation werden automatisch generiert:
  - Unit-Tests des Master Hearing Aid
  - Master Hearing Aid als Referenz-Implementierung
  - Matlab Implmentierung in Floating Point und Fixed Point (Bit-genau)





Binauraler breitbandiger Dynamikkompressor



- Oetting, Dirk; Adiloğlu, Kamil

Analyse der Energieaufnahme der Hardwarekomponenten zur Energieeffienzabschätzung

- Hardware Implementation Flow ist aufgesetzt
- GF 20 nm PDK ist verfügbar
- INVECAS IP Bibliotheken sind angefragt
- Racyics NDA unterzeichnet

Conclusion and Further Work

- 3 reference algorithm implemented on all Tensilica Cores
- Applied software optimizations
- Profiling results indicate that Fusion G6 might be a good choice

Further work

- Apply DSE flow for last reference algorithm
- Integrate Hardware acceleration
- Power Analysis

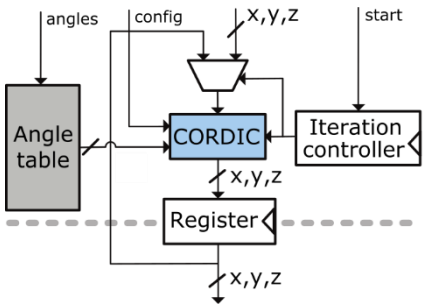
AP5 – Design und Optimierung der ISA des Prozessors
AP6 – Design von dedizierten Akzeleratoren
Statusupdate

- **Erstellung von TIEs**
- **Erstellung spezifischer Akzeleratoren mit Hilfe von VHDL (z.B. für Filterbänke, komplexe arithmetische Funktionen etc.)**

- Erste Ideen existieren
- Implementierung existieren
- Beispiele existieren aus Projekten wie Hearing4All
  - Co-Prozessoren und Funktionelle Einheiten für Audiosignalverarbeitung

CORDIC Coprocessor (COordinate Rotation DIgital Computer)

HW



SW

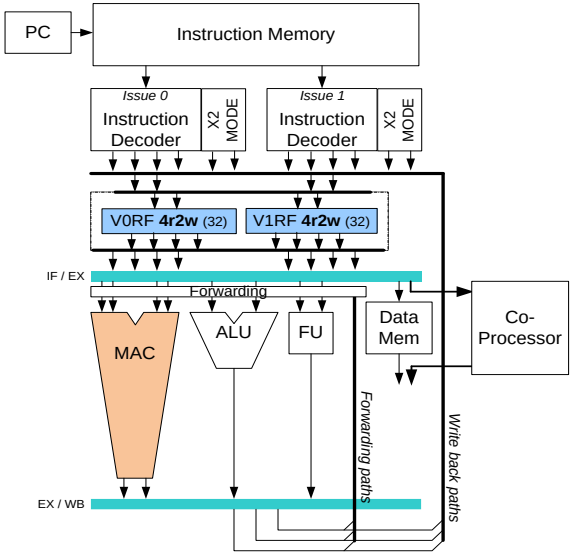
```
MVI_32  VxR70, #1
ANDCS_32 VxR70, VxR70, SHIFT_CNT
SMVI CONDSEL, #0b0010  // COND_Z
MIXL_32 VxR54, VxR53, VxR53
MIXR_32 VxR70, VxR53, VxR53
MVCR_32 VxR54, VxR70
SMVI VxR98, REG_Z_OUT, #0
CONDSEL, #0b0011  // COND_N
SUBICS_32
// Kernel equations
// x = x - sigma * 2^(-iteration) * y
// y = y + sigma * 2^(-iteration) * x
```

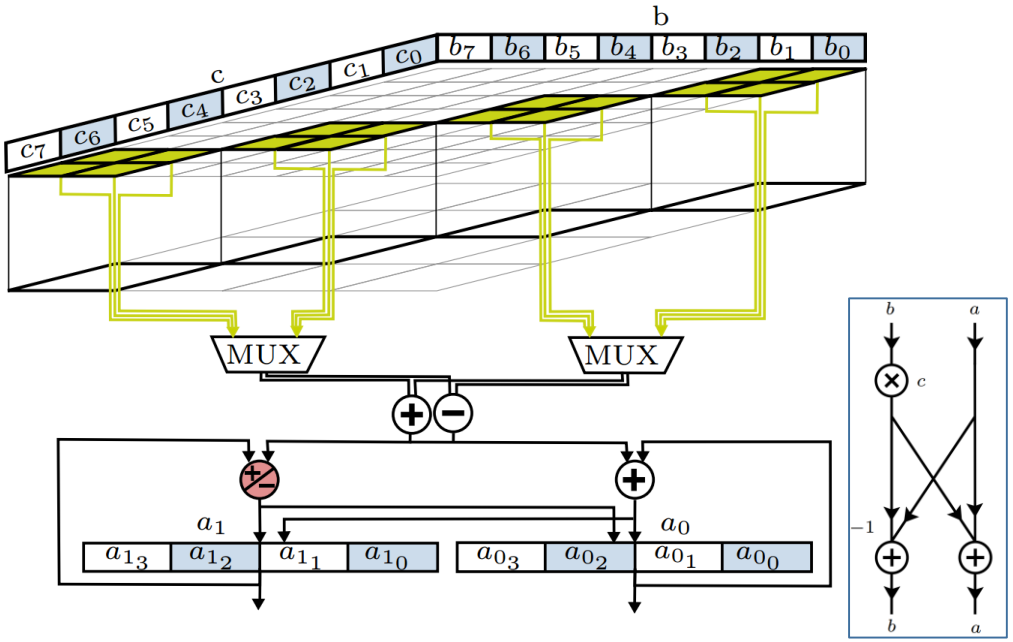
Fast computations of non-linear functions

Hyperbolic and trigonometric operations		Sine	Cosine	Exponential	Natural logarithm	Square root
Cycles	KAVUAKA+CORDIC (HW)	71	71	76	56	59
	KAVUAKA (SW)	621	621	668	664	649
	TI TMS320C6478	1259	1523	1529	1134	341

Complex-valued MAC

- Specialization Techniques





Processor: KAVUAKA	32 Point FFT Cycles	Core Area (*40 nm Low Power TSMC*)
Real-valued SIMD MAC	570	0.237 mm^2
Real- and Complex-valued SIMD MAC and Butterfly Operations	135 (Speedup: 4.22 x)	0.255 mm^2 (Overhead: 7%)

AP7 – Software-Mapping der Referenzalgorithmen auf neue Architektur

- **Optimierung der Referenzalgorithmen**
  - Algorithmische Optimierungen
  - Verwendung der vorhandenen Prozessor-spezifisch optimierten Bibliotheken
- **Umsetzung der Referenzalgorithmen auf die Tensilica-Architektur**
  - Zweiter Algorithmus ist vollständig implementiert
- **Zusammenstellung einer C oder C++ Bibliothek**
  - Automatisch generiert inklusive Datentypen + Tensilica Libs
- **Verifikation des Compilers-Frameworks**
  - Verifiziert mit Bit-genauen Matlab Modellen und dem MHA

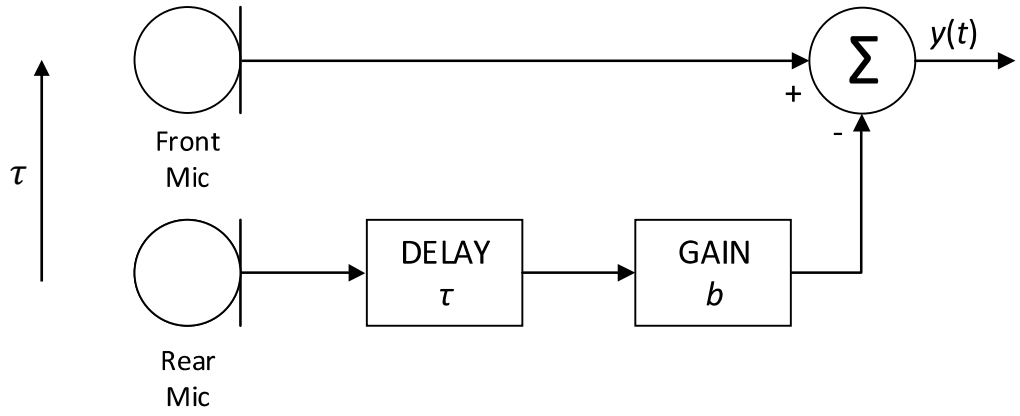
Virtual Prototyping

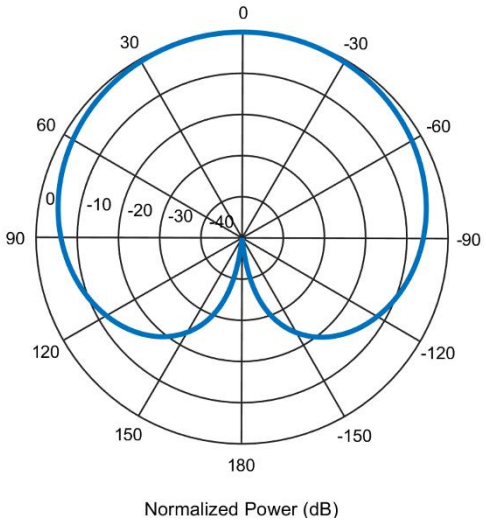
- Cadence hat den virtuellen Prototypen (VP) zur Verfügung gestellt
- Erster Test des VP erfolgreich:
  - 'Hello World' mit dem Fusion G3
  - DOA-SVM-MVDR Debugging im VP über GUI

- Der virtuellen Prototyp (VP) soll für die Hardware Integration (Co-Prozessoren + Interfaces) genutzt werden

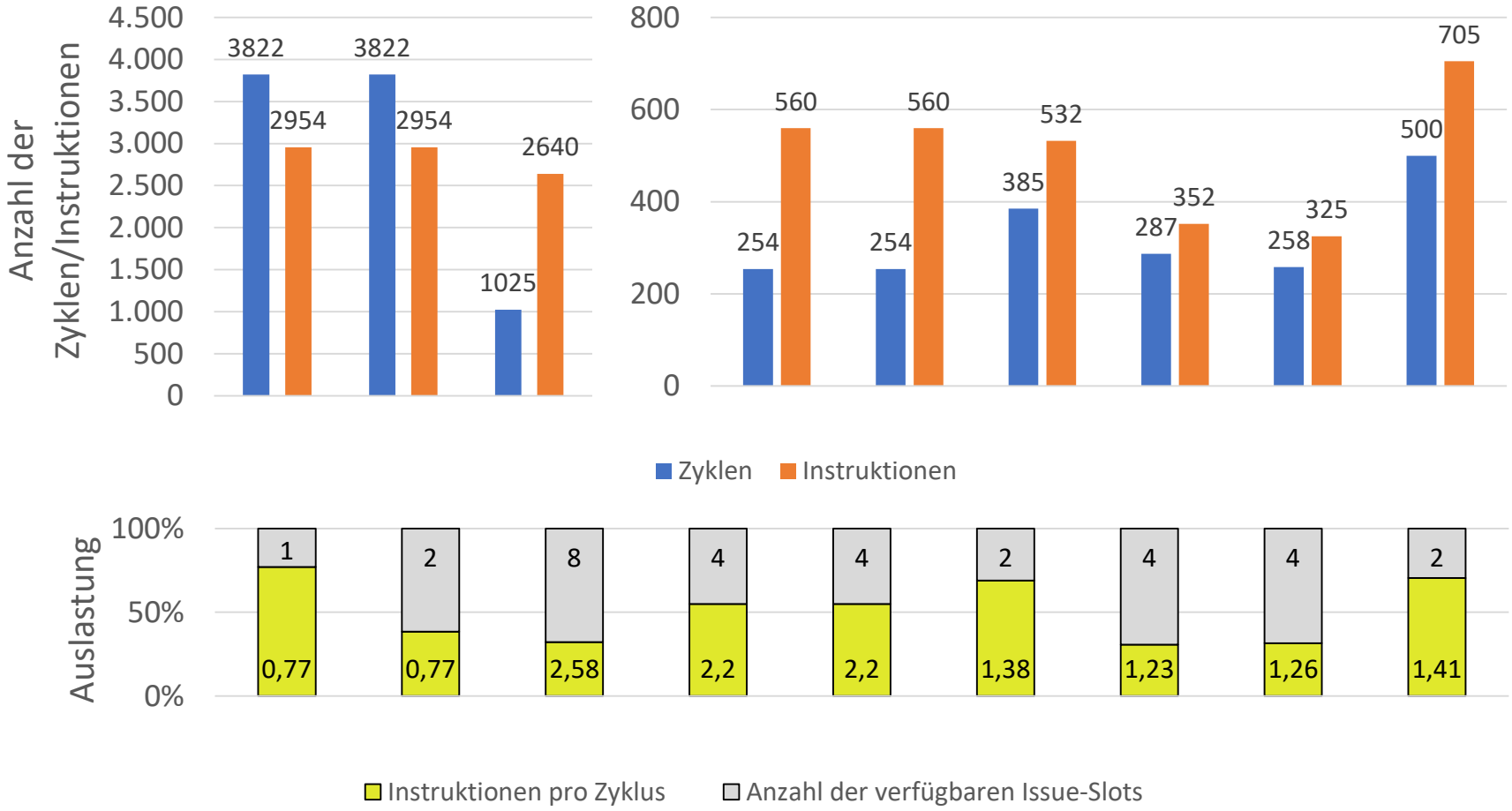
Profiling: Akustische Beamformer
Referenzalgorithmen

- Fixed Beamformer
- Adaptive Gain Beamformer
- **Adaptive Filter Beamformer**
- GSC Beamformer
- MVDR Beamformer



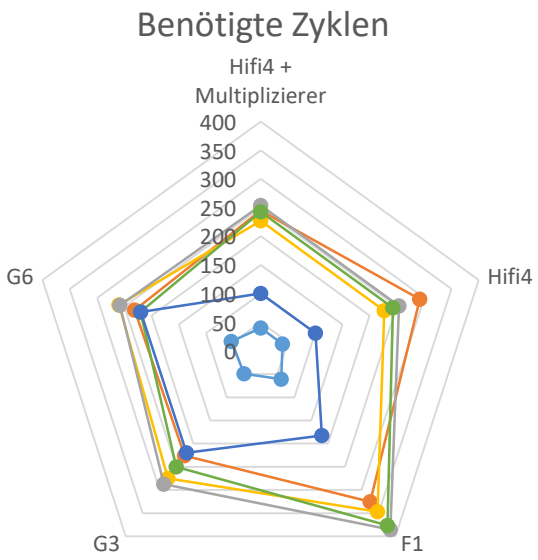


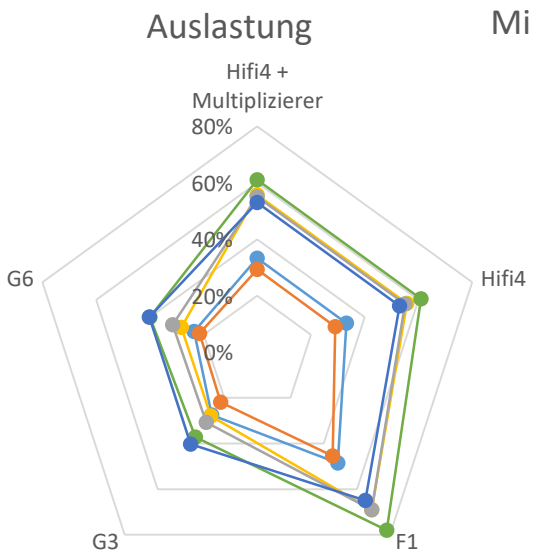
Tradeoff between Beamforming Algorithms and Tensilica Processor Configurations

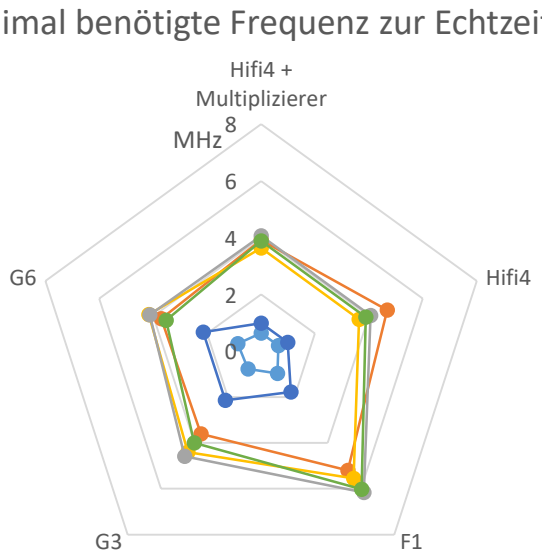


Adaptive Filter Beamformer

Tradeoff between Beamforming Algorithms and Tensilica Processor Configurations







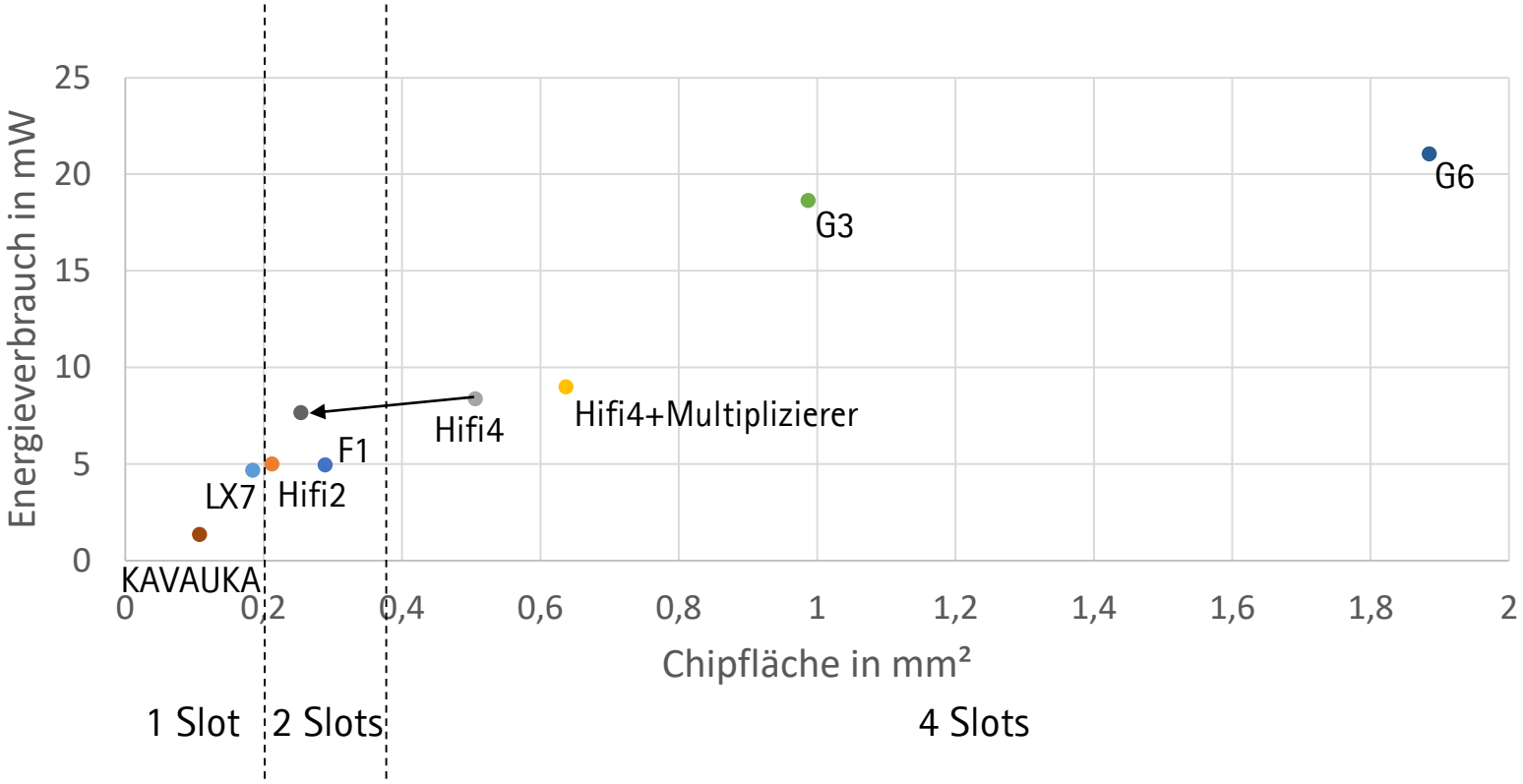
AP4 – Profiling der Referenzalgorithmen
Statusupdate

- **Analyse der Energieaufnahme der Hardwarekomponenten zur Energieeffienzabschätzung**
  - Hardware Implementation Flow ist aufgesetzt
  - GF 22FDX PDK ist vorhanden
  - INVECAS IP Bibliotheken sind angefragt

Requested IPs:

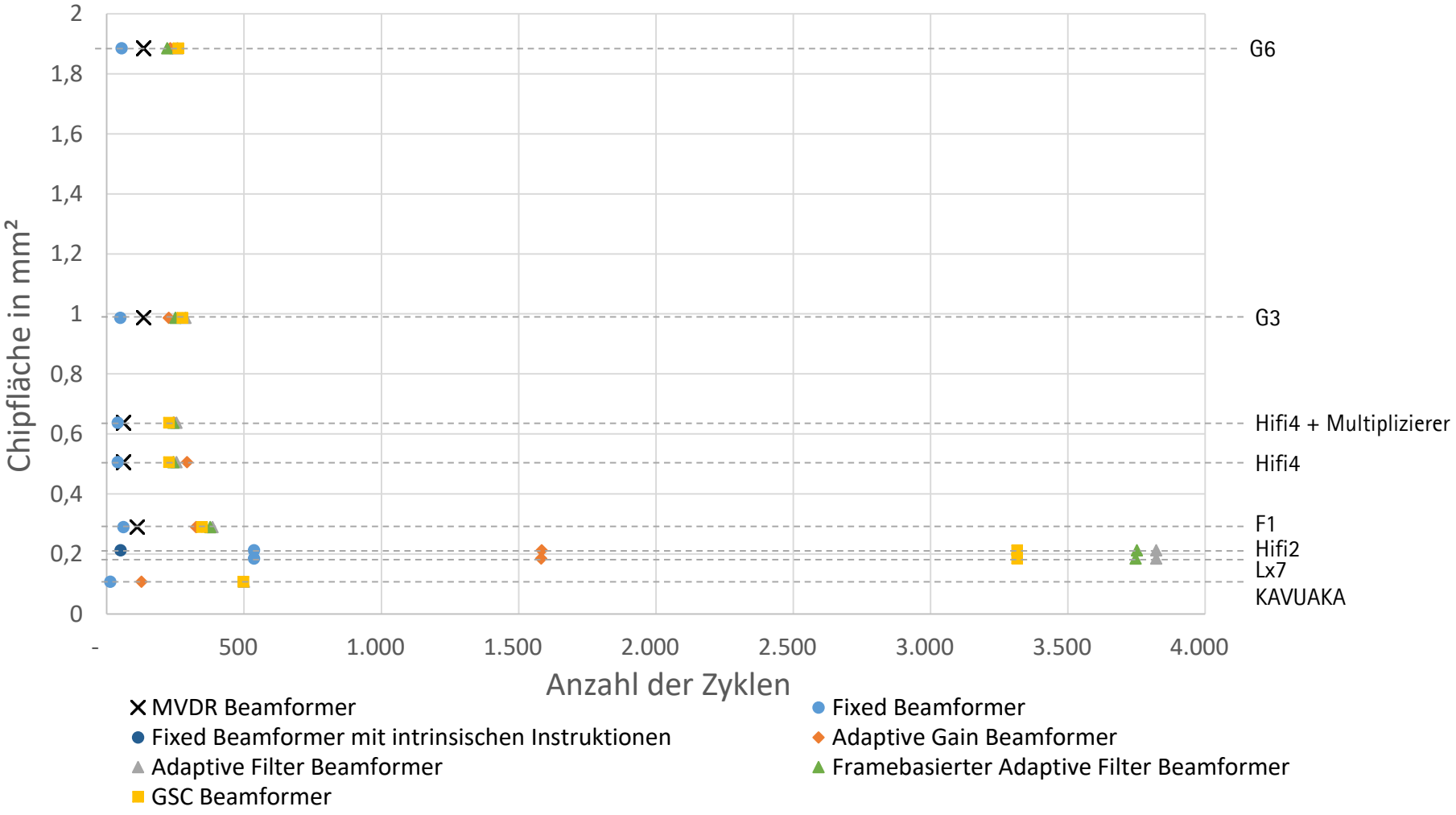
IP Type	Technology	IP	Flavour Name
Foundation IP	22FDX	GPIO	1.2V to 1.8V GPIO
Foundation IP	22FDX	Standard Cells	Ultra Low Power 116CPP 7.5T LVT/SLVT
Foundation IP	22FDX	Low Power Memories	Dual Port SRAM (SDPV)
Foundation IP	22FDX	Standard Cells	Ultra Low Leakage 116CPP 8T UHVT

Fusion vs HiFi
Chipfläche vs Energieverbrauch

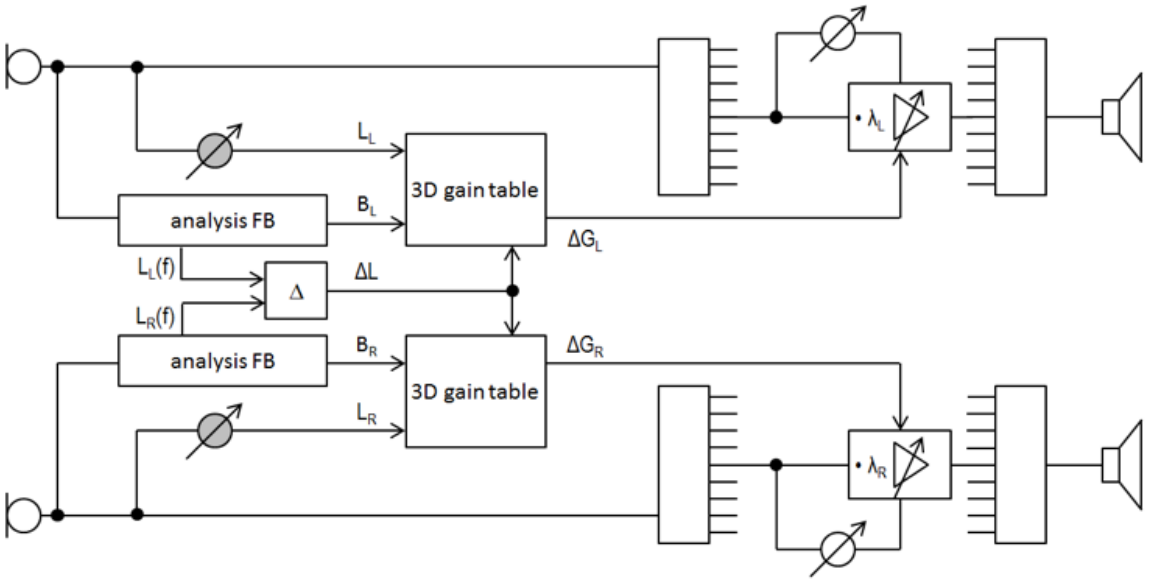


- 40 nm TSMC Technologie bei 100 MHz
- 28 nm HPL Technologie bei 100 MHz

Fusion vs HiFi
Chipfläche vs Performance



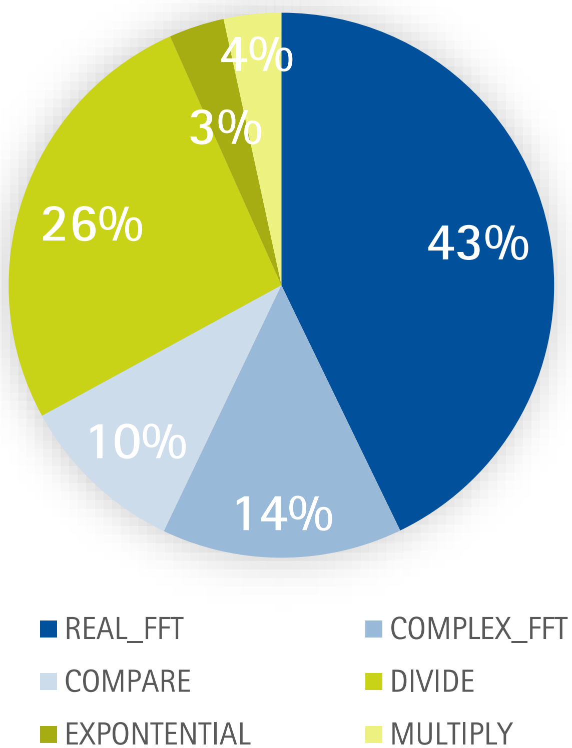
Binauraler breitbandiger Dynamikkompressor



Oetting, D., Hohmann, V., Appell, J. E., Kollmeier, B., & Ewert, S. D. (2018). Restoring perceived loudness for listeners with hearing loss. *Ear and hearing*, *39*(4), 664-678.

Profiling Ergebnisse auf Building Block- und Instruktions-Ebene

Function Profiling for Fusion G3



ISA Profiling for Fusion G3

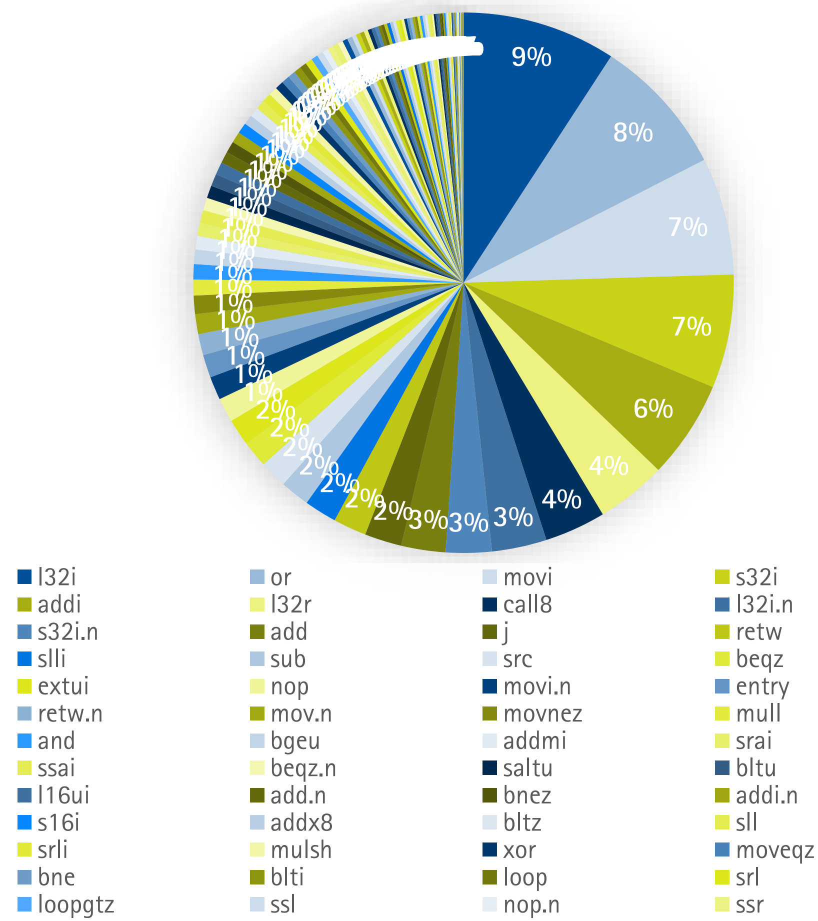
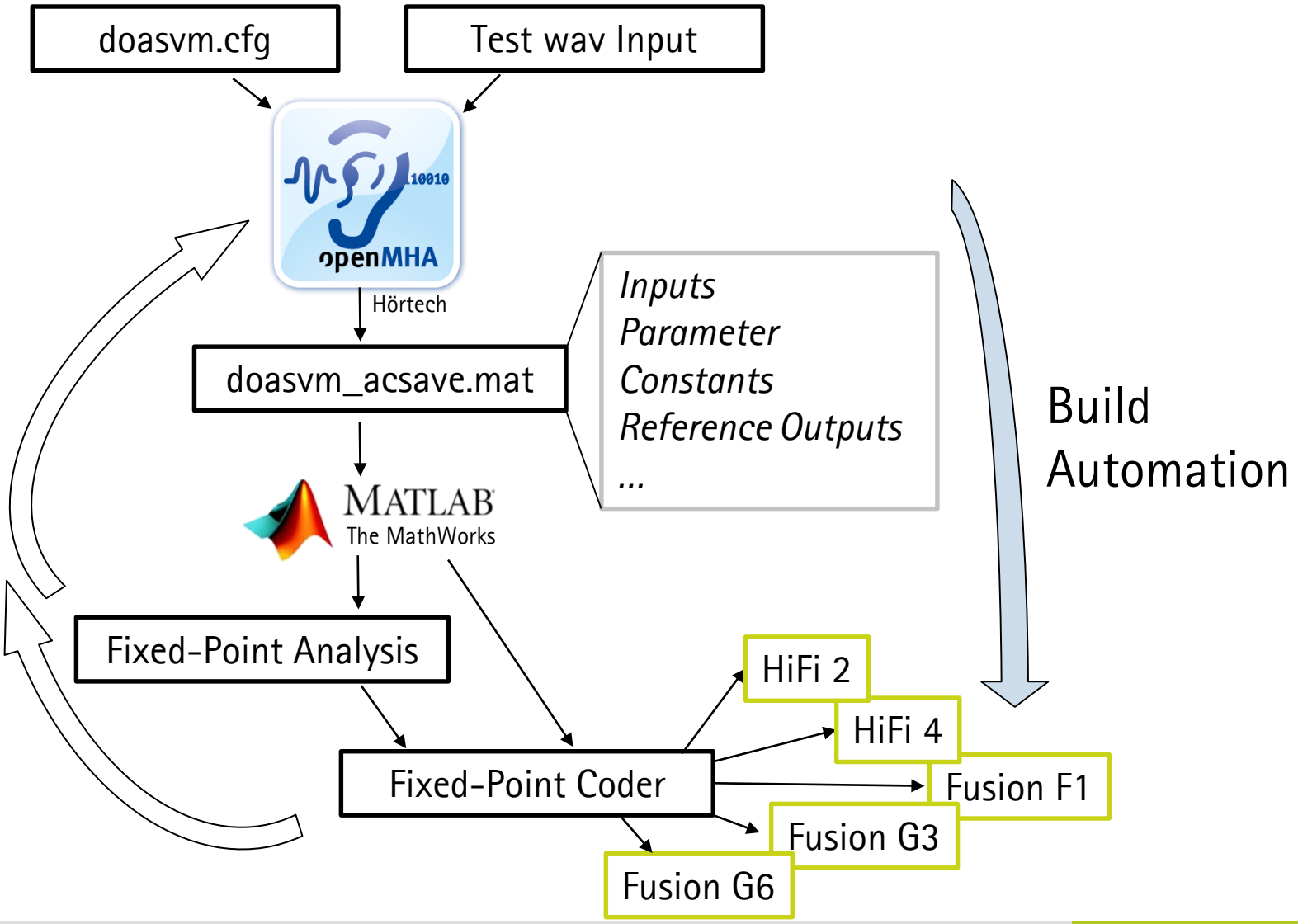


TABLE I. RADIX 2/4 COMPLEX FFT PERFORMANCE OF PROPOSED AND RELATED ARCHITECTURES

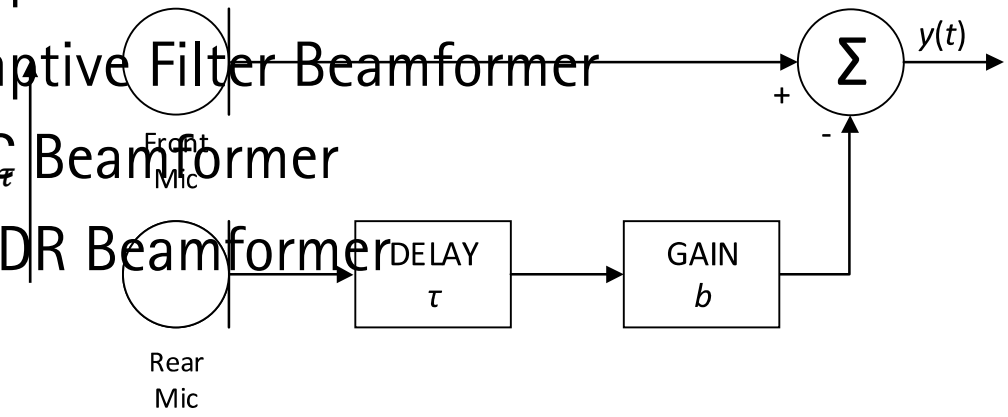
	MAC Arch.	Name	Clock frequency [MHz]	Number of MACs	Word/Subword [bit]	FFT Points [cycles] 32	64	128	256	512	1024
programmable	Single MAC	NXP CFX	300	2	24/24	997 (697%)	1657 (496%)	4129 (503%)	8393 (438%)	18654 (424%)	41385 (416%)
		TI C55x	200	2	16/16	-	-	-	4786 (250%)	-	-
		Hinrichs[3]	40	4	16/16	-	-	-	-	-	14440 (145%)
		Analog ADSP-21161N	100	2	32f	-	1156 (346%)	2158 (263%)	4316 (225%)	8770 (199%)	18288 (184%)
	SIMD MAC	Arm[7]	50	2	32/32	-	-	2700 (329%)	-	-	-
		TI C674x	456	2	32/32	258 (191%)	545 (163%)	953 (116%)	2216 (116%)	4664 (106%)	10055 (101%)
		Nadehara[8]	200	1	64/16	- -	839 (251%)	- -	4093 (214%)	- -	19257 (193%)
		Freescale SC3850	1000	4	64/16	212 (157%)	525 (157%)	1273 (155%)	2587 (135%)	5854 (133%)	11898 (119%)
	SIMD CMAC	KAVUAKA this work	50	1	64/32	135 (100%)	334 (100%)	821 (100%)	1915 (100%)	4397 (100%)	9959 (100%)
dedicated	Specialized CMAC	Al[4]	-	1	16/16	-	-	-	1024 (53%)	-	-
		Lee[5]	144	1	-/-	-	-	-	1536 (80%)	-	7680 (77%)
		Liu[6]	320	2	32/16	-	-	284 (35%)	568 (30%)	1188 (27%)	2496 (25%)

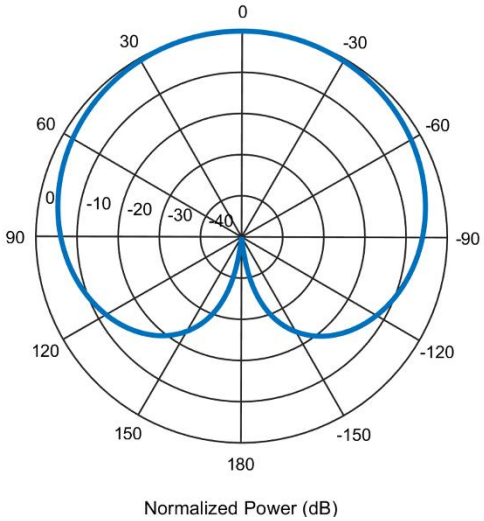
Fixed-Point Implementation Code Generation and DSE Flow



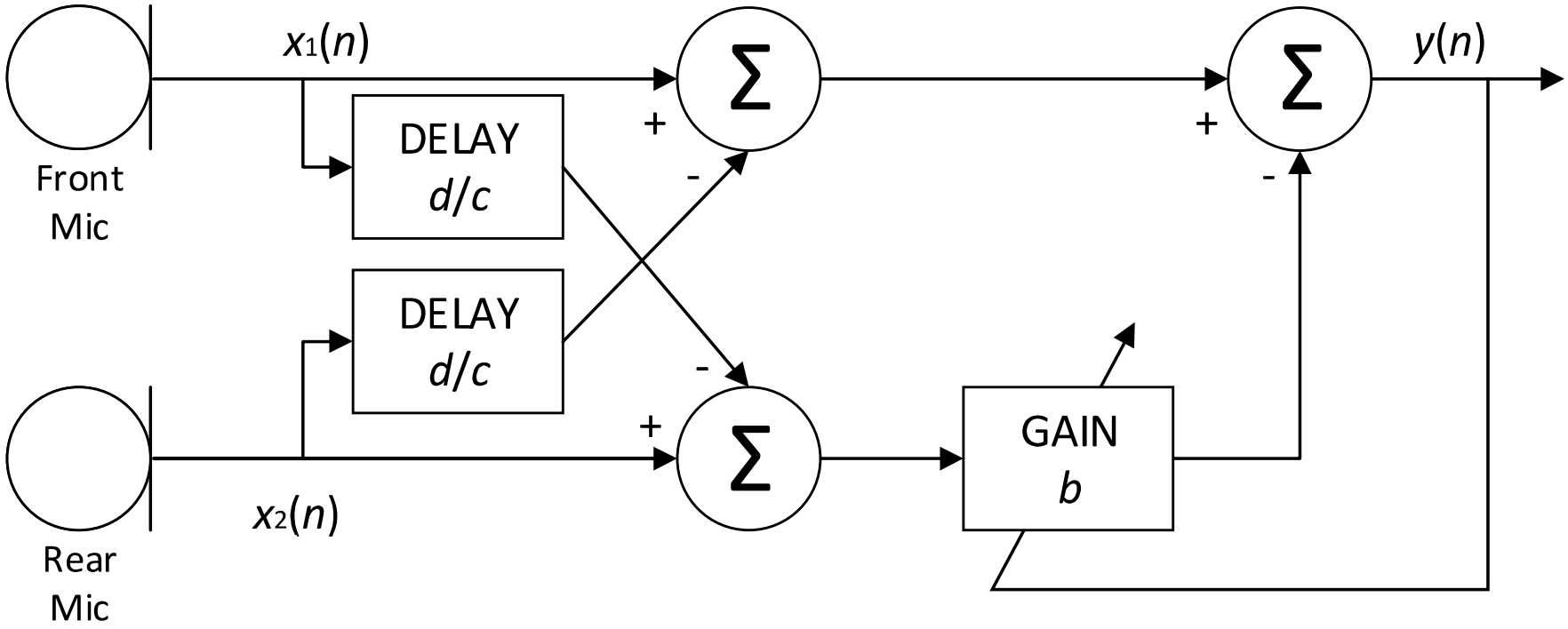
Akustische Beamformer
Referenzalgorithmen

- Fixed Beamformer
- Adaptive Gain Beamformer
- Adaptive Filter Beamformer
- GSC Beamformer
- MVDR Beamformer

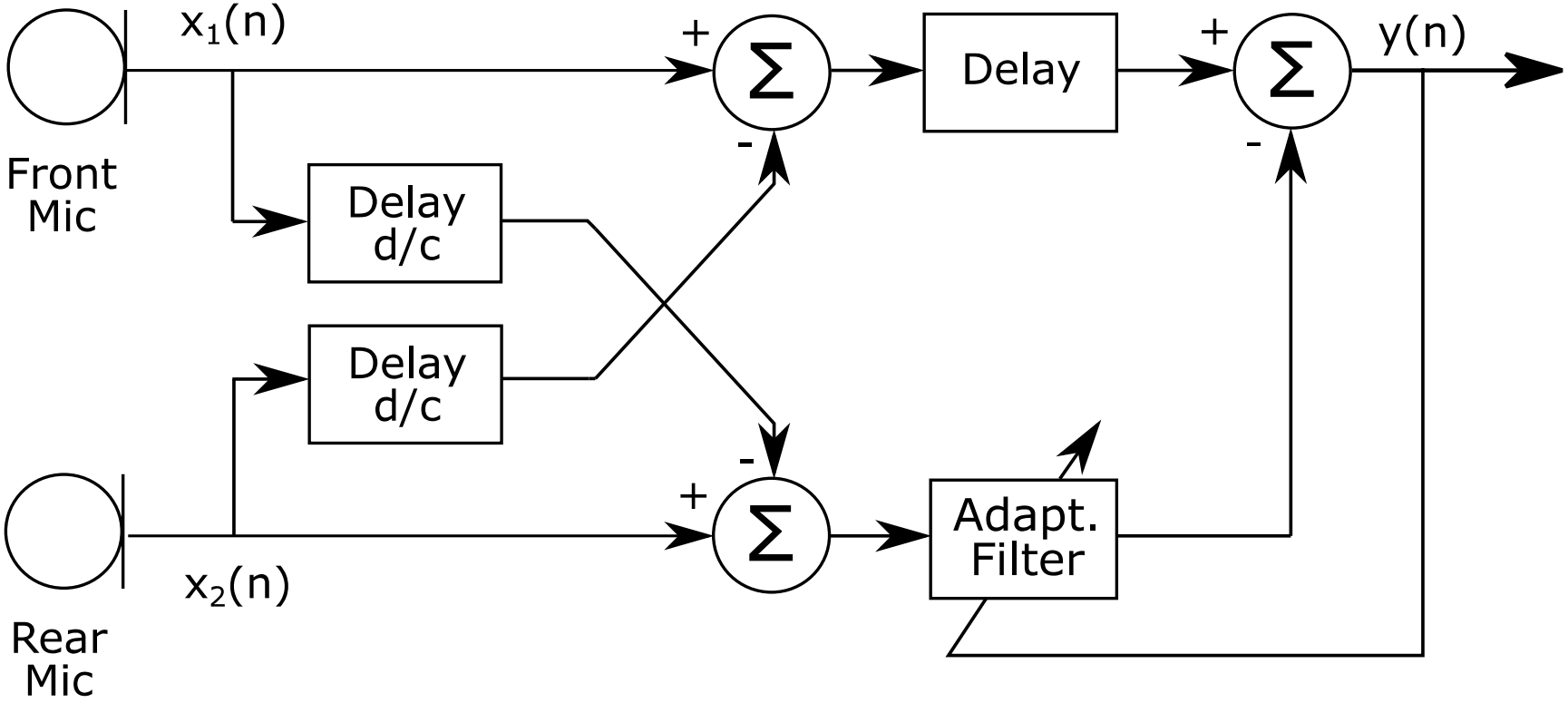




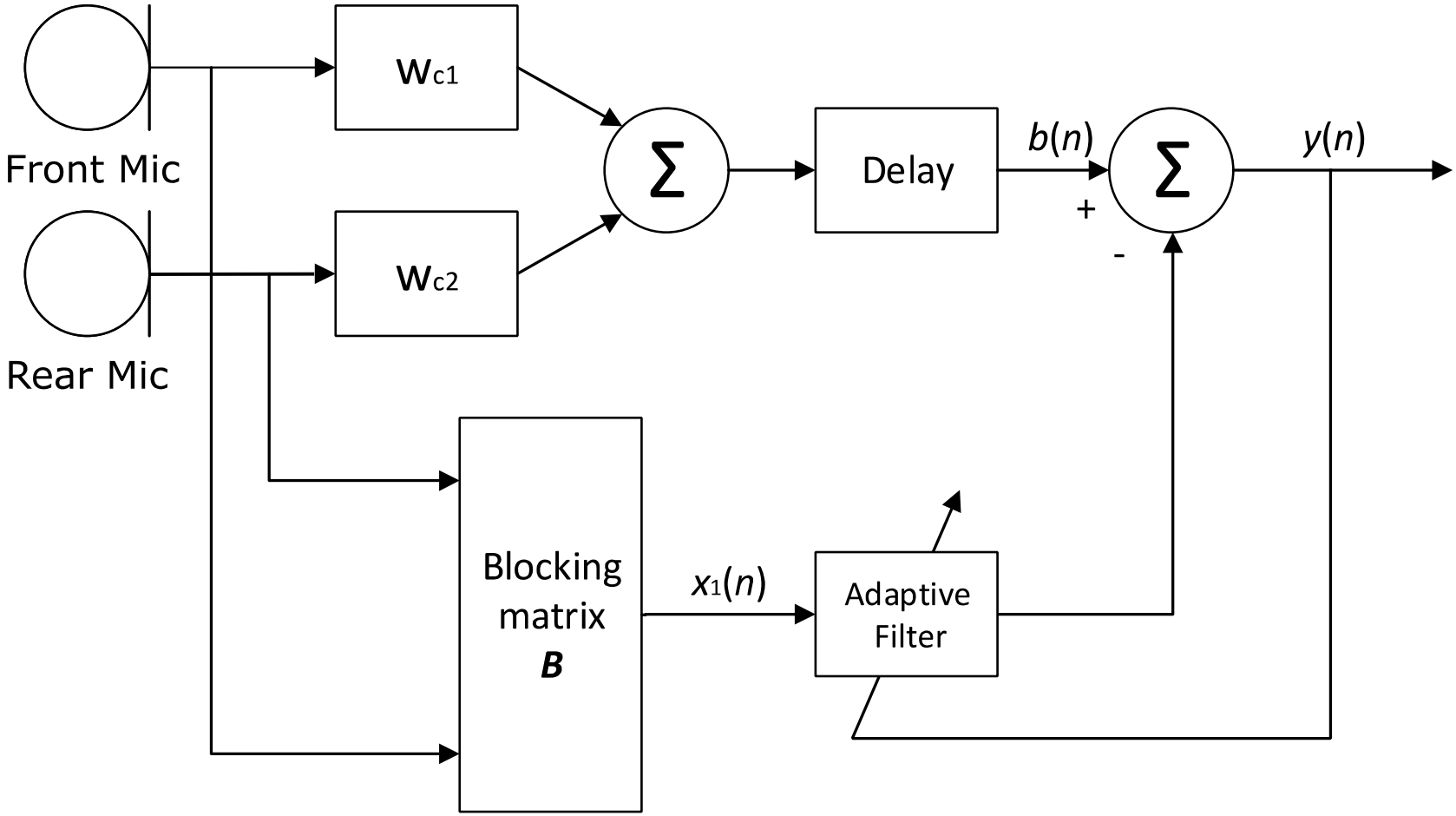
Akustische Beamformer
Adaptive Gain Beamformer



Akustische Beamformer
Adaptive Filter Beamformer



Akustische Beamformer
GSC Beamformer



Hyperbolic and trigonometric operations		Sine	Cosine	Arctangent	Divide	Exponential	Natural logarithm	Square root	Power
Cycles	KAVUAKA+CORDIC (HW)	71	71	47	63	76	56	59	129
	KAVUAKA (SW)	621	621	593	627	668	664	649	1309
	TI TMS320C6478	1259	1523	1319	637	1529	1134	341	5041

AP4 – Profiling der Referenzalgorithmen
Statusupdate

- **Analyse der Energieaufnahme der Hardwarekomponenten zur Energieeffienzabschätzung**
  - Hardware Implementation Flow ist aufgesetzt
  - Wird getestet mit TSMC 40 nm (wird unterstützt)
  - GF 20 nm PDK ist verfügbar
  - INVECAS IP Bibliotheken sind angefragt

Introduction to GLOBALFOUNDRIES 22FDSOI Technology

Erlangen, 4.10.-5.10.2018

Dr. Norbert Weber, Head of Integrated Circuits and Systems Department

Fraunhofer IIS

EUROPRACTICE IC SERVICE

AP7 – Software-Mapping der Referenzalgorithmen auf neue Architektur

- **Optimierung der Referenzalgorithmen**
  - Algorithmische Optimierungen werden nach der Auswertung der Profiling Ergebnisse durch Hörtech durchgeführt
  - Verwendung der vorhandenen Prozessor-spezifisch optimierten Bibliotheken
- **Umsetzung der Referenzalgorithmen auf die Tensilica-Architektur**
  - Erster Algorithmus ist fast vollständig implementiert
- **Zusammenstellung einer C oder C++ Bibliothek**
  - Bisher nur automatisch generiert inklusive Datentypen
- **Verifikation des Compilers-Frameworks**
  - Größtenteils nur mit Standardeinstellungen getestet

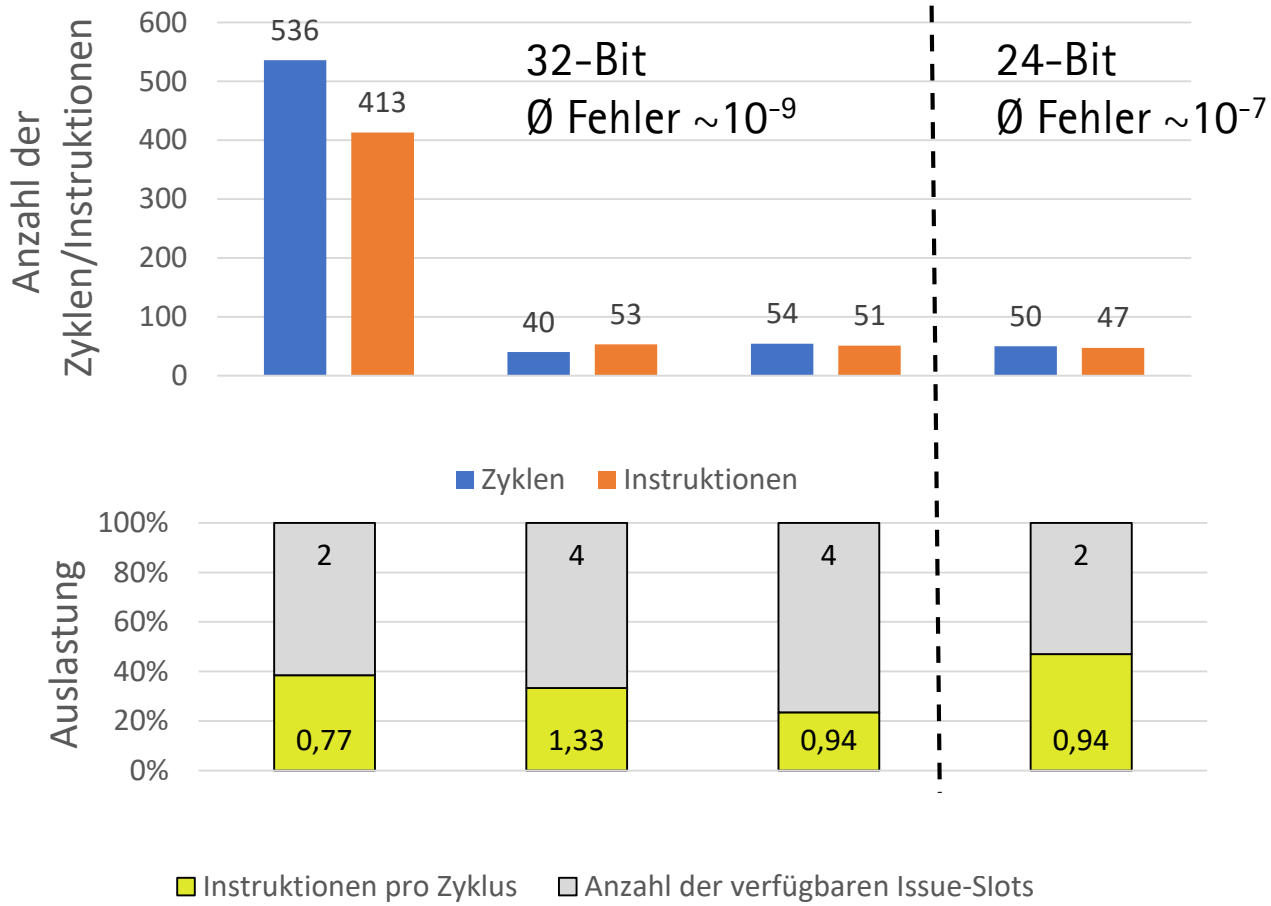
AP5 – Design und Optimierung der ISA des Prozessors
AP6 – Design von dedizierten Akzeleratoren
Statusupdate

- **Erstellung von TIEs**
- **Erstellung spezifischer Akzeleratoren mit Hilfe von VHDL (z.B. für Filterbänke, komplexe arithmetische Funktionen etc.)**

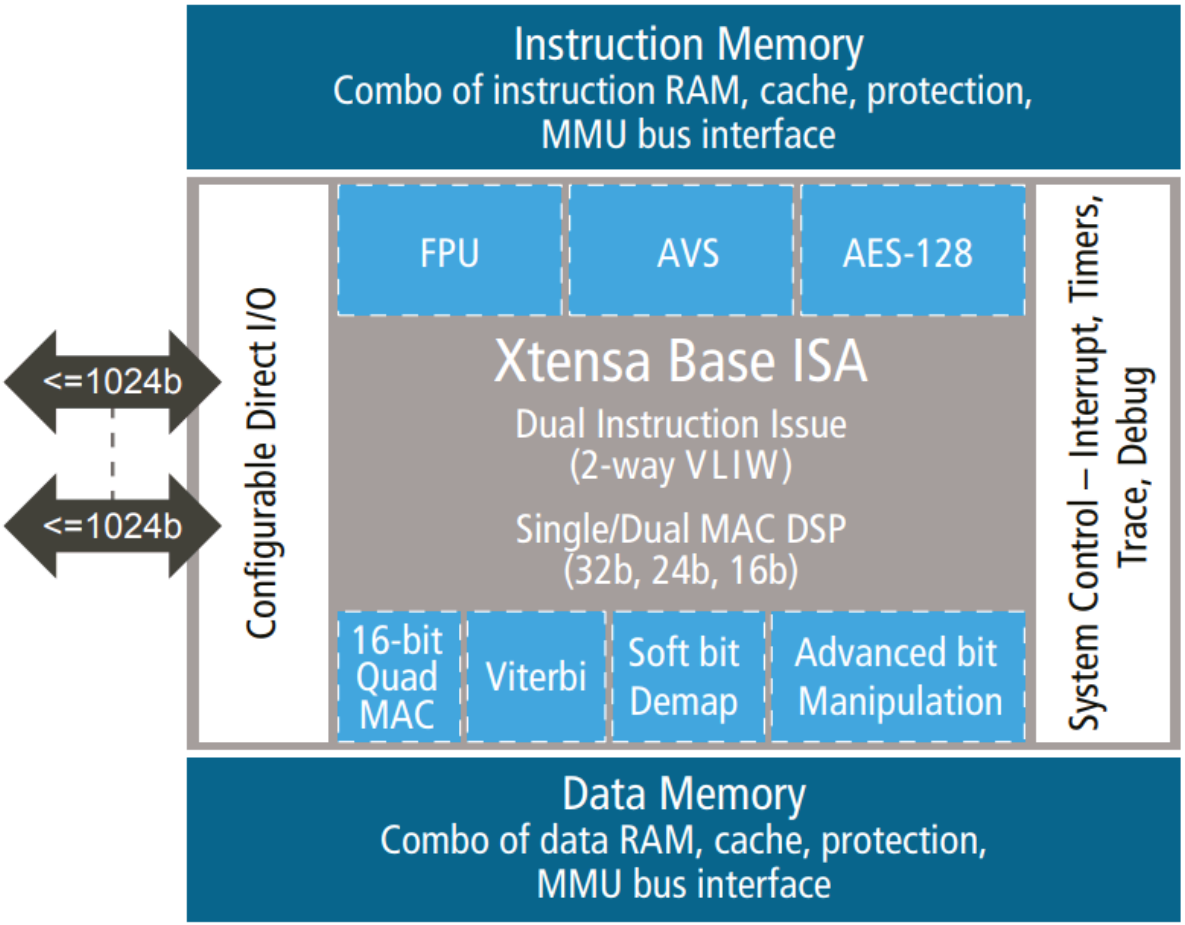
- Erste Ideen existieren
- Implementierung steht aus
- Beispiele existieren aus Projekten wie Hearing4All
  - Co-Prozessoren und Funktionelle Einheiten für Audiosignalverarbeitung

Evaluierung und Optimierung

Reduktion der Datenwortbreite und Verwendung von intrinsischen Instruktionen für den Hifi2



Fixed Beamformer



Key Tensilica Fusion F1 DSP Features

Feature	Supported	Benefits
VLIW Slots	2	Allows two concurrent operations
Fixed-Point MACs per Cycle	One 32x32 Two 24x24 Two 16x16 (four optional)	Flexible MAC architecture efficiently handles broad set of data types
Accumulator	64-bit	
FPU (optional)	Integrated scalar VLIW FPU	Tightly integrated for efficient performance
ITU/ETSI Intrinsics	Yes	Accelerates ITU voice codec performance
Circular Buffer	Yes	Efficiently manage streams of DSP data
Bit Stream VLD (incl. Huffman)	Yes	Accelerates audio codec performance
Prefetch for Cache-based Configurations	Optional	Improves cache memory performance
Customer-defined Instructions for Further Optimization	Yes	Extend the instruction set architecture (ISA) to efficiently meet customer requirements
Additional Options	AVS Advanced bit manipulation	Compatible with HiFi DSP software (option) ISA extensions to accelerate wireless communications standards (option)

https://ip.cadence.com

Tensilica HiFi DSP Family Specifications

ISA Subset Variant		HiFi24 ISA			HiFi32 ISA		
		HiFi 2	HiFi Mini	HiFi EP	HiFi 3	HiFi 3z	HiFi 4
		v24	v24	v24e	v32(vf)	v32(vf)	v32(vf)
VLIW Slots		2			3	3	4
VLIW Bundle Instruction Sizes		64-bit	40-bit	64-bit	64-bit	40-/ 48-/ 64-bit	48-/ 88-bit
Fixed Point MACs per Cycle	32x32	n/a			2	2	4
	24x24	2			4	4	4
	32x16	2			4	4	up to 8
	16x16	2			4	up to 8	up to 8
Accumulator		56-bit			64-bit	64-bit	64- / 72-bit
FPU (optional)		Scalar (Independent) FPU			Integrated 2-way SIMD VFPU	Integrated 2-way SIMD VFPU	2 integrated 2-way SIMD VFPU
ITU Intrinsic Support		No			Yes		
Circular Buffer Support		None	None	1	1	1	2
Bitstream VLE/VLD Support		Yes			Yes	Yes	Yes
User defined instructions		Yes	Limited	Yes	Yes	Yes	Yes
Other Notes				Includes 32x24 precision MAC			

ISA Variant Definitions		
	v24	Base HiFI instruction set, 24-bit and 16-bit data types
	v24e	Enhanced precision 32x24 support, circular buffer
	v32	Superset of v24 with additional operations on 32-bit types
	v32(vf)	Single-precision integrated IEEE vector floating point

https://ip.cadence.com

Key Tensilica Fusion F1 DSP Features

Feature	Supported	Benefits
VLIW Slots	2	Allows two concurrent operations
Fixed-Point MACs per Cycle	One 32x32	Flexible MAC architecture efficiently handles broad set of data types
	Two 24x24	
	Two 16x16 (four optional)	
Accumulator	64-bit	
FPU (optional)	Integrated scalar VLIW FPU	Tightly integrated for efficient performance
ITU/ETSI Intrinsics	Yes	Accelerates ITU voice codec performance
Circular Buffer	Yes	Efficiently manage streams of DSP data
Bit Stream VLD (incl. Huffman)	Yes	Accelerates audio codec performance
Prefetch for Cache-based Configurations	Optional	Improves cache memory performance
Customer-defined Instructions for Further Optimization	Yes	Extend the instruction set architecture (ISA) to efficiently meet customer requirements
Additional Options	AVS Advanced bit manipulation	Compatible with HiFi DSP software (option) ISA extensions to accelerate wireless communications standards (option)

Key Fusion G DSP Family Features

Specification	Fusion G3 DSP	Fusion G6 DSP
Load/store	Dual 128-bit load/store and 128-bit load units	Dual 256-bit load/store and 256-bit load units
Instruction lengths (bits)	128, 64, 24, 16	
VLIW	128 bits wide, 2- or 4-issue slots	
SIMD	2/4/8/16-way for integer, 2/4-way for floating point	4/8/16/32-way for integer, 4/8-way for floating point
Data formats	8/16/20/32/40/64/80 real integer, 32/64 real and complex floating-point	
Vector register files	32 entries x 128-bit, 4 entries x 320-bit (wide data)	32 entries x 256-bit, 4 entries x 640-bit (wide data)
ALU operations/cycle	2, 4, 8, 16 for 64, 32, 16, 8-bit	4, 8, 16, 32 for 64, 32, 16, 8-bit
Multiply bit width	8 x 8, 16 x 16, 32 x 32, 64 x 64	
MACs/cycle	1, 4, 8, 16 for 64, 32, 16, 8-bit	2, 8, 16, 32 for 64, 32, 16, 8-bit
16-bit complex MAC throughput	2 per cycle	4 per cycle
Guard bits	4/8/16-bit on 16/32/64 types (utilizing 20/40/80-bit accumulators)	
Aligning load	Available in either of two slots	
Addressing modes	Post-increment, reverse, circular, and general	
Floating-point configuration options	IEEE-754 compliant, single-precision vector floating-point or single- and double-precision vector floating-point	
Floating-point operations	Vector MAC, FMA, ADDSUB, ALU, and type conversion, including to/from half-precision	
MAC/FMA/ADDSUB per cycle	2, 4 for double, single-precision floating-point	4, 8 for double, single-precision floating-point
Division/reciprocal/RSQRT	Included in ISA: integer division and floating-point division/reciprocal/RSQRT	
Vector-programming support	M-way, 32-bit based	
FFT acceleration	ADDSUB included for 2X FFT speedup (fixed- and floating-point)	
Image processing	Scatter gather and histogram	
Vector predication	ALU, MAC, LOAD/STORE, etc.	
DMA support	Integrated DMA controller option and support for external DMA controller	