HALO: A Flexible and Low Power Processing Fabric for Brain-Computer Interfaces

Abhishek Bhattacharjee, Computer Science

Rajit Manohar, Electrical Engineering

Yale University

Ioannis Karageorgos

Karthik Sriram

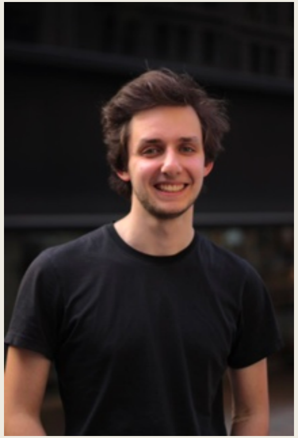
Ján Veselý

Michael Wu

Xiayuan Wen

Nick Lindsay

Lenny Khazan



ID-19 Science

NEWS | HEALTH

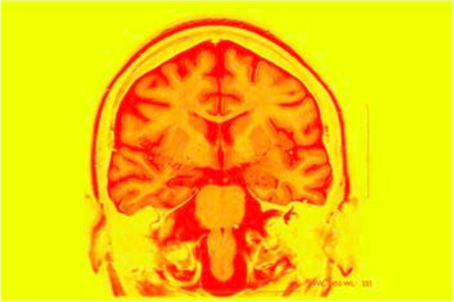
In a first, brain implant lets man with complete paralysis spell out thoughts: 'I love my cool son.'

Surgically placed electrodes enable person with late-stage ALS to communicate via neural signals

22 MAR 2022 • 12:00 PM • BY KELLY SERVICK

A Brain Implant Improved Memory, Scientists Report

Give this article 12



A magnetic resonance image of an epileptic brain. Scientists have tested a brain implant on people with epilepsy that aided memory. Bsip/UIG, via Getty Images

Subscribe Latest Issues SCIENTIFIC AMERICAN. Cart 0 Sign In | Newsletters

MENTAL HEALTH

Experimental Brain Implant Could Personalize Depression Therapy

Symptoms subsided for one woman after a carefully targeted neural circuit was stimulated

By Gary Stix on October 4, 2021

DARPA's BCI Chip Allows Pilots to Control Drones Telepathically

In a breakthrough discovery, DARPA researchers have developed a BCI chip that can control multiple drones with the use of brainwaves. | Image By Antiv Shutterstock

The New York Times

Brain Implant Allows Fully Paralyzed Patient to Communicate

Letter by painstaking letter, a man in a completely locked-in state was able to formulate words and sentences using only his thoughts.

The New York Times

A 'Pacemaker for the Brain': No Treatment Helped Her Depression — Until This

It's the first study of individualized brain stimulation to treat severe depression. Sarah's case raises the possibility the method may help people who don't respond to other therapies.

HEALTHCARE

Brain Implants With The Potential To Restore Vision To The Blind

William A. Haseltine Contributor

Follow

Nov 5, 2021, 12:24pm EDT

Subscribe Latest Issues SCIENTIFIC AMERICAN. Cart 0 Sign In | Newsletters

NEUROSCIENCE

New Brain Implant Transmits Full Words from Neural Signals

No spelling out of letters is needed for a paralyzed person to use the first-of-a-kind neuroprosthesis

By Emily Willingham on July 15, 2021

The New York Times Magazine

THE HEALTH ISSUE

The Man Who Controls Computers With His Mind

16 years ago, Dennis DeGray was paralyzed in an accident. Now, implants in his brain allow him some semblance of control.

kernel

synchron

Blackrock Neurotech

NEIRY

NEURALINK

NEUROPACE

InteraXon

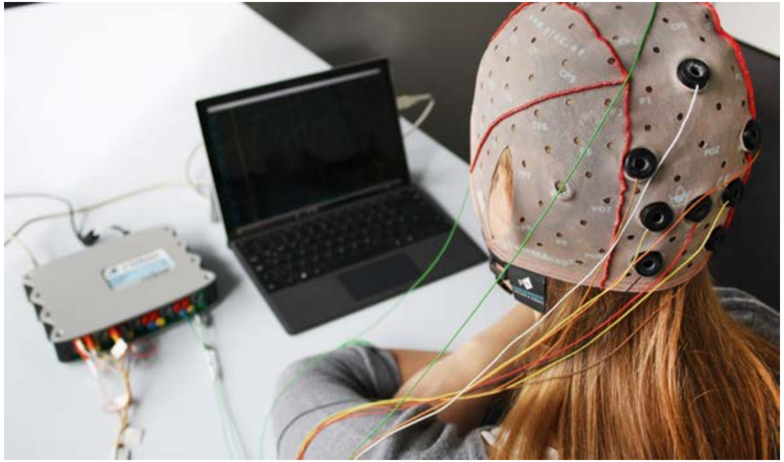
NEUROLUTIONS
Restoring function by thought

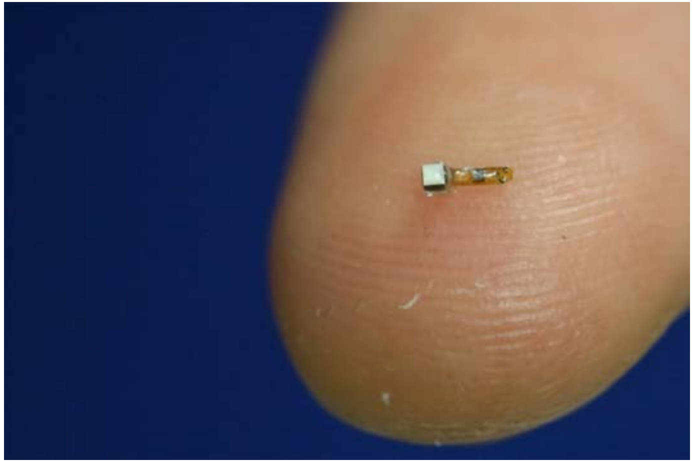
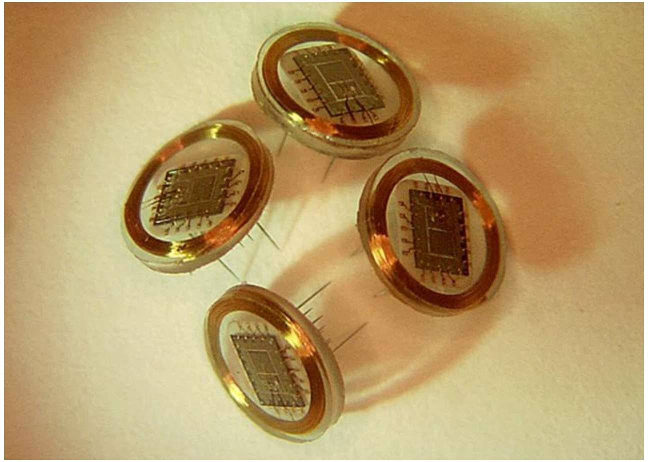
Medtronic

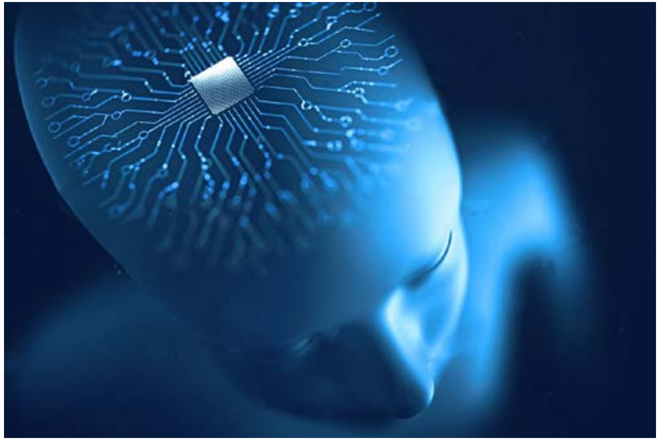
Meta

mindmaze

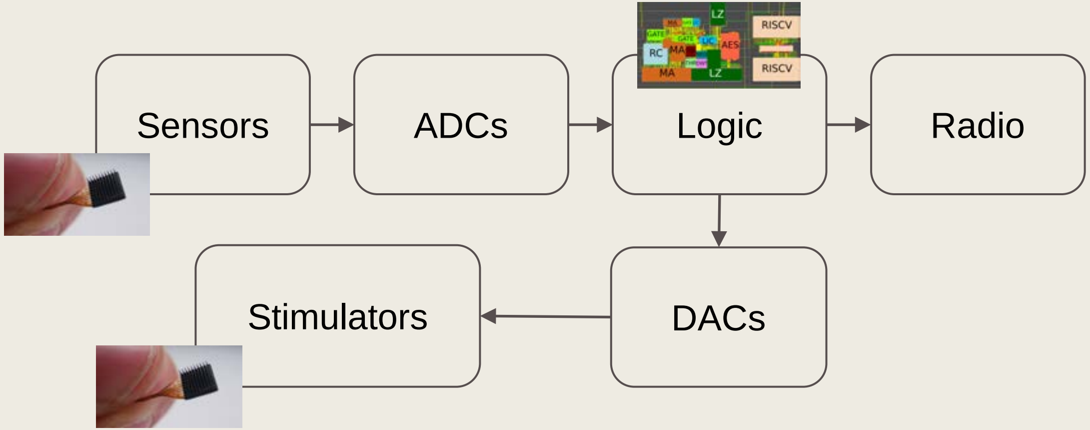
Cognixion





How are implantable brain-computer interfaces implemented?



Implantable brain-computer interfaces trade processing, power, real-time processing, and flexibility

The FDA warns against overheating cellular tissue beyond 1℃ → 15-40mW

DARPA NESD targets 100s Mbps - 10s Gbps to read/stimulate biological neurons

Responses within 10s of milliseconds to treat epilepsy or movement disorders

Flexibility for new computational methods, use cases, for personalization, and to build standards for wider computational stack

TASKS	Medtronic	Neuropace	Aziz et al.	Chen et al.	Kassiri et al.	NURIP
Spike Detection						
Compression			✓			
Seizure Prediction		✓		✓	✓	✓
Movement Intent	✓					
Encryption						

FEATURES						
Programmable	✓	Limited		Limited	✓	Limited
Read Bandwidth	10Kbps	20Kbps	10Mbps	8Kbps		4Mbps
Stimulation Bandwidth	10Kbps	20Kbps				
Safety (<15mW)	✓	✓	✓		✓	✓

TASKS	Medtronic	Neuropace	Aziz et al.	Chen et al.	Kassiri et al.	NURIP
Spike Detection						
Compression			✓			
Seizure Prediction		✓		✓	✓	✓
Movement Intent	✓					
Encryption						

FEATURES						
Programmable	✓	Limited		Limited	✓	Limited
Read Bandwidth	10Kbps	20Kbps	10Mbps	8Kbps		4Mbps
Stimulation Bandwidth	10Kbps	20Kbps				
Safety (<15mW)	✓	✓	✓		✓	✓

TASKS	Medtronic	Neuropace	Aziz et al.	Chen et al.	Kassiri et al.	NURIP
Spike Detection						
Compression			✓			
Seizure Prediction		✓		✓	✓	✓
Movement Intent	✓					
Encryption						

FEATURES						
Programmable	✓	Limited		Limited	✓	Limited
Read Bandwidth	10Kbps	20Kbps	10Mbps	8Kbps		4Mbps
Stimulation Bandwidth	10Kbps	20Kbps				
Safety (<15mW)	✓	✓	✓		✓	✓

TASKS	Medtronic	Neuropace	Aziz et al.	Chen et al.	Kassiri et al.	NURIP
Spike Detection						
Compression			✓			
Seizure Prediction		✓		✓	✓	✓
Movement Intent	✓					
Encryption						

FEATURES						
Programmable	✓	Limited		Limited	✓	Limited
Read Bandwidth	10Kbps	20Kbps	10Mbps	8Kbps		4Mbps
Stimulation Bandwidth	10Kbps	20Kbps				
Safety (<15mW)	✓	✓	✓		✓	✓

TASKS	Medtronic	Neuropace	Aziz et al.	Chen et al.	Kassiri et al.	NURIP	HALO
Spike Detection							✓
Compression			✓				✓
Seizure Prediction		✓		✓	✓	✓	✓
Movement Intent	✓						✓
Encryption							✓
FEATURES							
Programmable	✓	Limited		Limited	✓	Limited	✓
Read Bandwidth	10Kbps	20Kbps	10Mbps	8Kbps		4Mbps	46Mpbs
Stimulation Bandwidth	10Kbps	20Kbps					8Mpbs
Safety (<15mW)	✓	✓	✓		✓	✓	✓

Identifying computational capabilities

Important computational methods for both clinical and research

Support for reading and stimulation of biological neurons

Supported computational kernels representative of methods used across brain regions and depths

Some computational kernels need to meet real-time processing needs

Support for parameter tuning to personalize algorithms to subject

Support for emerging algorithms and computational methods

Identifying a standard set of computational capabilities

Miscellaneous Algorithms

2-stage, in-order 32-bit modified ibex (RV32E)

RISC-V µcontroller

Widely-Used Algorithms Amenable to Specialization

Compression

Movement Intent

Seizure Treatment

Spike Detection

Encryption

Building monolithic ASICs

Miscellaneous Algorithms

2-stage, in-order 32-bit modified ibex (RV32E)

RISC-V µcontroller

Widely-Used Algorithms Amenable to Specialization

Compression	LZ4	LZMA	DWT MA
Movement Intent	Movement Intent		
Seizure Treatment	Seizure Treatment		
Spike Detection	DWT	NEO	
Encryption	AES		

Baseline: Monolithic ASIC

HALO: Processing Elements

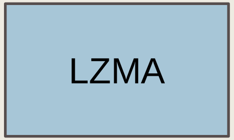
28nm FD-SOI CMOS process, physical synthesis flow with standard cells from STMicroelectronics

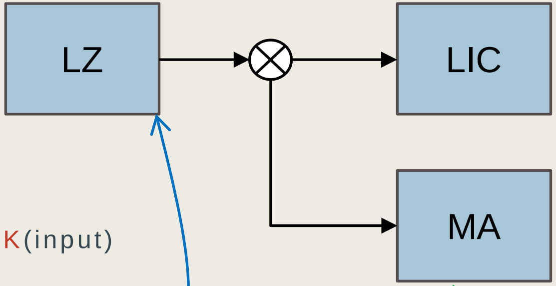
Baseline: Monolithic ASIC | **HALO: Processing Elements**

Baseline: Monolithic ASIC

HALO: Processing Elements

Compression

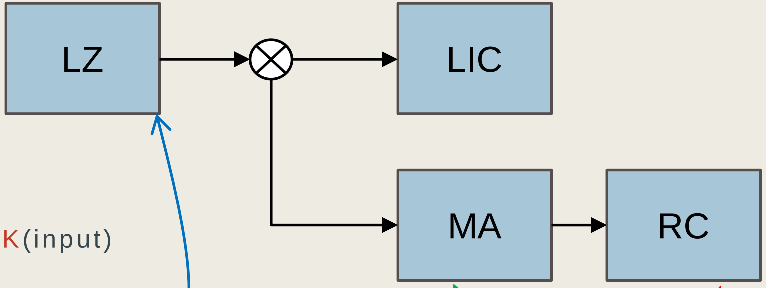





```
function LZMA_COMPRESS_BLOCK(input)
    output = list(lzma header)
    while data = input.get() do
        best_match = find_best_match(data)
        match_prob = count(match_table, best_match)
            / count_total(match_table)
        r1 = range_encode(match_prob)
        output.push(r1)
        increment_counter(match_table, best_match)
    end while
    ret output
```

Baseline: Monolithic ASIC

HALO: Processing Elements

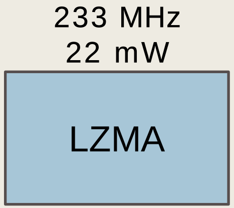



```
function LZMA_COMPRESS_BLOCK(input)
    output = list(lzma header)
    while data = input.get() do
        best_match = find_best_match(data)
        match_prob = count(match_table, best_match)
            / count_total(match_table)
        r1 = range_encode(match_prob)
        output.push(r1)
        increment_counter(match_table, best_match)
    end while
    ret output
```

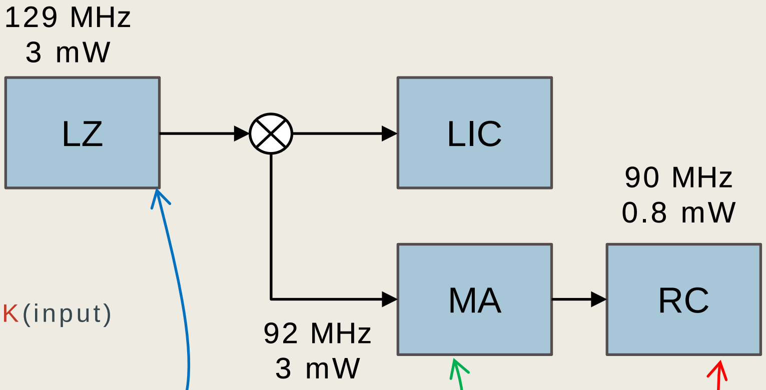
28nm FD-SOI CMOS process, physical synthesis flow with standard cells from STMicroelectronics

Compression

Baseline: Monolithic ASIC



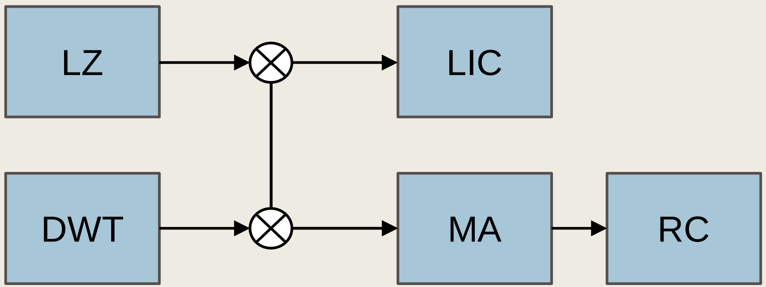
HALO: Processing Elements




```
function LZMA_COMPRESS_BLOCK(input)
    output = list(lzma header)
    while data = input.get() do
        best_match = find_best_match(data)
        match_prob = count(match_table, best_match)
            / count_total(match_table)
        r1 = range_encode(match_prob)
        output.push(r1)
        increment_counter(match_table, best_match)
    end while
    ret output
```

Baseline: Monolithic ASIC

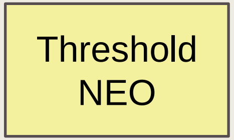
HALO: Processing Elements



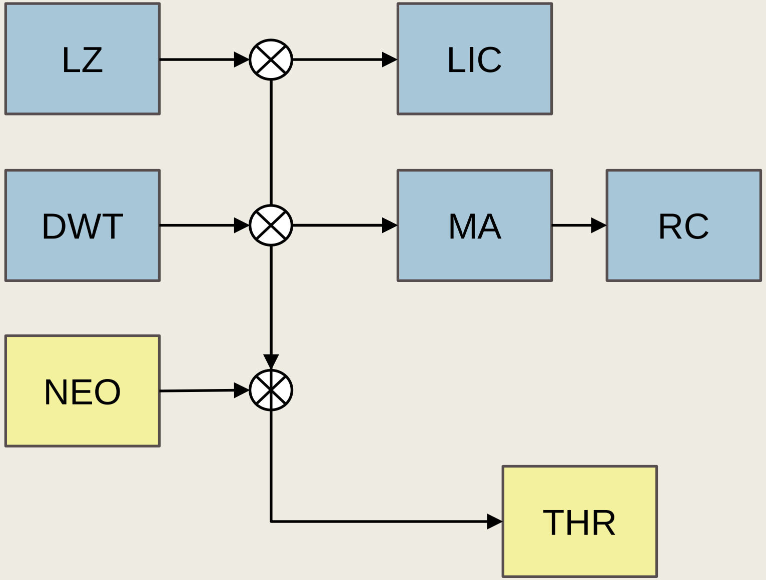
Compression

Spike Detection

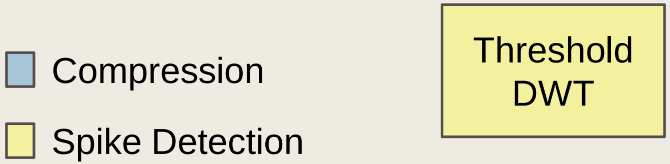
Baseline: Monolithic ASIC



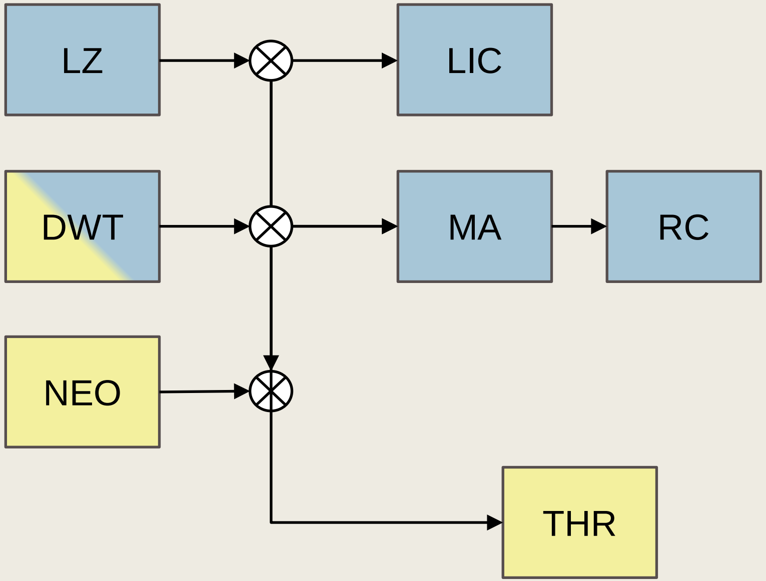
HALO: Processing Elements



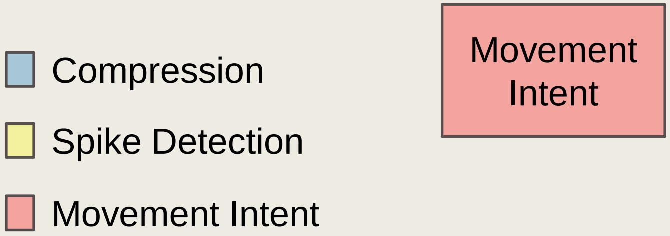
Baseline: Monolithic ASIC



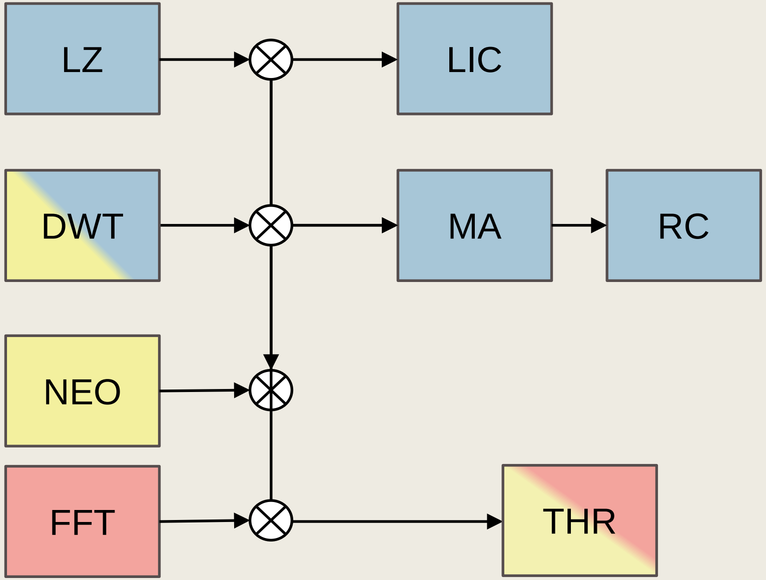
HALO: Processing Elements



Baseline: Monolithic ASIC

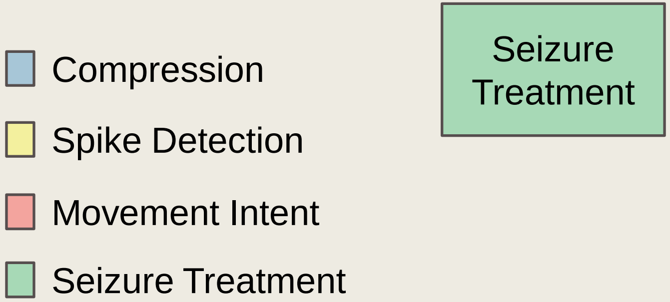


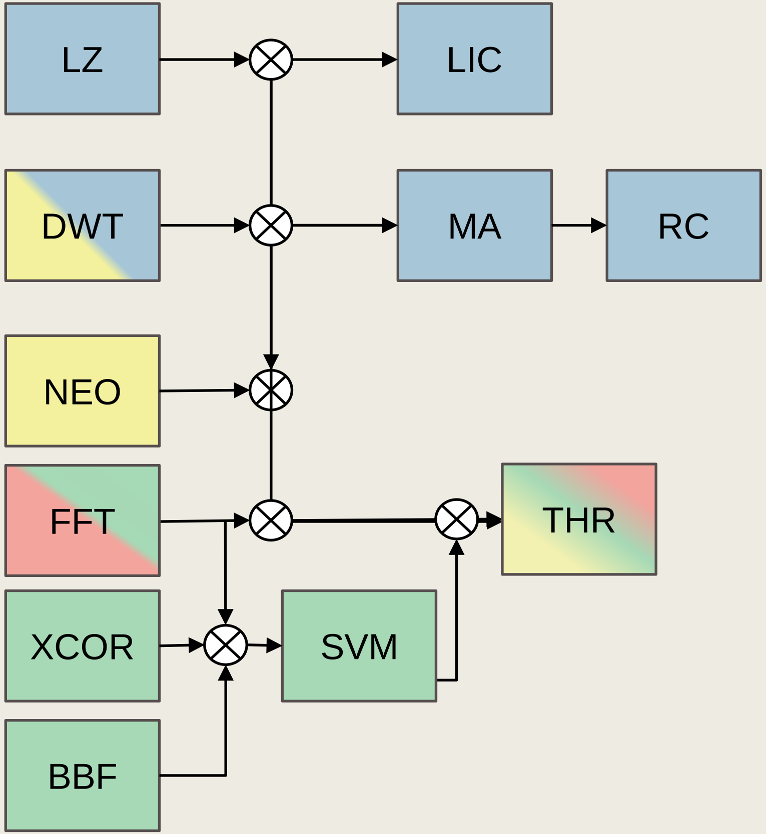
HALO: Processing Elements



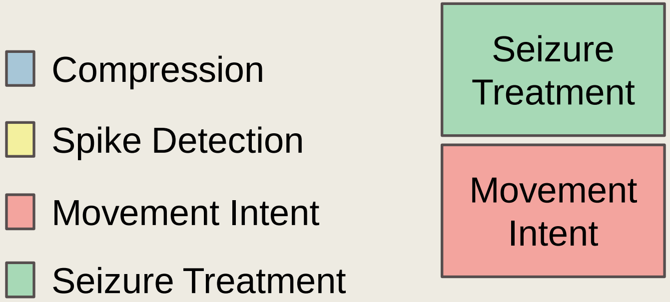
Baseline: Monolithic ASIC

HALO: Processing Elements



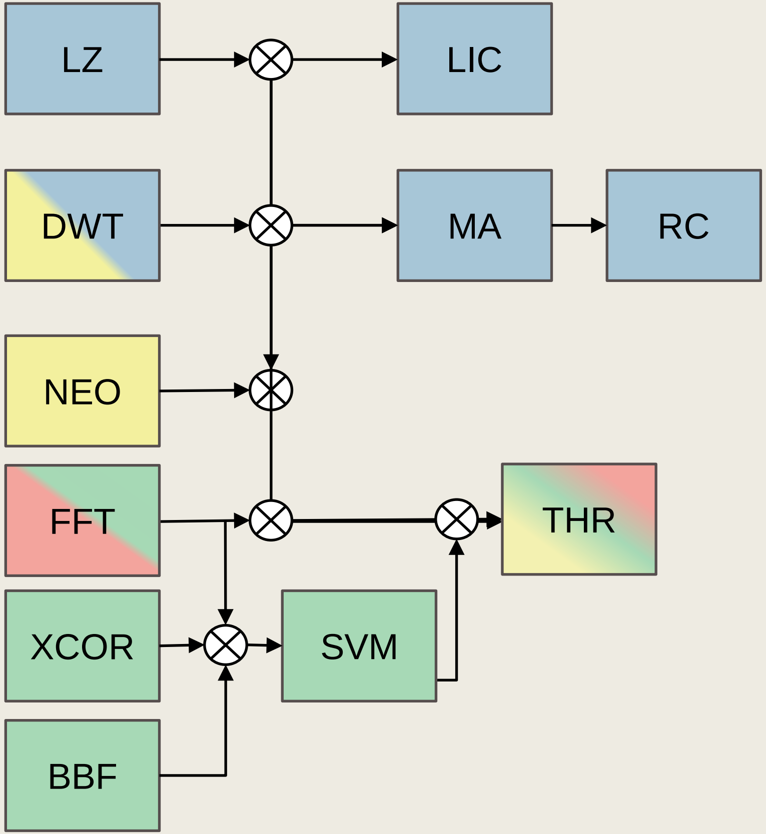


Baseline: Monolithic ASIC

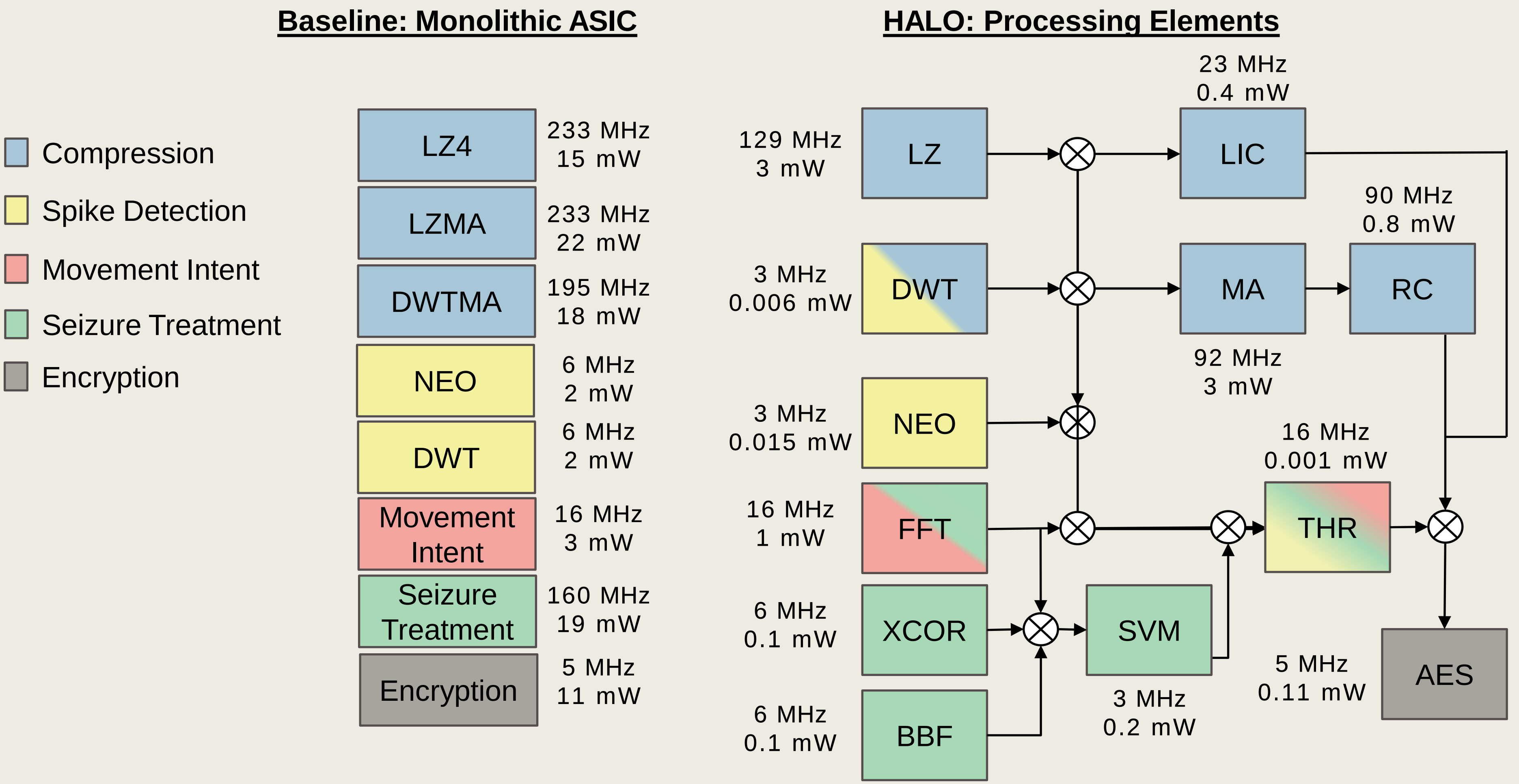


HALO: Processing Elements

Movement intent 1024 points
Seizure treatment 25 points

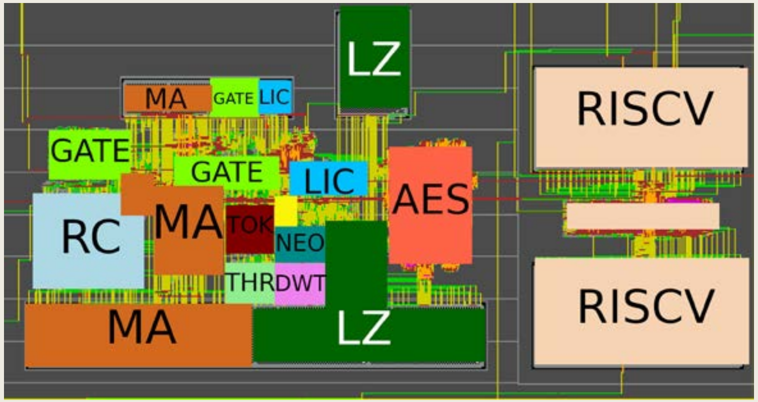


28nm FD-SOI CMOS process, physical synthesis flow with standard cells from STMicroelectronics

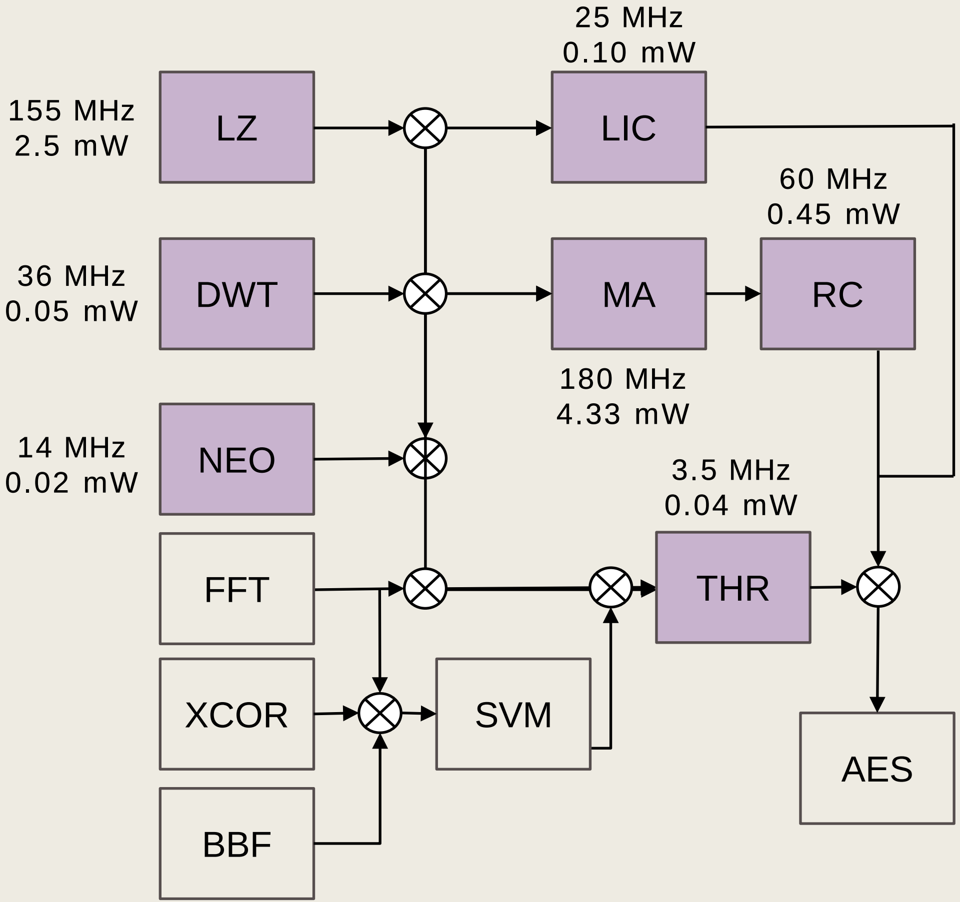


Waiting for vendor to package chip for measurement results; physical synthesis results shown

Chip tape-out in 12nm CMOS process



HALO: Processing Elements



Summary of the HALO approach

Break each computational task into individual kernels

Instead of monolithic ASIC, build a hardware PE per kernel

Clock each PE at no more than its necessary frequency

Avoid overly fine-grained PEs to reduce communication

Avoid overly coarse-grained PEs to facilitate sharing, reuse, and lower clock speed

Designing a module

Computation needs are still being investigated by neuroscience researchers

→ For rapid prototyping, we used a high-level synthesis (HLS) flow

HLS structure

- Standardized parameter settings “config” interface for µcontroller
- Elastic I/O interface from HLS tools

HLS optimizations

- Fixed-point v/s floating-point
- Choice of loop pipelining
- Re-structuring input to make it more “HLS-friendly"

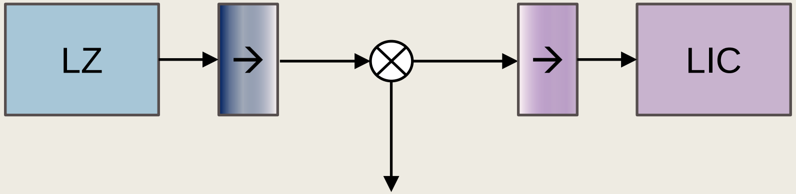
Interconnect design

Current implementation

- PE frequency to/from interconnect frequency adaptor
- Interconnect frequency selected to support “full throughput”

Standard synchronizer structure for interface to interconnect

Similar configuration interface to set configuration bits for switches



Handling bursty data

Flow of data tokens is bursty *and* data-dependent

- Example: compression produces a variable number of output data tokens
- Each component has a *peak* token consumption rate (set by its frequency)
  - → FIFOs needed at some interfaces to buffer data tokens

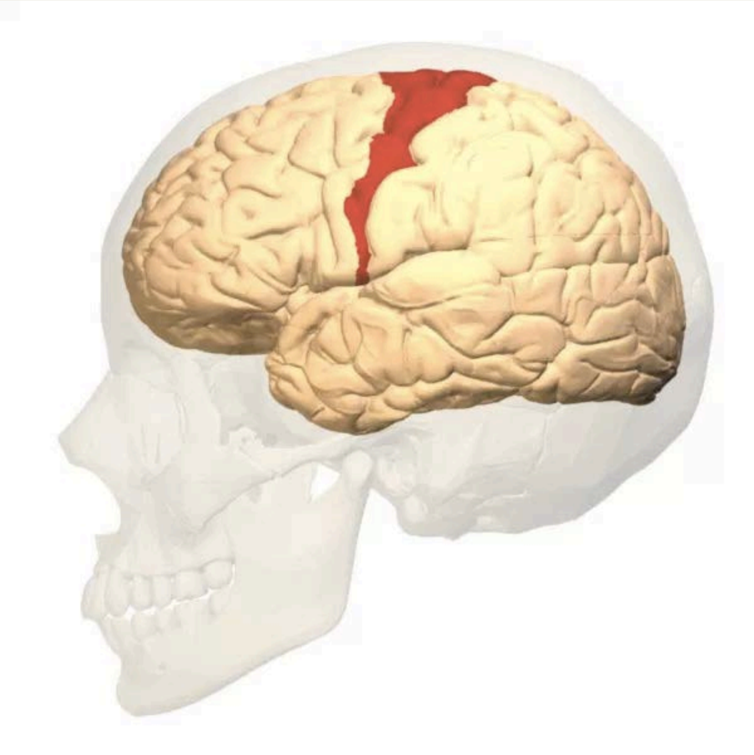
FIFOs sized based on frequency of PEs + worst-case data patterns

Management and configuration interface

Each element of architecture exports a standardized “config” port

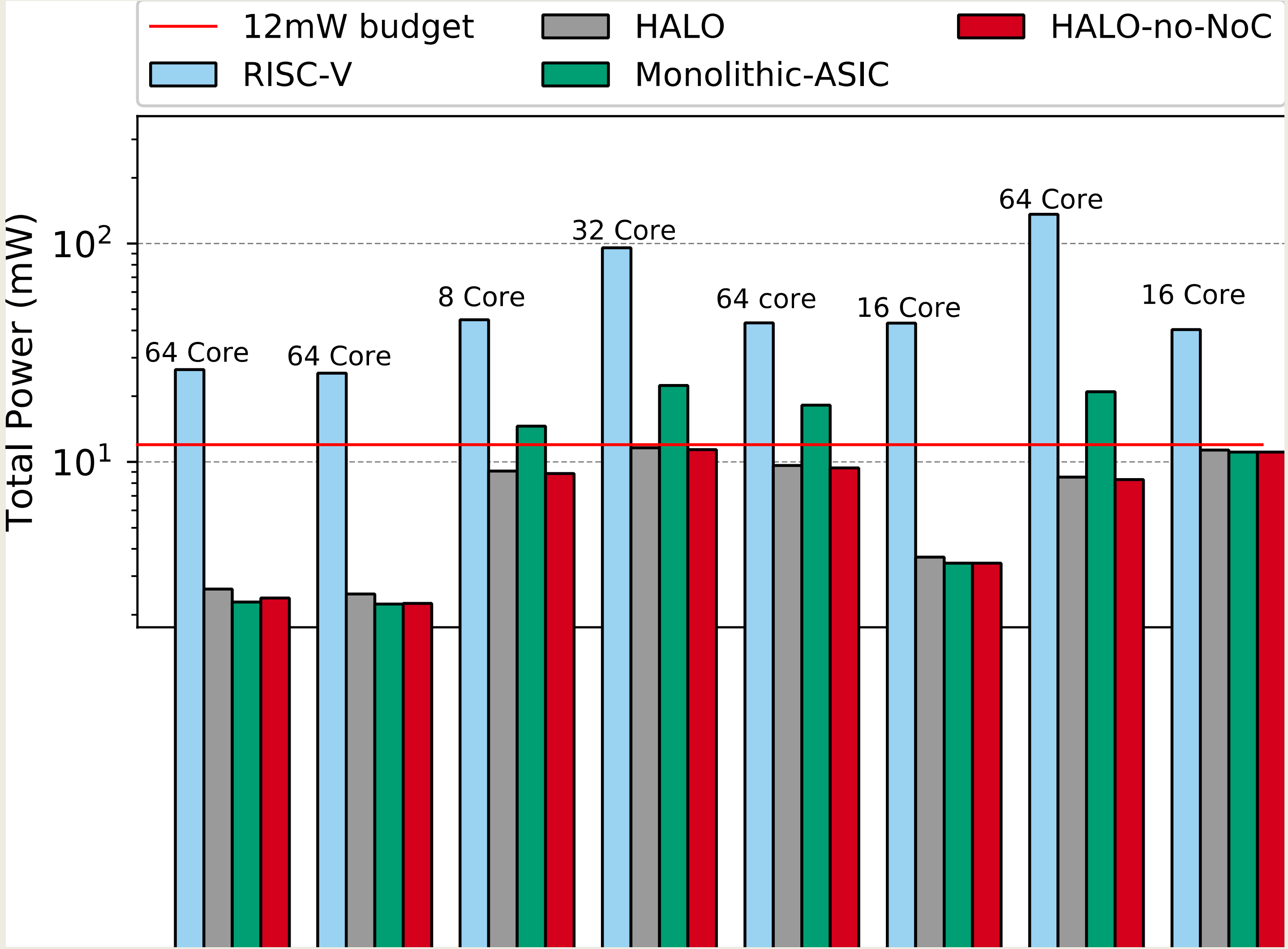
- Parameter settings
- Pipeline configuration
- Reading debugging information from PE

Config module added to RISC-V core, under software control



Evaluations using neuronal recordings of a non-human primate’s motor cortex collected by the Borton Lab lab at Brown

More recent evaluations using recordings from human patients with epilepsy collected by the Yale Epilepsy Research Center



28nm FD-SOI CMOS estimates

Results for worst case variation corner at VDD_{MAX}, Tr_{FF}, RC_{BEST} at V_{DD} of 1V

Standard cell and macro libraries characterized or interpolated to 40℃

Total power budget of 15mW, with 2mW devoted to ADCs, amplifiers, and radio


Home > Conferences > ISCA > Proceedings > ISCA '20 > Hardware-software co-design for brain-computer interfaces

RESEARCH-ARTICLE

Hardware-software co-design for brain-computer interfaces

Authors: Ioannis Karageorgos, Karthik Sriram, Ján Veselý, Michael Wu, Marc Powell, David Borton, Rajit Manohar, Abhishek Bhattacharjee Authors Info & Claims

ISCA '20: Proceedings of the ACM/IEEE 47th Annual International Symposium on Computer Architecture • May 2020 • Pages 391–404 • https://doi.org/10.1109/ISCA45697.2020.00041

Online: 23 September 2020 Publication History

Also selected for inclusion in IEEE Micro's Top Picks in Computer Architecture, article titled: "Balancing Specialized Versus Flexible Computation in Brain-Computer Interfaces"

Our focus is on more complete tape-outs, designing an asynchronous vector processor, building support for long-term storage, and distributed BCI scenarios

Also exploring potential in-vivo tests with swine with collaborators at Yale's Epilepsy Research Center

Karthik Sriram

Xiayuan Wen

Zach Taylor

Oliver Ye

Michal Gerasimiuk

Raghavendra Pothukuchi

Anurag Khandelwal

Hitten Zaveri

Dennis Spencer