Learning-based compression

Nature

Value / Knowledge

Volkan Cevher
Laboratory for Information and Inference Systems
http://lions.epfl.ch

lions@epfl

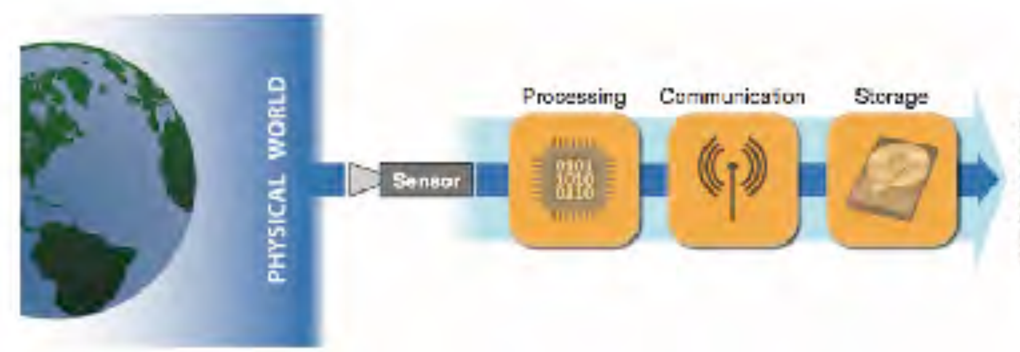
FN SNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

erc

EPFL
ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

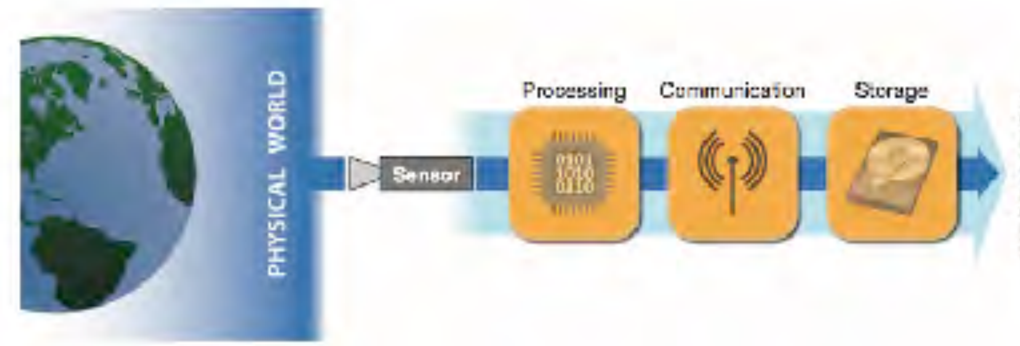
A paradigm shift in data generation

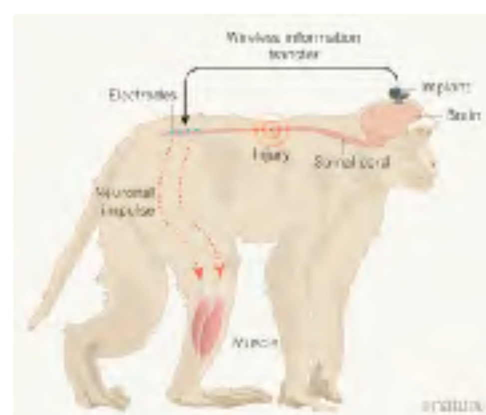
1977 - 5hours

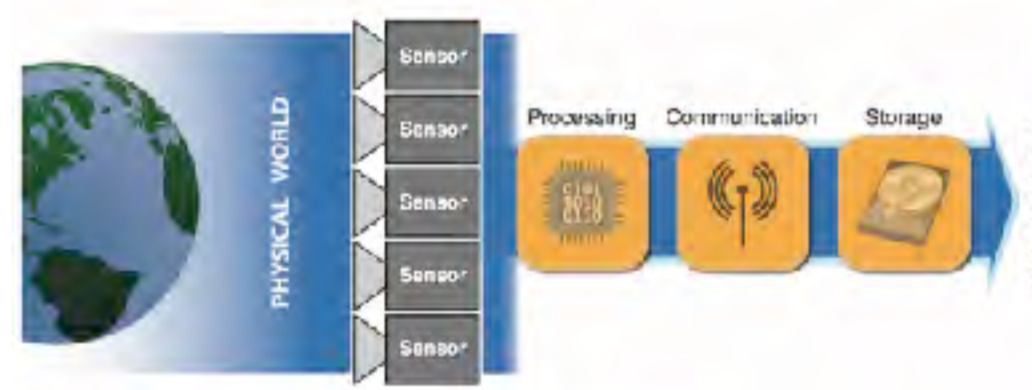


A paradigm shift in data generation

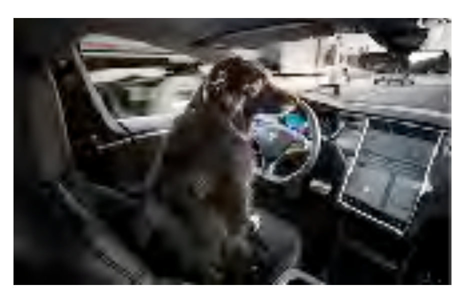
1977 - 5hours





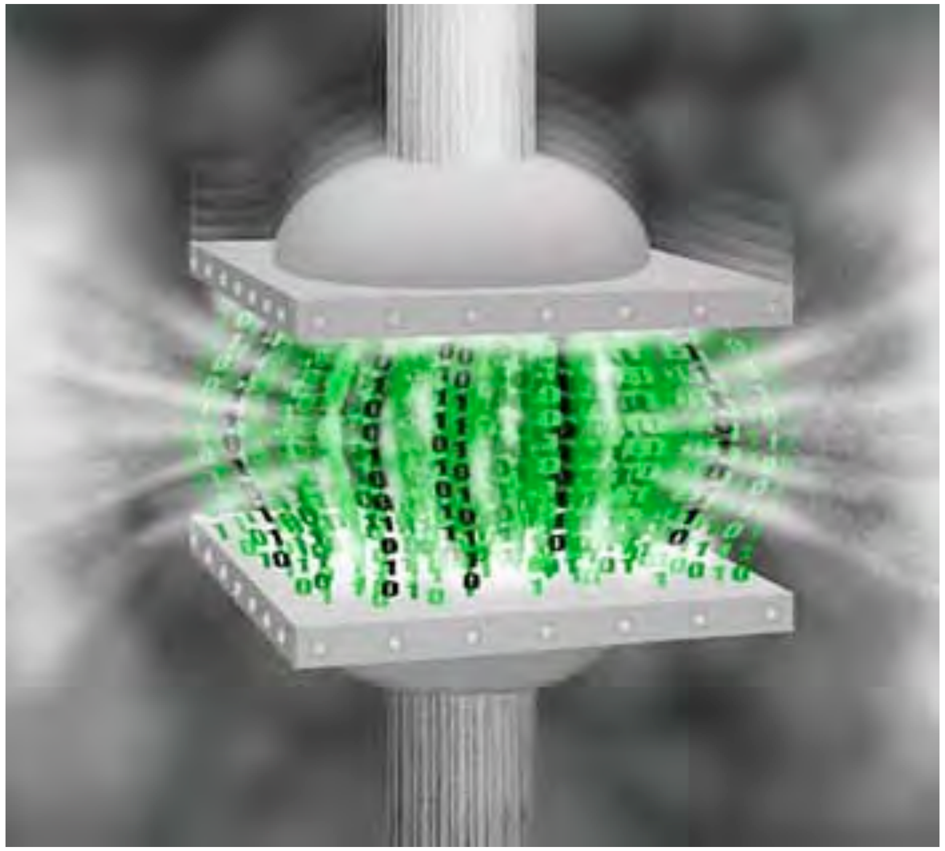


SKA AFRICA
SQUARE KILOMETRE ARRAY



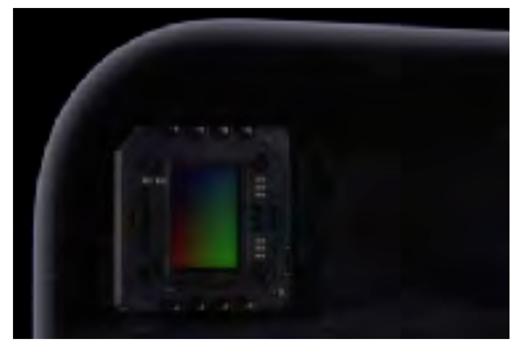
Key tool:

Compression



A familiar example

12MPix



Power: OK

Talk time (wireless):
Up to 21 hours on 3G

Standby:
Up to 16 days

Internet use:
Up to 13 hours on 3G

Up to 13 hours on LTE

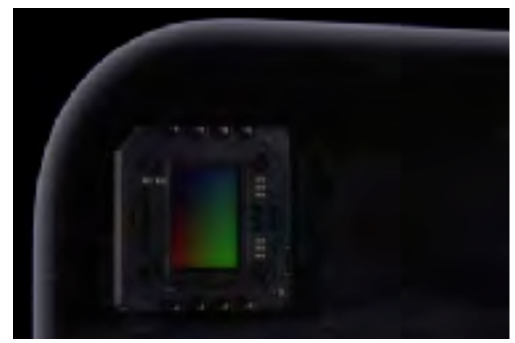
Up to 15 hours on Wi-Fi

Wireless video playback:
Up to 14 hours

Wireless audio playback:
Up to 60 hours

A familiar example

12MPix & 24bits/pixel
= 36MB



Power: OK
Storage: NO

Talk time (wireless):
Up to 21 hours on 3G
Standby:
Up to 16 days
Internet use:
Up to 13 hours on 3G
Up to 13 hours on LTE
Up to 15 hours on Wi-Fi
Wireless video playback:
Up to 14 hours
Wireless audio playback:
Up to 60 hours

iPhone 7 Plus 32GB Price in Switzerland :- 837CHF
iPhone 7 Plus 128GB Price in Switzerland :- 947CHF
iPhone 7 Plus 256GB Price in Switzerland :- 1057CHF

$\approx$ 1000 images

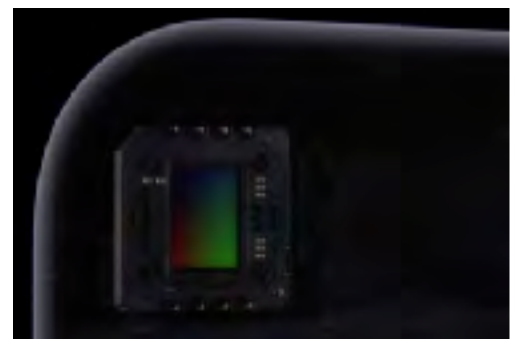
$+$ no apps

A familiar example

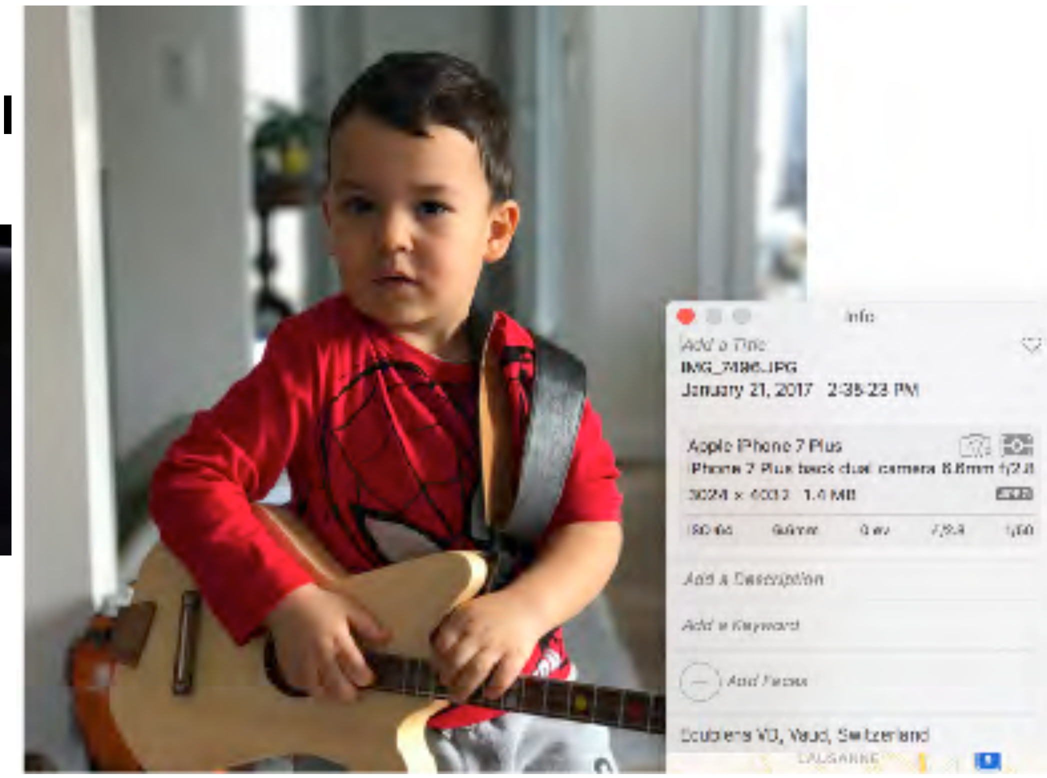
Power: OK
Storage: OK

Talk time (wireless):
Up to 21 hours on 3G
Standby:
Up to 16 days
Internet use:
Up to 13 hours on 3G
Up to 13 hours on LTE
Up to 15 hours on Wi-Fi
Wireless video playback:
Up to 14 hours
Wireless audio playback:
Up to 60 hours

12MPix & 24bits/pixel = 36MB



Compression ➥

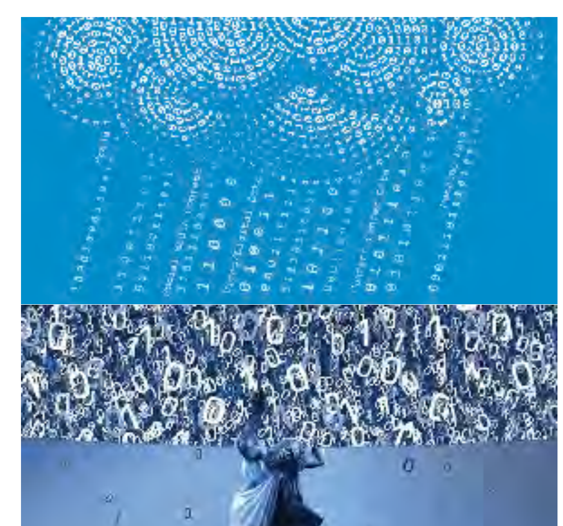


iPhone 7 Plus 32GB Price in Switzerland :- 837CHF
iPhone 7 Plus 128GB Price in Switzerland :- 947CHF
iPhone 7 Plus 256GB Price in Switzerland :- 1057CHF

actual: 1.4MB

$\approx$ 25000 images

+ no apps

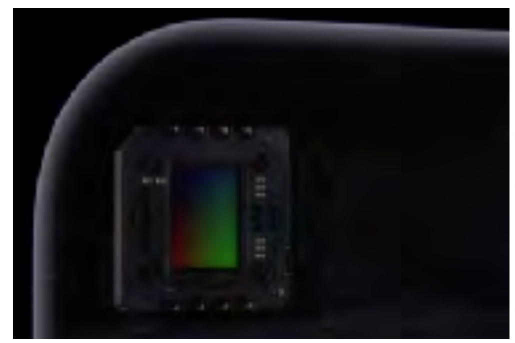


(vs 1000 images)

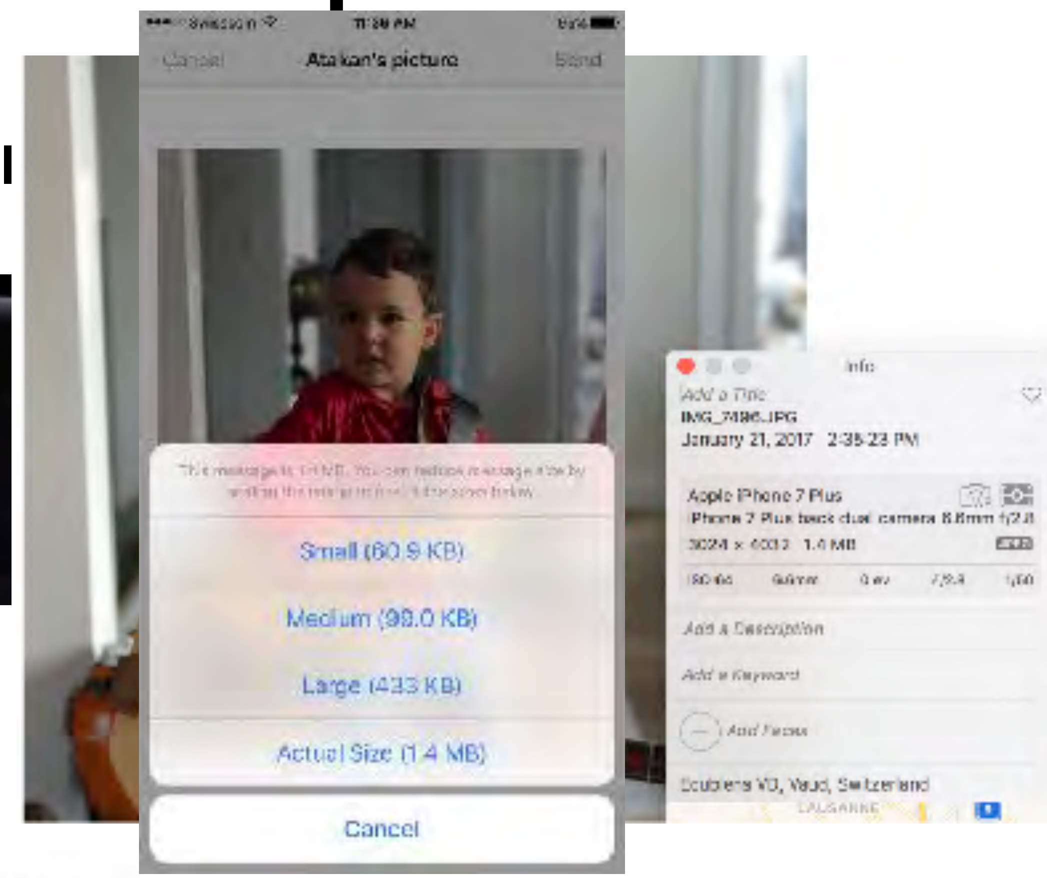
A familiar example

Bandwidth: OK

12MPix & 24bits/pixel = 36MB



Compression ➥



Power: OK
Storage: OK

Talk time (wireless):
Up to 21 hours on 3G

Standby:
Up to 16 days

Internet use:
Up to 13 hours on 3G
Up to 13 hours on LTE
Up to 15 hours on Wi-Fi

Wireless video playback:
Up to 14 hours

Wireless audio playback:
Up to 60 hours

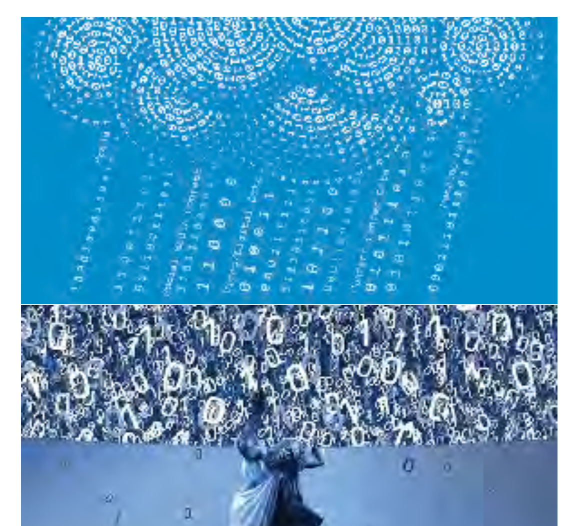
iPhone 7 Plus 32GB Price in Switzerland :- 837CHF
iPhone 7 Plus 128GB Price in Switzerland :- 947CHF
iPhone 7 Plus 256GB Price in Switzerland :- 1057CHF

$\approx$ 25000 images

+ no apps

(vs 1000 images)

actual: 1.4MB



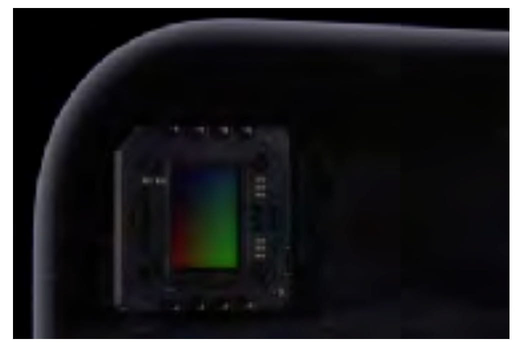
Compression helps!

Bandwidth: OK

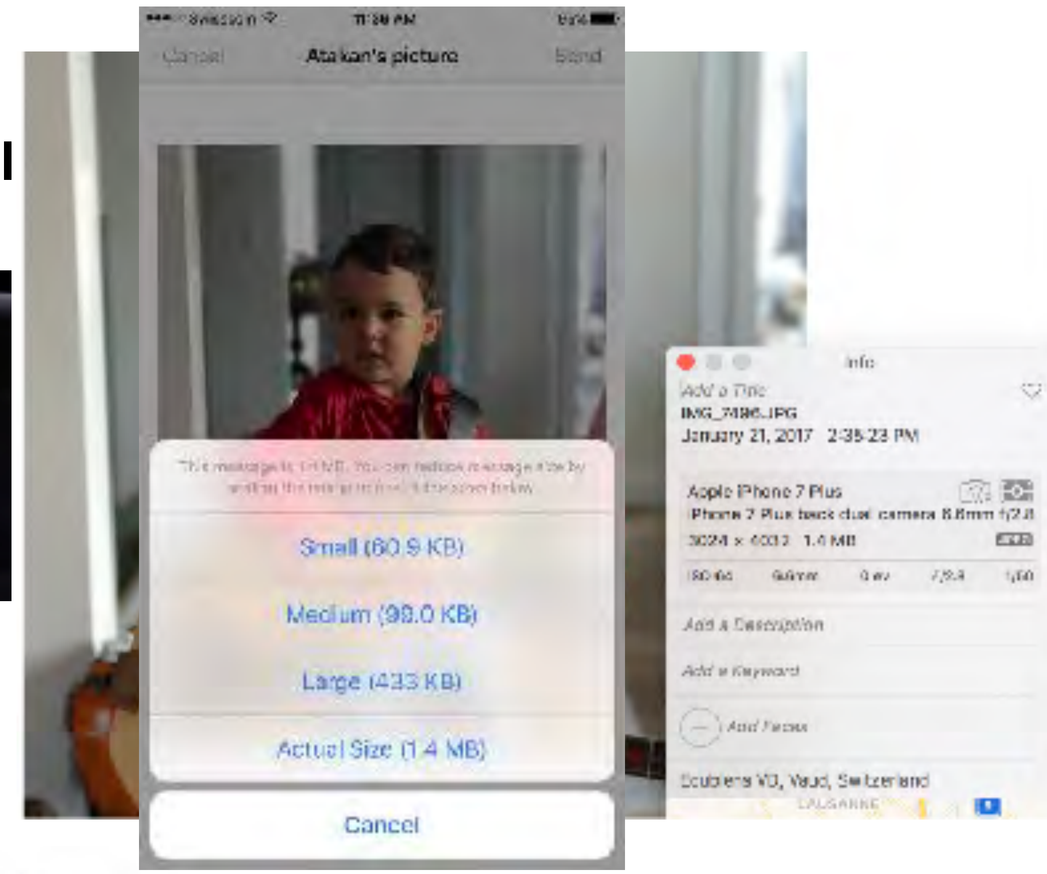
Power: OK
Storage: OK

Talk time (wireless):
Up to 21 hours on 3G
Standby:
Up to 16 days
Internet use:
Up to 13 hours on 3G
Up to 13 hours on LTE
Up to 15 hours on Wi-Fi
Wireless video playback:
Up to 14 hours
Wireless audio playback:
Up to 60 hours

12MPix & 24bits/pixel = 36MB



Compression ➥



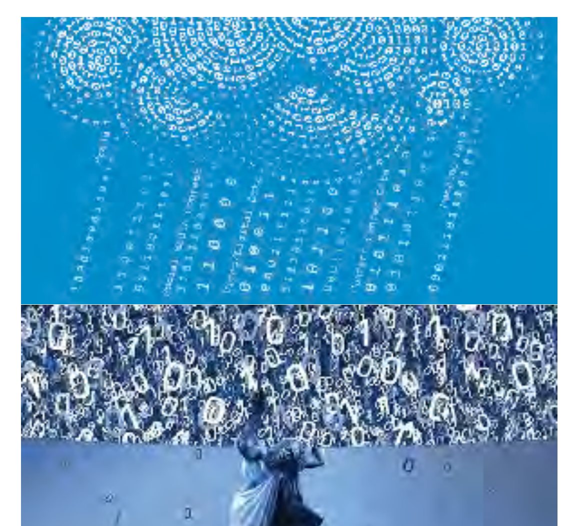
actual: 1.4MB

iPhone 7 Plus 32GB Price in Switzerland :- 837CHF
iPhone 7 Plus 128GB Price in Switzerland :- 947CHF
iPhone 7 Plus 256GB Price in Switzerland :- 1057CHF

≈ 25000 images

$+$ no apps

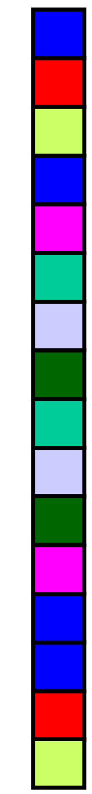
(vs 1000 images)



Compression: The basics

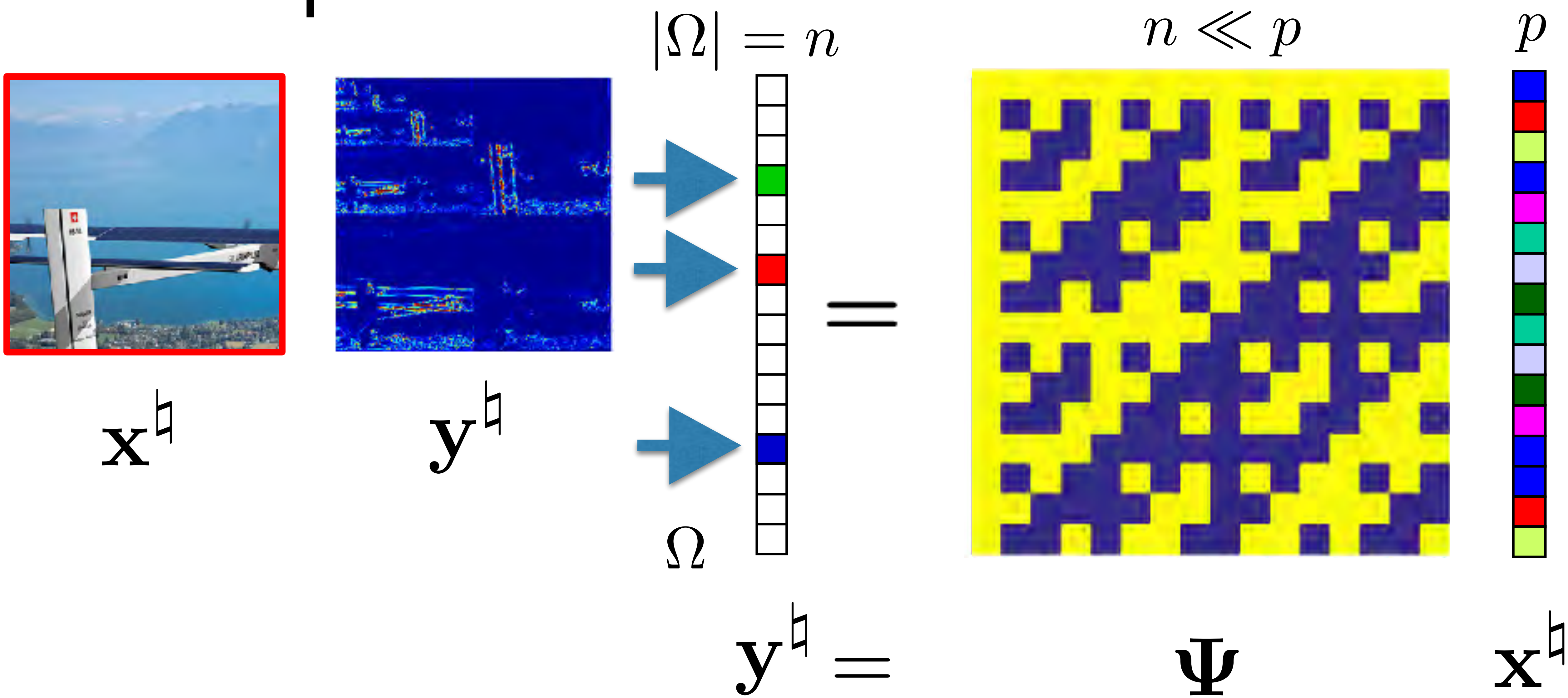
$\mathbf{x}^\natural$

p



$\mathbf{x}^\natural$

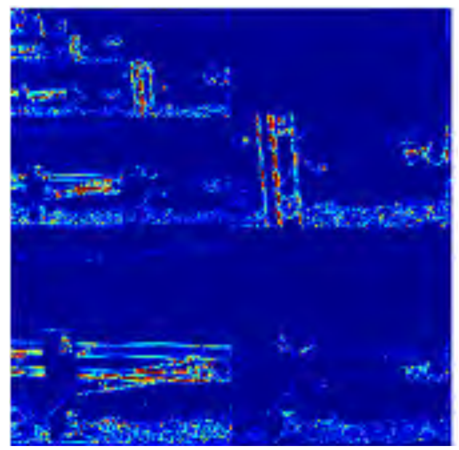
Compression: The basics

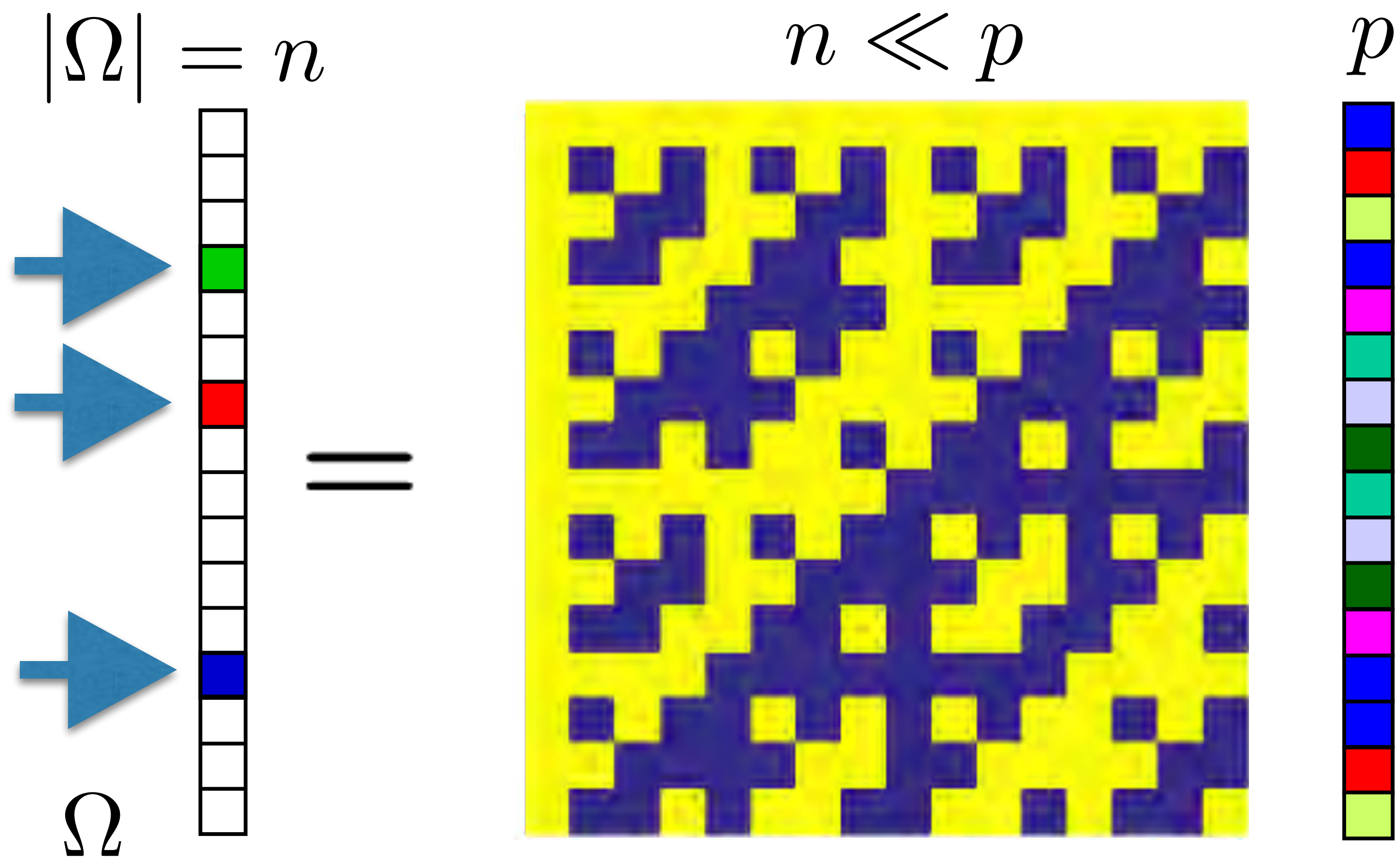


JPEG2000: Wavelets

sparsity

Compression: The basics





$\mathbf{x}^\natural$ $\mathbf{y}^\natural$

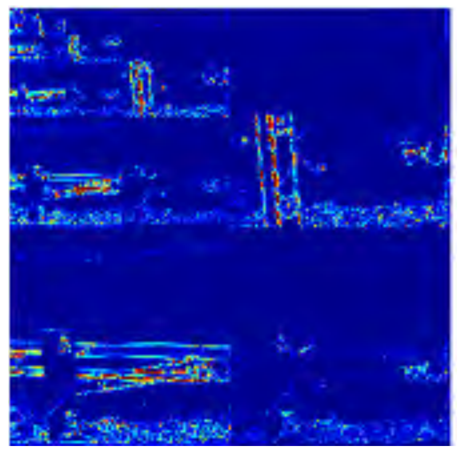
Add decades
&
math + eng

$$\mathbf{y}^\natural = \mathbf{\Psi} \mathbf{x}^\natural$$

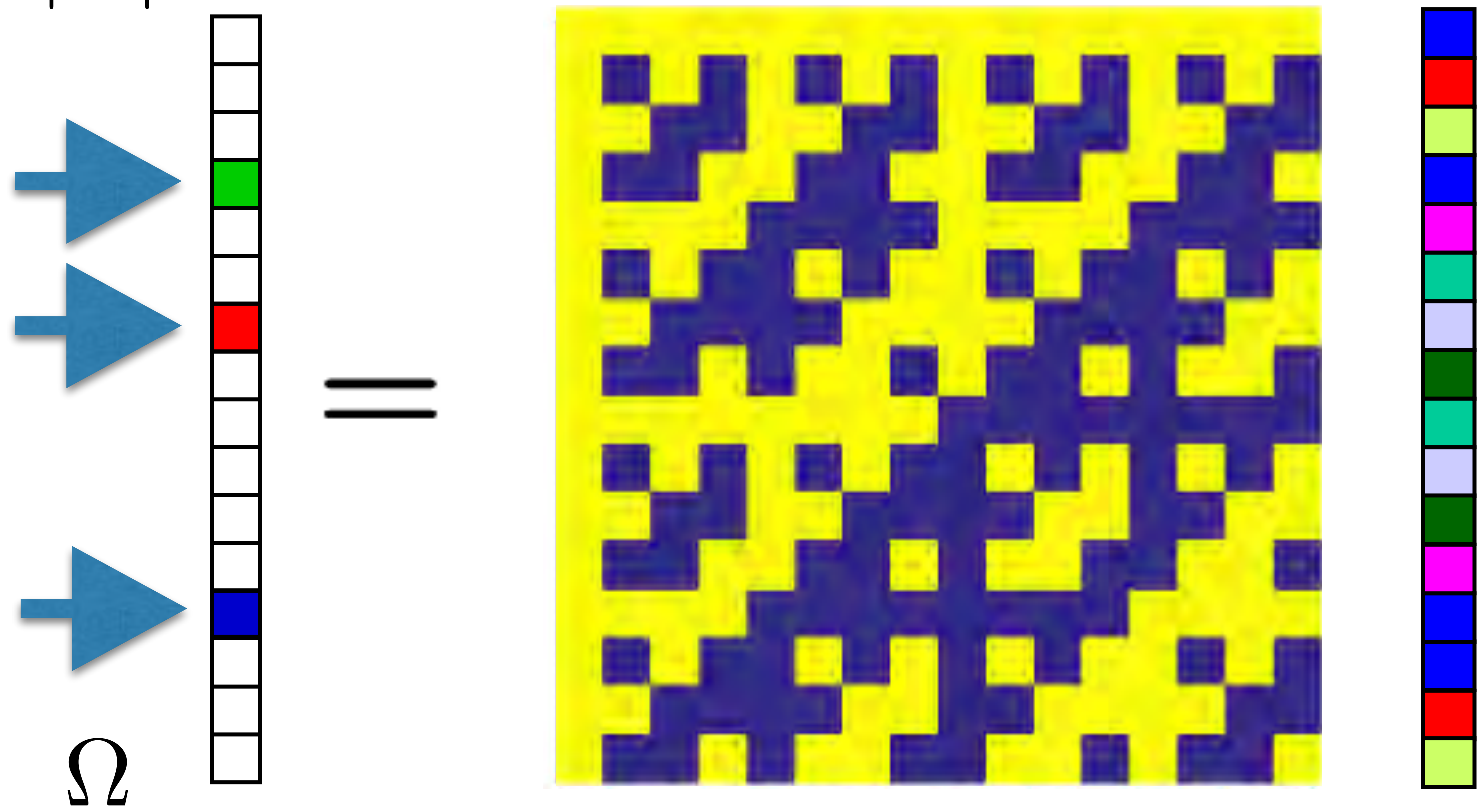
JPEG2000: Wavelets

sparsity

Compression: The basics



$\mathbf{x}^\natural$ $\mathbf{y}^\natural$



Add decades
&
math + eng

$$\mathbf{y}^\natural = \mathbf{\Psi} \quad \mathbf{x}^\natural$$

JPEG2000: Wavelets

Strategy: Encode $b = P_\Omega \Psi x^\natural$

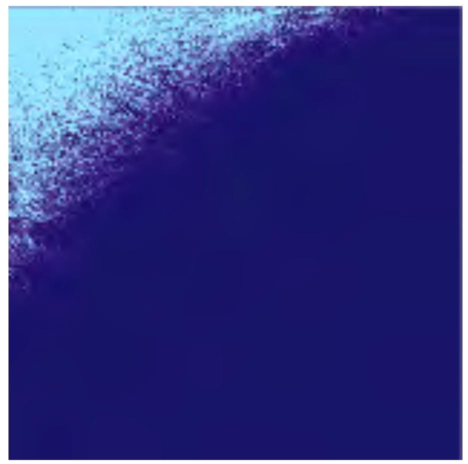
P_Ω : Subset selector

Decode $\hat{x} = \Psi^* P_\Omega^* b$

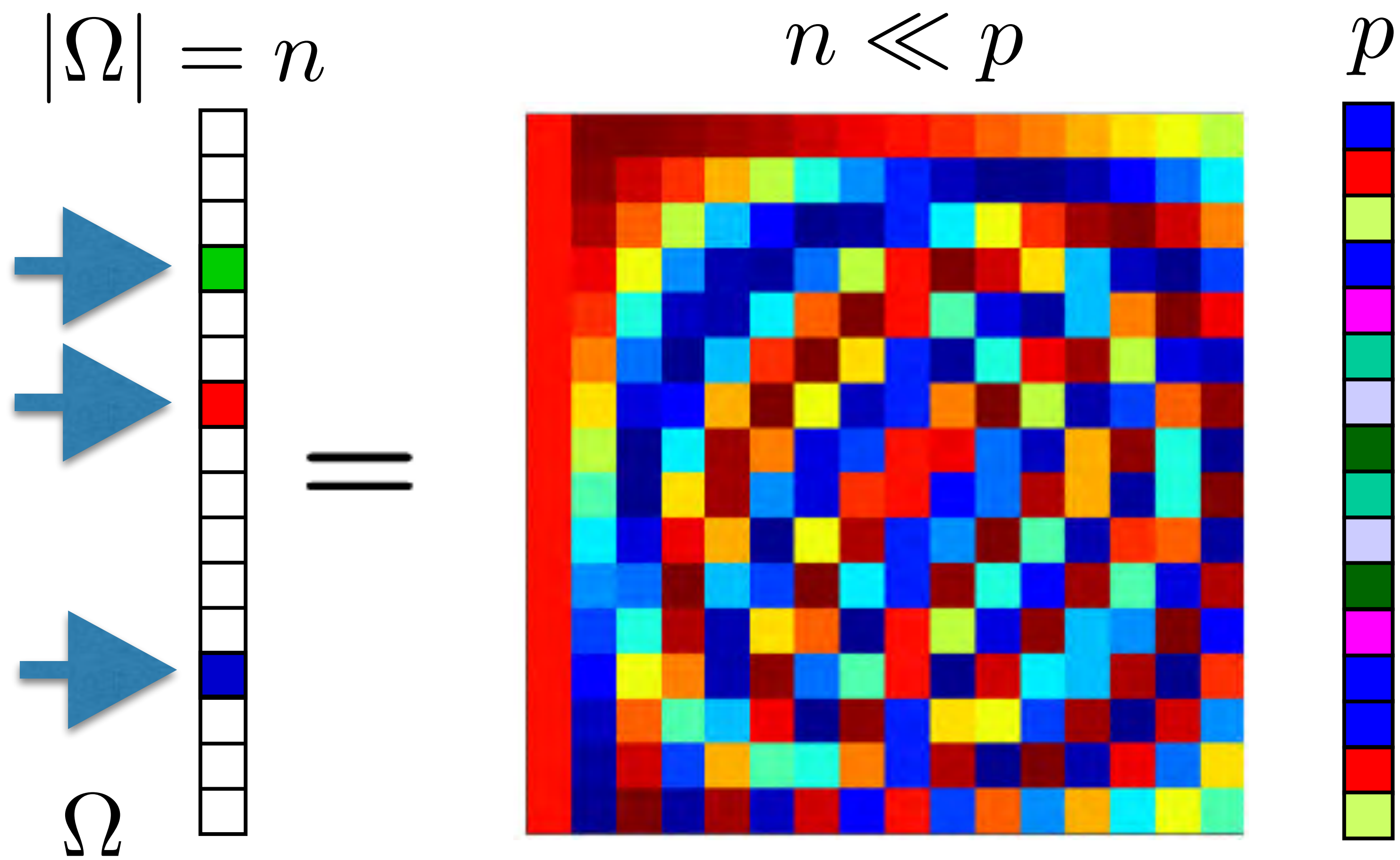
sparsity

Compression: The basics

$\mathbf{x}^\natural$



$\mathbf{y}^\natural$



$$\mathbf{y}^\natural = \boldsymbol{\Psi} \quad \mathbf{x}^\natural$$

Add decades
&
math + eng

Strategy: Encode $b = P_\Omega \Psi x^\natural$

P_Ω : Subset selector

Decode $\hat{x} = \Psi^* P_\Omega^* b$

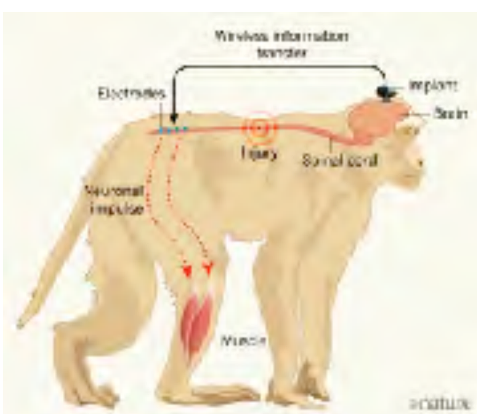
JPEG: DCT

sparsity

The core challenge:

"Can we automatically teach any sensor how to compress its own data well?"

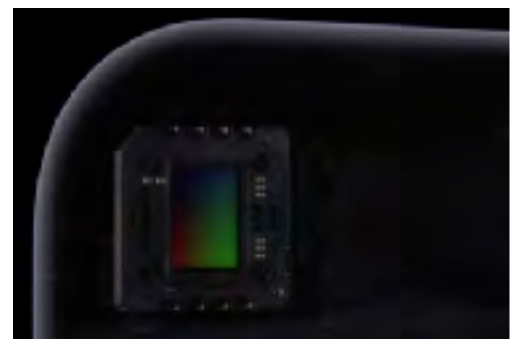
SKA AFRICA
SQUARE KILOMETRE ARRAY



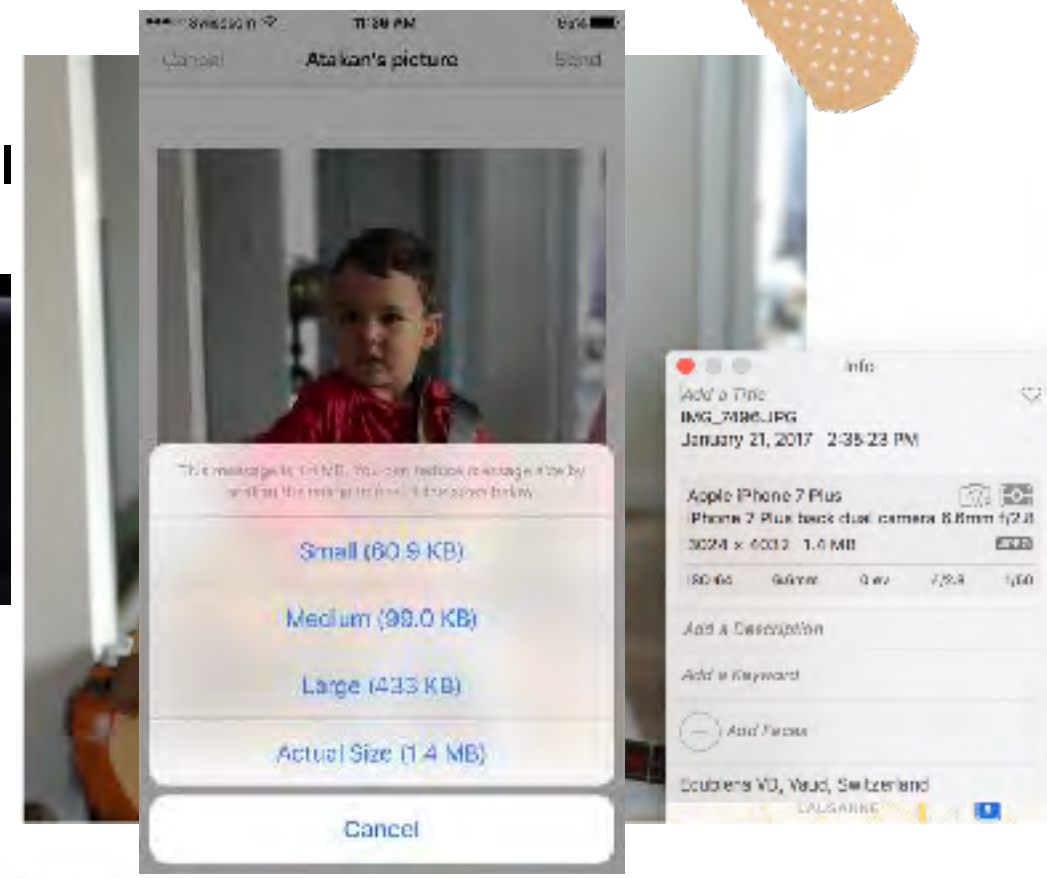
Compression helps!

Bandwidth: OK

12MPix & 24bits/pixel = 36MB



Compression ➥



actual: 1.4MB

Power: OK
Storage: OK

Talk time (wireless):
Up to 21 hours on 3G
Standby:
Up to 16 days
Internet use:
Up to 13 hours on 3G
Up to 13 hours on LTE
Up to 15 hours on Wi-Fi
Wireless video playback:
Up to 14 hours
Wireless audio playback:
Up to 60 hours

iPhone 7 Plus 32GB Price in Switzerland :- 837CHF
iPhone 7 Plus 128GB Price in Switzerland :- 947CHF
iPhone 7 Plus 256GB Price in Switzerland :- 1057CHF

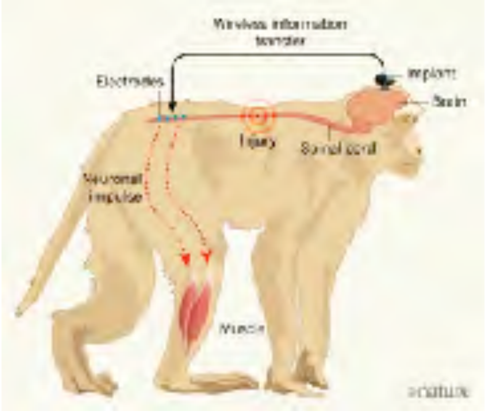
Caveats for generalization:

Collected the full data &
Performed a full transformation!

The core challenge:

“Can we automatically teach any sensor how to compress its own data well?”

SKA AFRICA
SQUARE KILOMETRE ARRAY



Our twist:

Compress without transforming or sampling the whole data!


KEEP
CALM
AND
FASTEN YOUR
SEAT BELTS

(Old) Compressive sensing (CS)

- Goal: *Directly obtain the compressed version*
- Off-load the difficulty to computation
  - encoding model: $b = P_\Omega \mathcal{F} x^\natural$ & $x^\natural$ is s sparse in Ψ
  - decoding algorithm: convex optimization

$$\hat{x} = \arg\min_x \{\|\Psi^* x\|_1 : b = P_\Omega \mathcal{F} x\}$$

  - Theorem: If $|\Omega| \geq s\,(\log p)^\gamma$ & Ω is sufficiently random then $\hat{x} = x^\natural$ with hp

[Candes et al 2006;…]

Challenges to the old CS

- High computational cost & latency: $\mathcal{O}(n^2 p^{1.5})$
- Oversampling: p vs s vs $s(\log p)^\gamma$
- Dictionary Ψ: ***hidden*** need for training data

``When solving a given problem, try to avoid a more general problem as an intermediate step.''

–Vladimir Vapnik
[main developer of statistical learning theory (along with Alexey Chervonenkis)]

Given training data, we will bypass dictionary learning & design the whole compressive sampling system directly

Statistical Learning Theory
meets Compressive Sensing

Ω

Learning data triage (simplified)

A statistical learning framework for CS with sample signals

- *Probabilistic model:* $y = P_\Omega \mathcal{F} x^\natural$
  - $x^\natural$ follows some **unknown** probability distribution $\mathbb{P}$.

- *Sample signals:* $\{x_i\}_{i \le m}$, i.i.d. random vectors following $\mathbb{P}$

- *Fix* an estimator: $\hat{x} = \mathcal{F}^H P_\Omega^T y = (P_\Omega \mathcal{F})^\dagger y$

- *Loss function:* $\mathcal{L}(x^\natural; \Omega) = \frac{\|\hat{x} - x^\natural\|_2^2}{\|x^\natural\|_2^2}$

- *Goal:* Fix $|\Omega| = n$. Find a sub-sampling pattern Ω, given $\{x_i\}_{i \le m}$, such that the risk $\mathsf{E}\,\mathcal{L}(x^\natural; \Omega)$ is minimized.

A statistical learning framework for CS with sample signals

- *Probabilistic model:* $y = P_\Omega \mathcal{F} x^\natural$
  - $x^\natural$ follows some **unknown** probability distribution $\mathbb{P}$.
- *Sample signals:* $\{x_i\}_{i \le m}$, i.i.d. random vectors following $\mathbb{P}$
- *Fix* an estimator: $\hat{x} = \mathcal{F}^H P_\Omega^T y = (P_\Omega \mathcal{F})^\dagger y$ ☜ *simplification is here*
- *Loss function:* $\mathcal{L}(x^\natural; \Omega) = \frac{\|\hat{x} - x^\natural\|_2^2}{\|x^\natural\|_2^2}$
- *Goal:* Fix $|\Omega| = n$. Find a sub-sampling pattern Ω, given $\{x_i\}_{i \le m}$, such that the risk $\mathsf{E}\,\mathcal{L}(x^\natural; \Omega)$ is minimized. ☜ *simplification is here*

Empirical risk minimization-I

If $\mathbb{P}$ *were* known, the optimal Ω is given by solving the discrete optimization problem:

$$\Omega_{\mathsf{opt}} \in \operatorname*{arg\,min}_{\Omega : |\Omega| \le n} \mathsf{E}\, \mathcal{L}(x^\natural; \Omega)$$

Proposition

We have $\mathcal{L}(x^\natural; \Omega) = 1 - \frac{\|P_\Omega \mathcal{F} x^\natural\|_2^2}{\|x^\natural\|_2^2} =: 1 - \ell(x^\natural; \Omega)$.

Therefore, we can write

$$\Omega_{\mathsf{opt}} \in \operatorname*{arg\,max}_{\Omega : |\Omega| \le n} \mathsf{E}\, \ell(x^\natural; \Omega),$$

and we have

$$\mathsf{E}\, \mathcal{L}(x^\natural; \Omega_{\mathsf{opt}}) = \min_{\Omega : |\Omega| \le n} \mathsf{E}\, \mathcal{L}(x^\natural; \Omega) = 1 - \mathsf{E}\, \frac{\|P_{\Omega_{\mathsf{opt}}} \mathcal{F} x^\natural\|_2^2}{\|x^\natural\|_2^2} =: 1 - \varepsilon_{\mathbb{P}}.$$

Empirical risk minimization-II

While $\mathbb{P}$ is unknown, we have i.i.d. samples $\{x_i\}_{i\leq n}$ from $\mathbb{P}$.

Hence we may consider the *empirical risk minimizer* given by:

$$\hat{\Omega} \in \arg\max_{\Omega:|\Omega|\leq n} \frac{1}{m} \sum_{i\leq m} \ell(x_i;\Omega).$$

Since in general $\hat{\Omega} \neq \Omega_{\mathsf{opt}}$, we can only expect that

$$\mathsf{E}\,\mathcal{L}(x^\natural;\hat{\Omega}) \leq \mathsf{E}\,\mathcal{L}(x^\natural;\Omega_{\mathsf{opt}}) + \varepsilon_m = 1 - \varepsilon_{\mathbb{P}} + \varepsilon_m.$$

Statistical analysis

Recall that $\mathsf{E}\,\mathcal{L}(x^\natural; \hat{\Omega}) \leq \mathsf{E}\,\mathcal{L}(x^\natural; \Omega_{\mathsf{opt}}) + \varepsilon_m = 1 - \varepsilon_{\mathbb{P}} + \varepsilon_m$.

Theorem

For any $\beta \in (0, 1)$, we have

$$\varepsilon_m \leq \sqrt{\frac{2}{m}\left[\log\binom{p}{n} + \log\left(\frac{2}{\beta}\right)\right]},$$

with probability at least $1 - \beta$.

Corollary

Number of sample signals required is of $O(n \log p)$.

Solving the discrete optimization problem

Define $\tilde{x}_i = x_i / \|x_i\|_2$. Recall that

$$\hat{\Omega} \in = \arg\max_{\Omega:|\Omega| \le n} \sum_{i \le m} \frac{\|P_\Omega \mathcal{F} x_i\|_2^2}{\|x_i\|_2^2} = \arg\max_{\Omega:|\Omega| \le n} \sum_{i \le m} \|P_\Omega \mathcal{F} \tilde{x}_i\|_2^2.$$

Proposition (Existence of a simple greedy algorithm)

Let ϕ_i be the i-th row of $\mathcal{F}$. We can compute $\hat{\Omega}$ *exactly* by the following greedy algorithm.

1. *For all $i \le p$, compute $v_i = \sum_{j \le m} |\langle \phi_i, \tilde{x}_j \rangle|^2$.*
2. *Let Ω be the set of indices of the n largest v_i's.*

- If $\mathcal{F}$ is the Fourier transform, then the computational complexity is $O(mp \log p)$, *nearly linear time*.

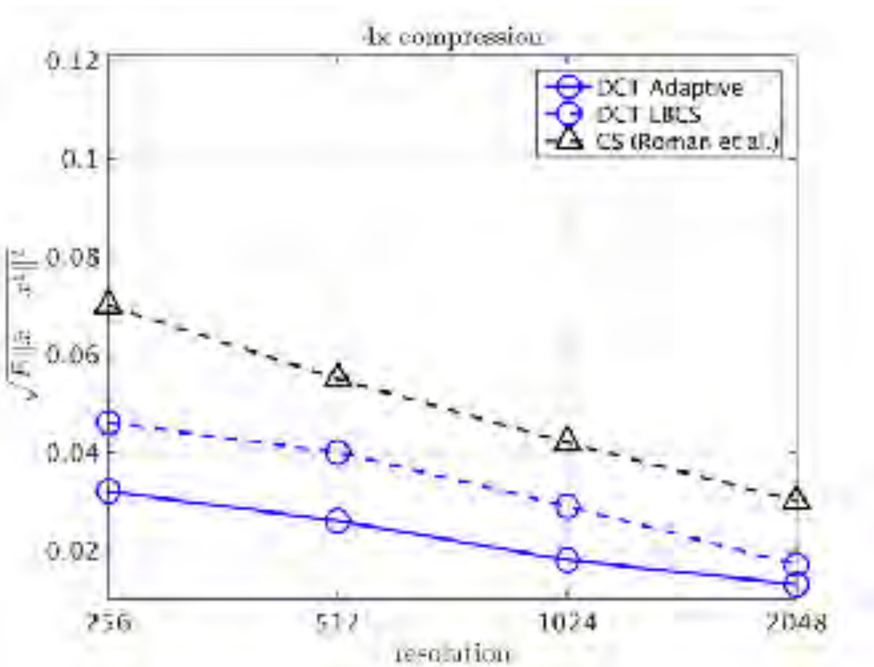
Applications

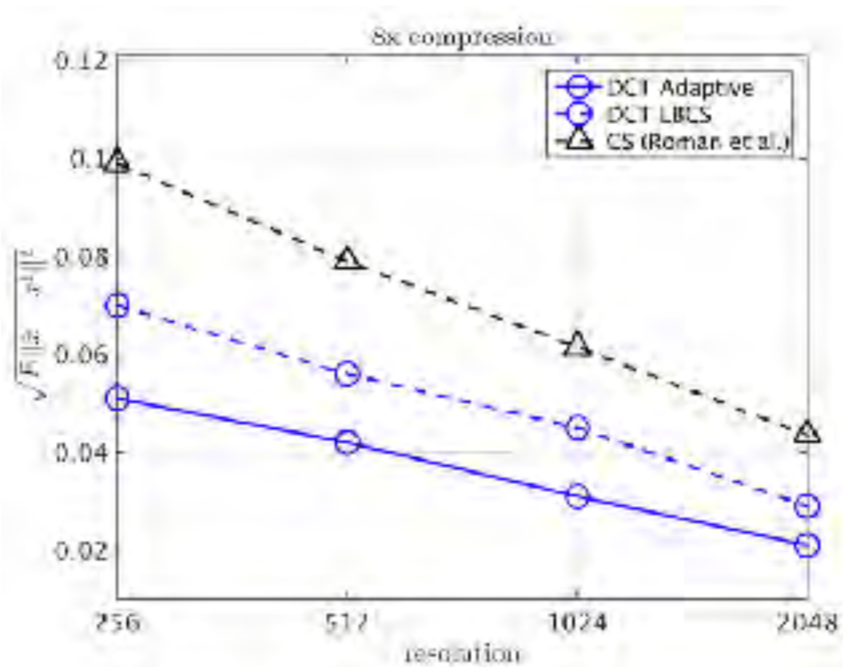
Images - I

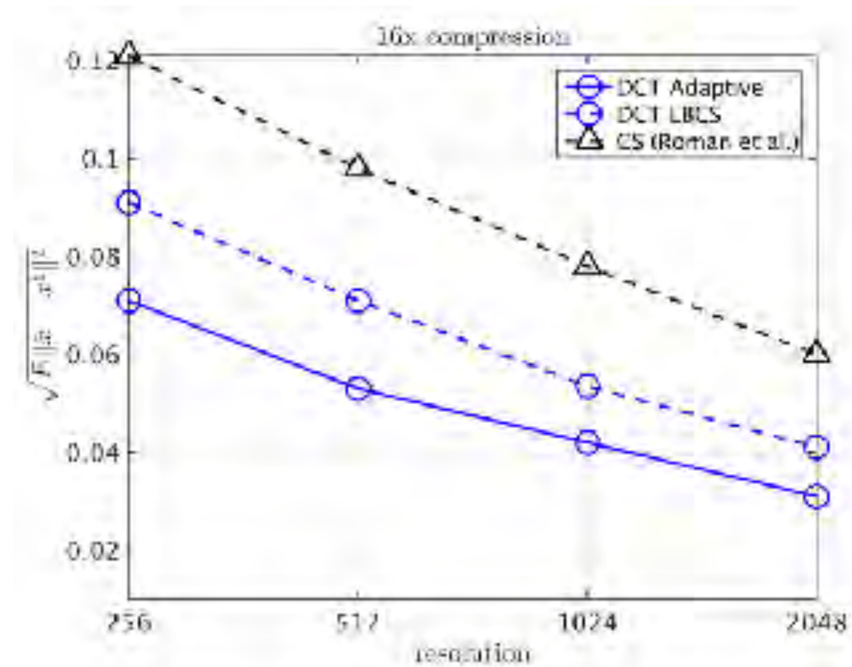




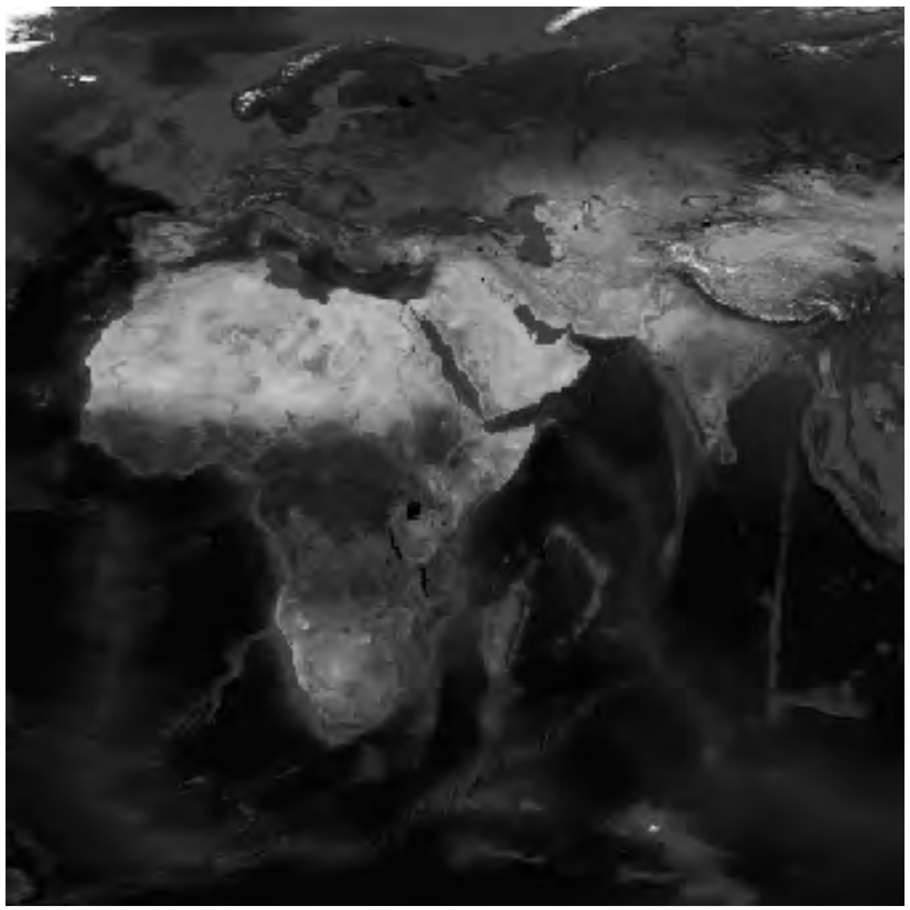
Old CS vs Learning based CS vs JPEG







Images - II

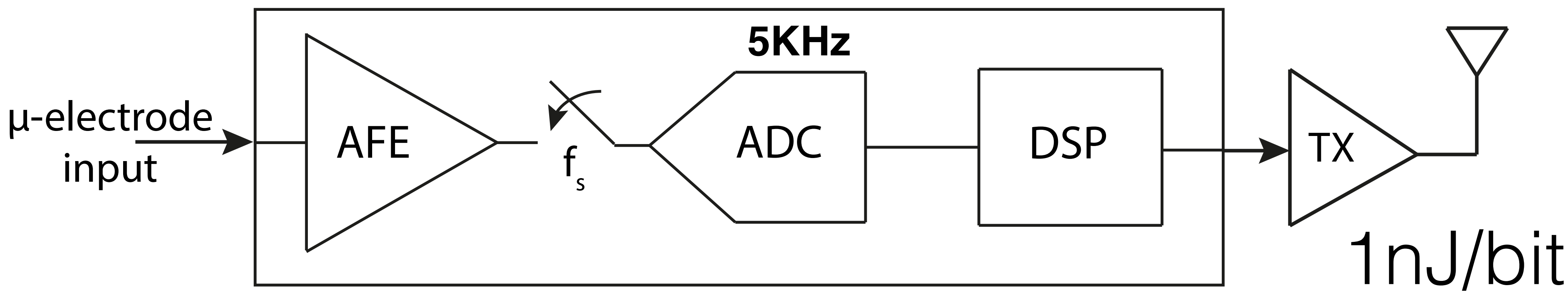


Resolution	Recovery	Sampling rate		
		6.25%	12.50%	25%
256	BP	0.102 / 6s	0.083 / 6s	0.063 / 6s
	TV	0.102 / 27s	0.082 / 22s	0.062 / 20s
	Adjoint	0.103 / 0.01s	0.084 / 0.01s	0.064 / 0.01s
512	BP	0.080 / 23s	0.063 / 22s	0.048 / 22s
	TV	0.080 / 151s	0.063 / 162s	0.047 / 153s
	Adjoint	0.081 / 0.03s	0.064 / 0.03s	0.049 / 0.02s
1024	BP	0.062 / 85s	0.049 / 85s	0.036 / 93s
	TV	0.062 / 340s	0.049 / 614s	0.036 / 65s
	Adjoint	0.063 / 0.08s	0.050 / 0.08s	0.037 / 0.09s
2048	BP	0.047 / 381s	0.036 / 366s	0.026 / 333s
	TV	0.047 / 1561s	0.036 / 2501s	0.025 / 2560s
	Adjoint	0.048 / 0.26s	0.037 / 0.29s	0.027 / 0.28s

1Gpix at 1MPix rate!

Opens up the possibility of streaming video at 30FPS

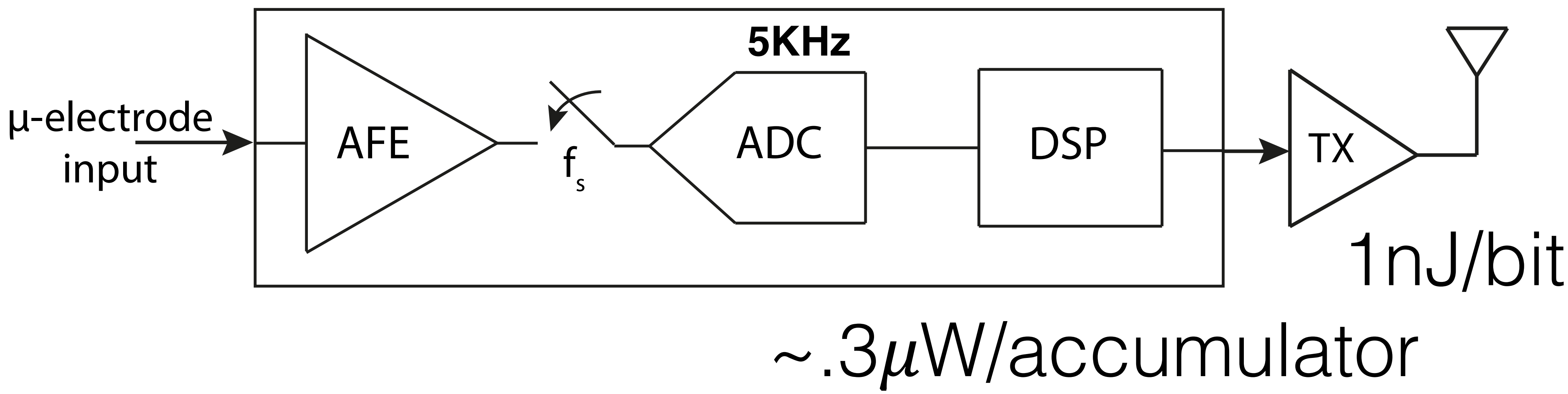
Wireless neural implants - I



~.3μW/accumulator

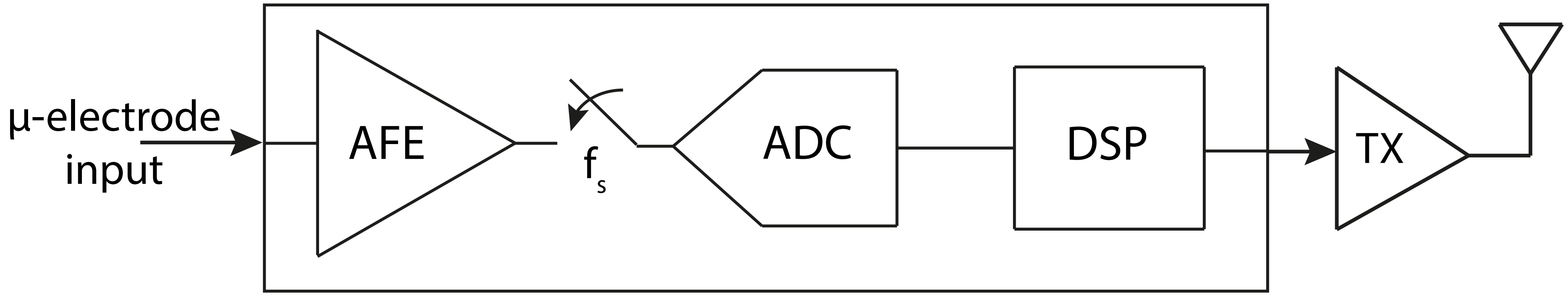
> 30dB quality	Stream out
AFE + ADC	10μW
DSP	0
TX	50μW

Wireless neural implants - I



> 30dB quality	Stream out	Full comp.
AFE + ADC	10μW	10μW
DSP	0	80μW
TX	50μW	~2.5μW

Wireless neural implants - II



- Dataset: billion samples length from `iEEG.org`

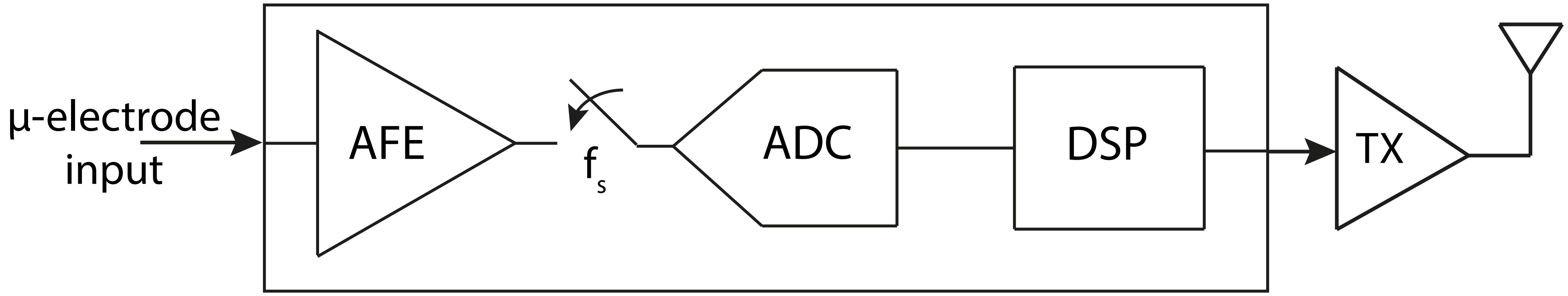
SNR comparison in [**dB**], for **N=256** and **B_i = 10**

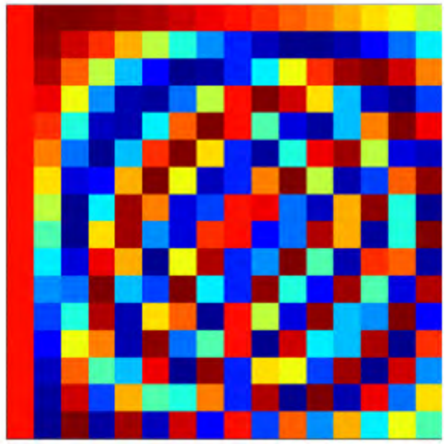
Method	Compression rate					
	2	4	8	16	32	64
LBCS	**40.79**	**37.64**	**33.27**	**28.48**	**23.27**	**18.06**
SHS	36.92	27.96	23.89	20.26	18.53	14.49
BERN	37.48	26.69	20.49	16.87	13.53	11.15
MCS	28.96	24.40	20.92	17.48	n.a.	n.a.

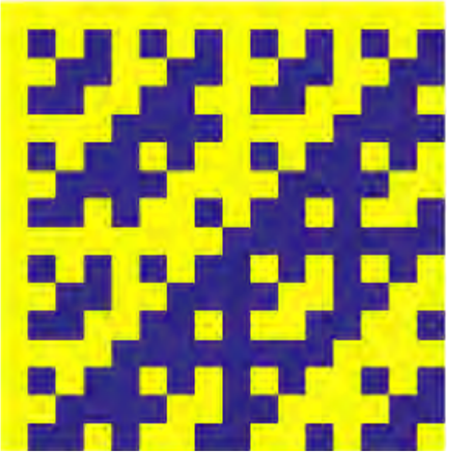
Old CS

- SHS: Structured Hadamard Sampling [Baldassarre, '15]
- BERN: Random Bernoulli [Chen, JSSC '12]
- MCS: Multi-Channel Sampling [Shoaran, TBioCAS'15]

Wireless neural implants - II

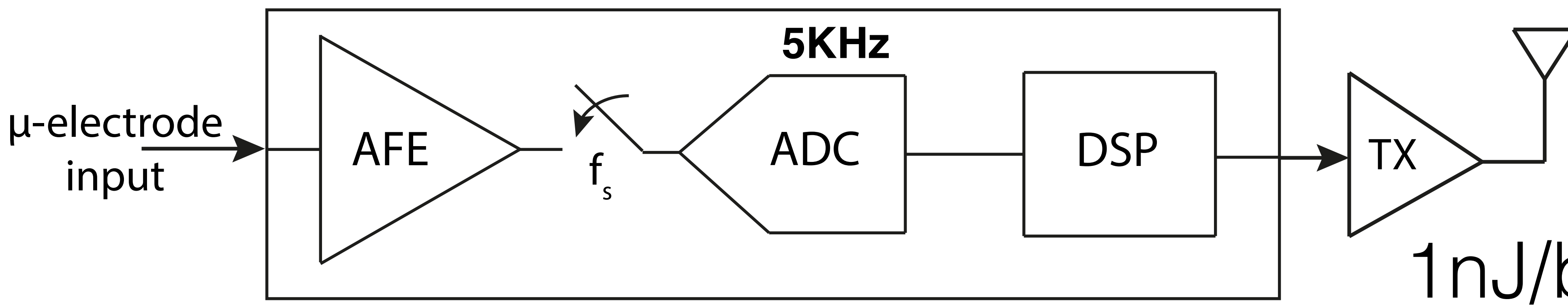






Method	Compression rate					
	2	4	8	16	32	64
DCT Adaptive	42.03	41.96	40.16	37.36	32.88	25.63
DCT LBCS	**41.65**	**40.66**	**38.59**	**35.55**	**31.00**	**23.97**
Had-Adaptive	41.60	39.86	36.38	31.40	25.42	19.43
Had-LBCS	40.79	37.64	33.27	28.48	23.27	18.06
SHS HGL	36.92	27.96	23.89	20.26	18.53	14.49
BERN HGL	37.48	26.69	20.49	16.87	13.53	11.15
MCS HGL	28.96	24.40	20.92	17.48	n.a.	n.a.

Wireless neural implants - II



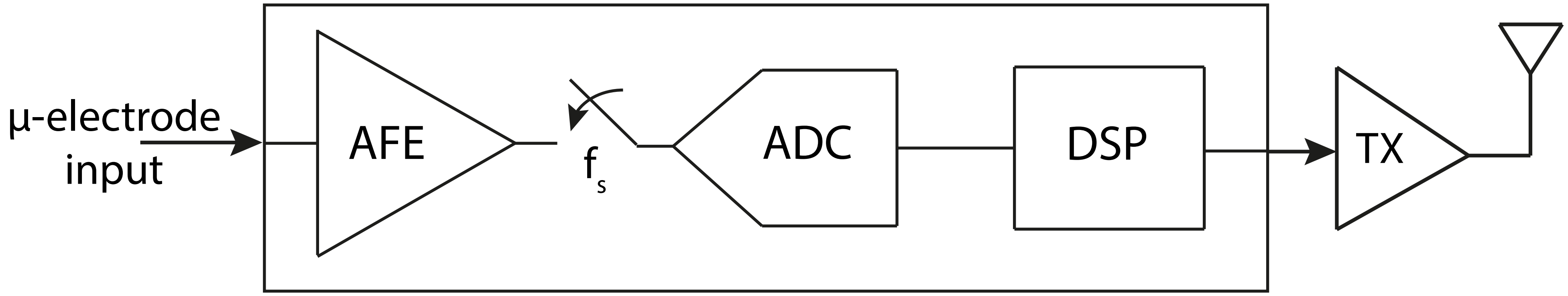
1nJ/bit

~.3μW/accumulator

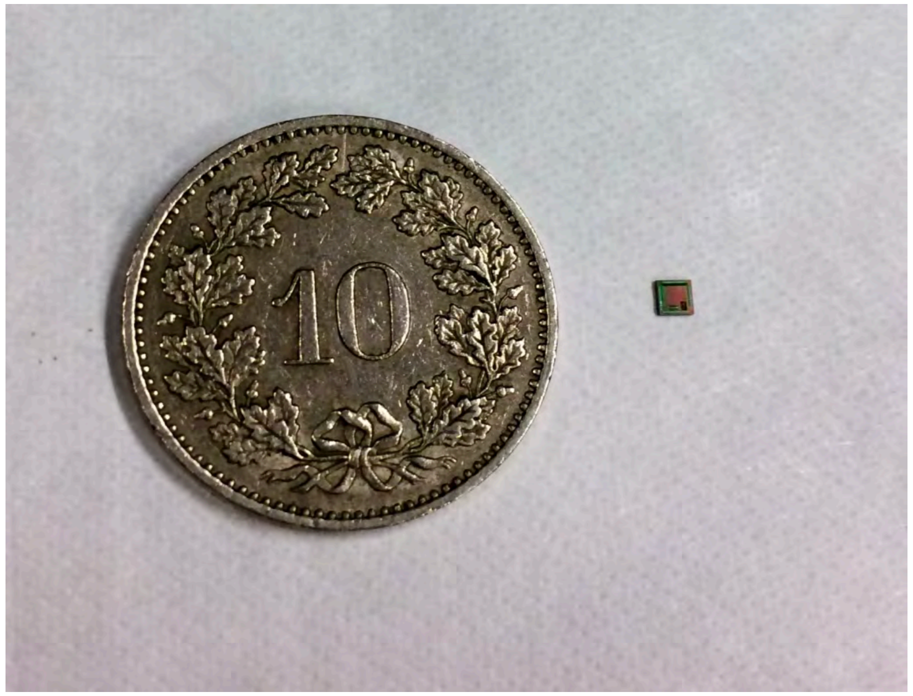
> 30dB quality	Stream out	Full comp.	LBCS
AFE + ADC	10μW	10μW	10μW
DSP	0	80μW	~2.5μW
TX	50μW	~2.5μW	~3μW

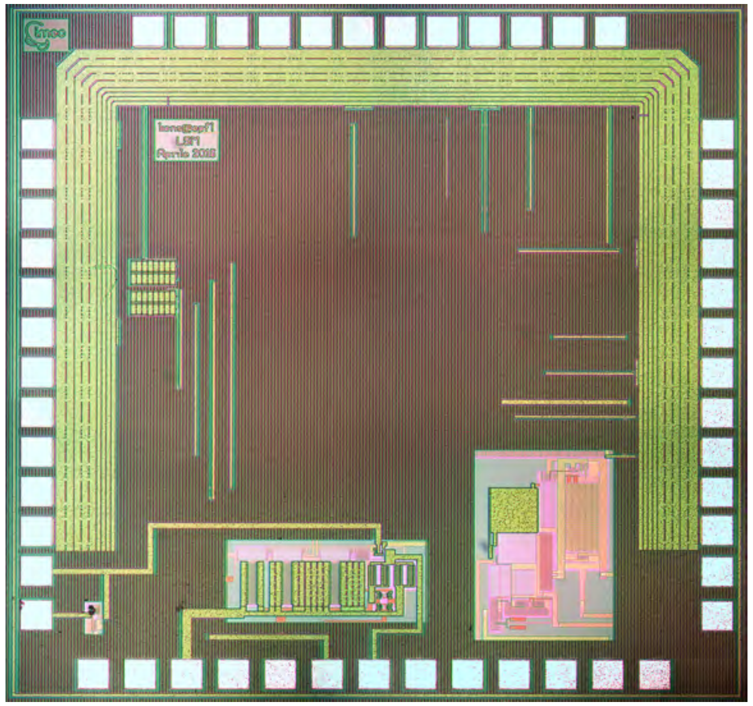
other trade-offs are possible!

Wireless neural implants - III

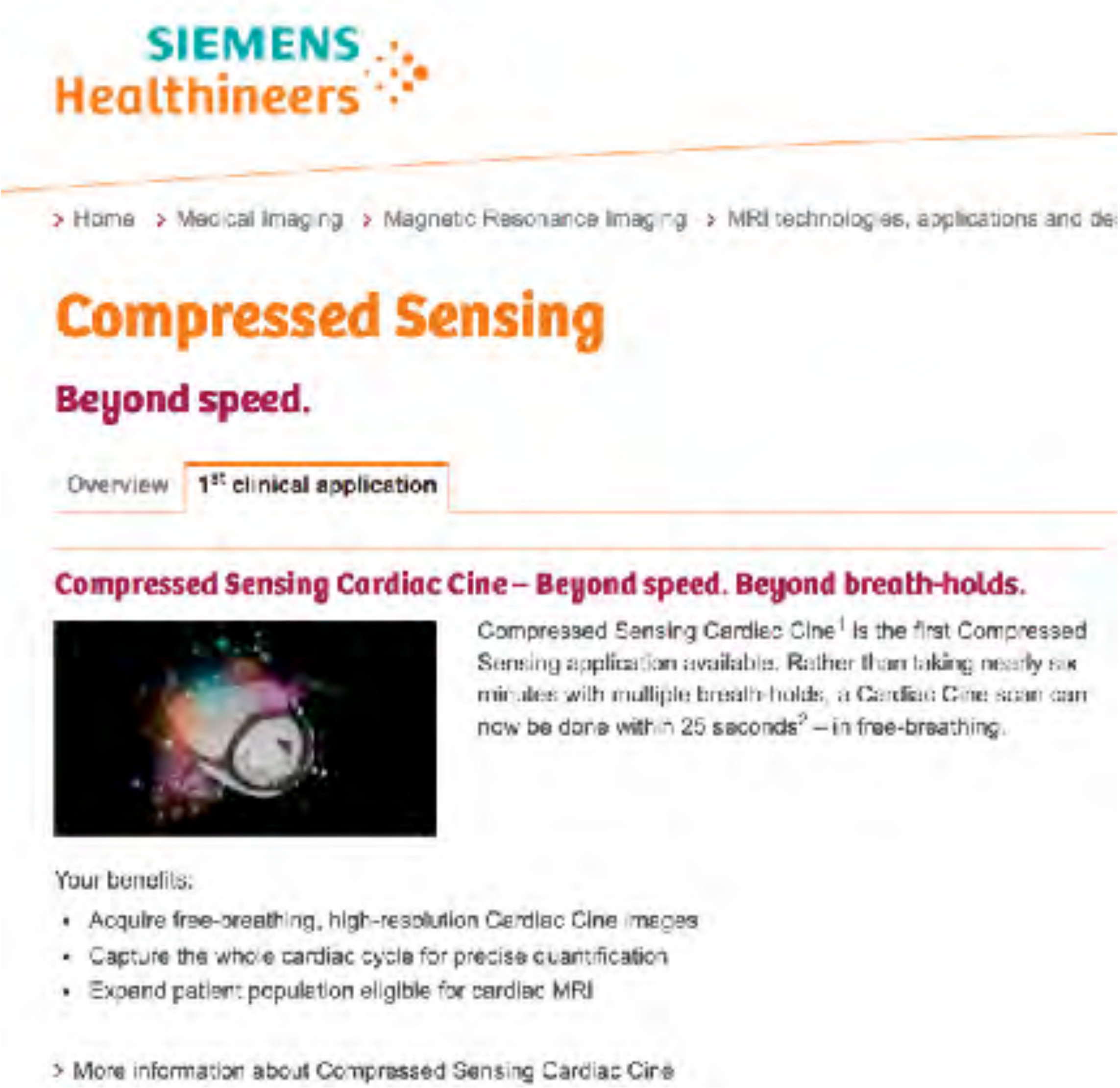


Actual circuit:





Magnetic Resonance Imaging


SIEMENS Healthineers

> Home > Medical Imaging > Magnetic Resonance Imaging > MRI technologies, applications and de

Compressed Sensing

Beyond speed.

Overview | 1st clinical application

Compressed Sensing Cardiac Cine – Beyond speed. Beyond breath-holds.

Compressed Sensing Cardiac Cine[1] is the first Compressed Sensing application available. Rather than taking nearly six minutes with multiple breath-holds, a Cardiac Cine scan can now be done within 25 seconds[2] – in free-breathing.

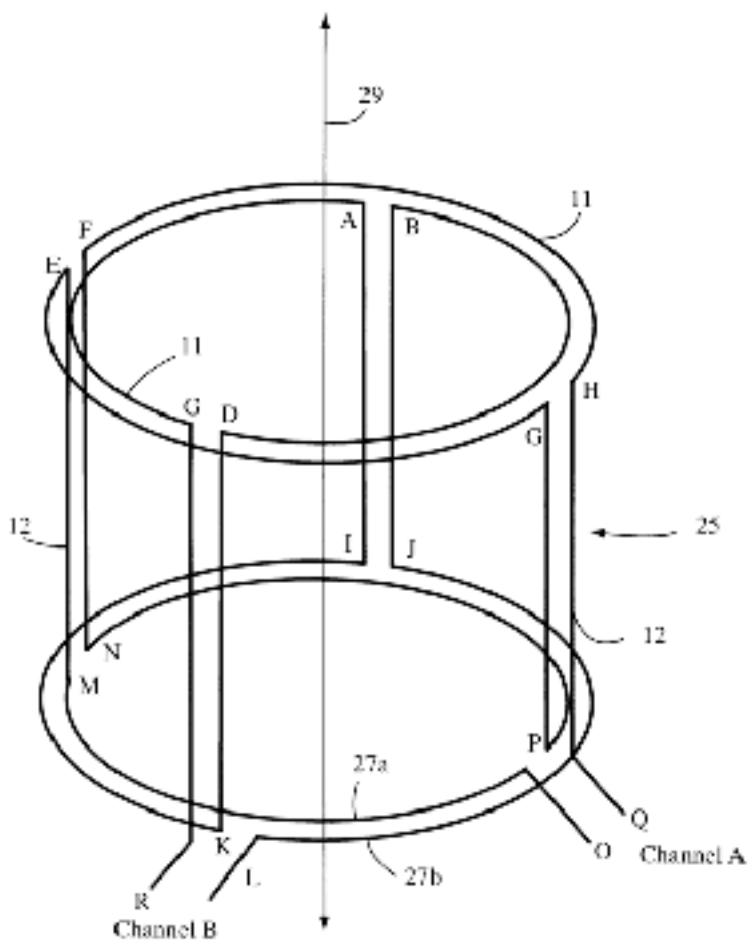
Your benefits:

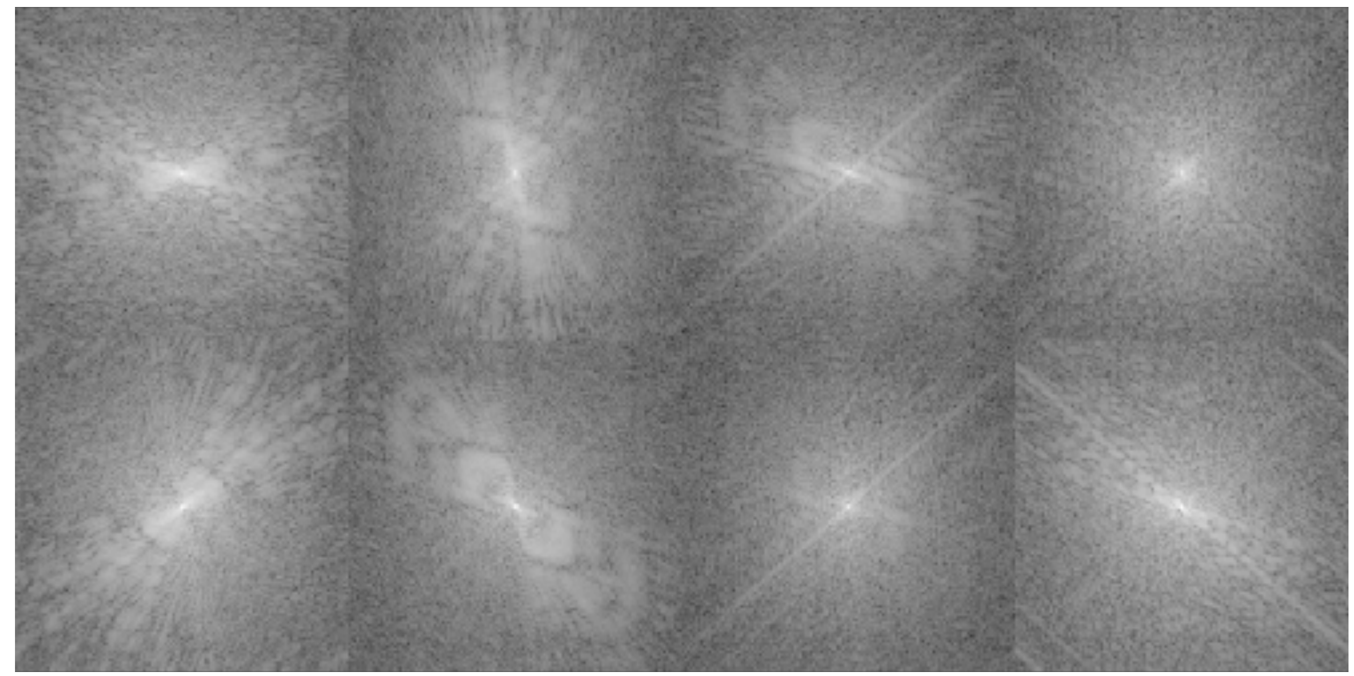
- Acquire free-breathing, high-resolution Cardiac Cine images
- Capture the whole cardiac cycle for precise quantification
- Expand patient population eligible for cardiac MRI

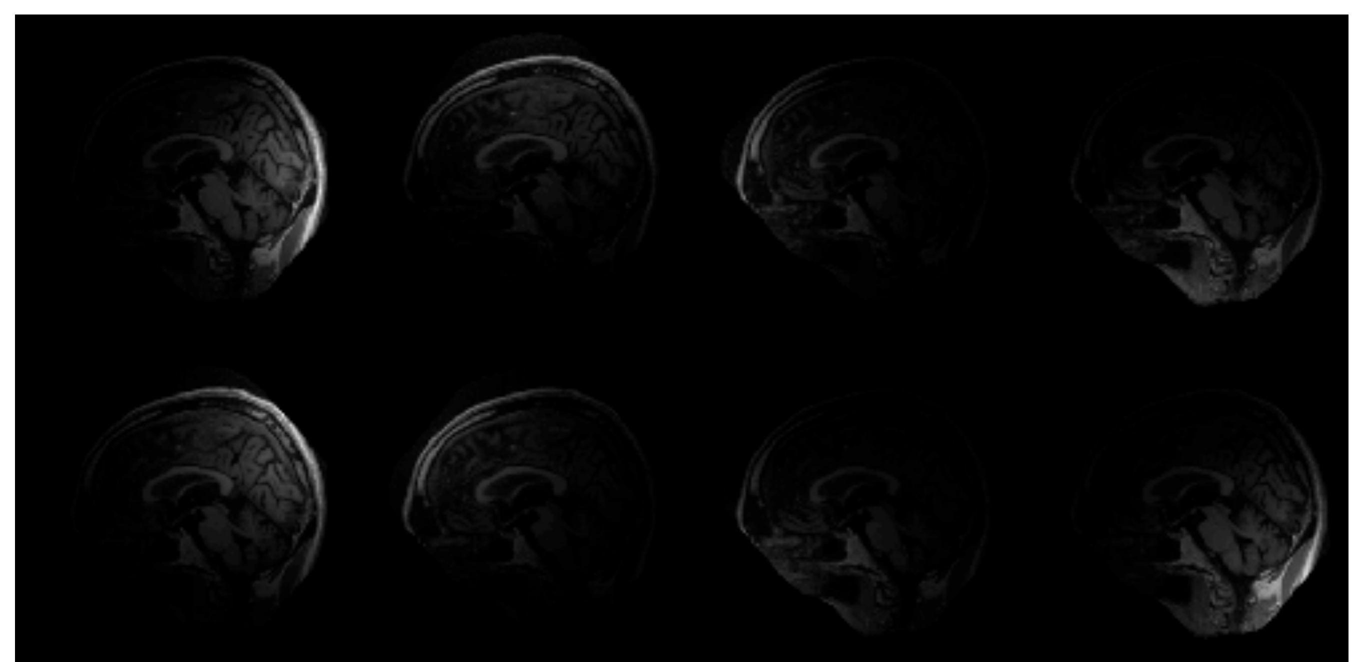
> More information about Compressed Sensing Cardiac Cine

- MRI of the Brain … 20-45 minute scan time.
- MRI of the Orbits … 20-35 minute scan time.
- MRI of the TMJ … 45-60 minute scan time.
- MRI of the Soft Tissue Neck … 25-35 minute scan time.
- MRI of the Cervical Spine … 20-35 minute scan time.
- MRI of the Upper Extremity … 20-45 minute scan time.
- MRI of the Thoracic Spine … 25-45 minute scan time.
- MRI of the Chest … 25-45 minute scan time.
- MRI of the Abdomen … 25-45 minute scan time.

MRI - multi coil (4xaccel)

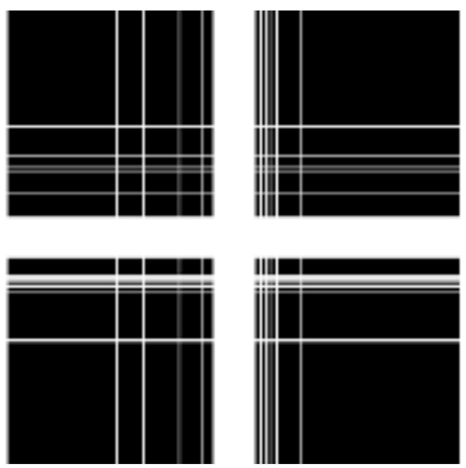


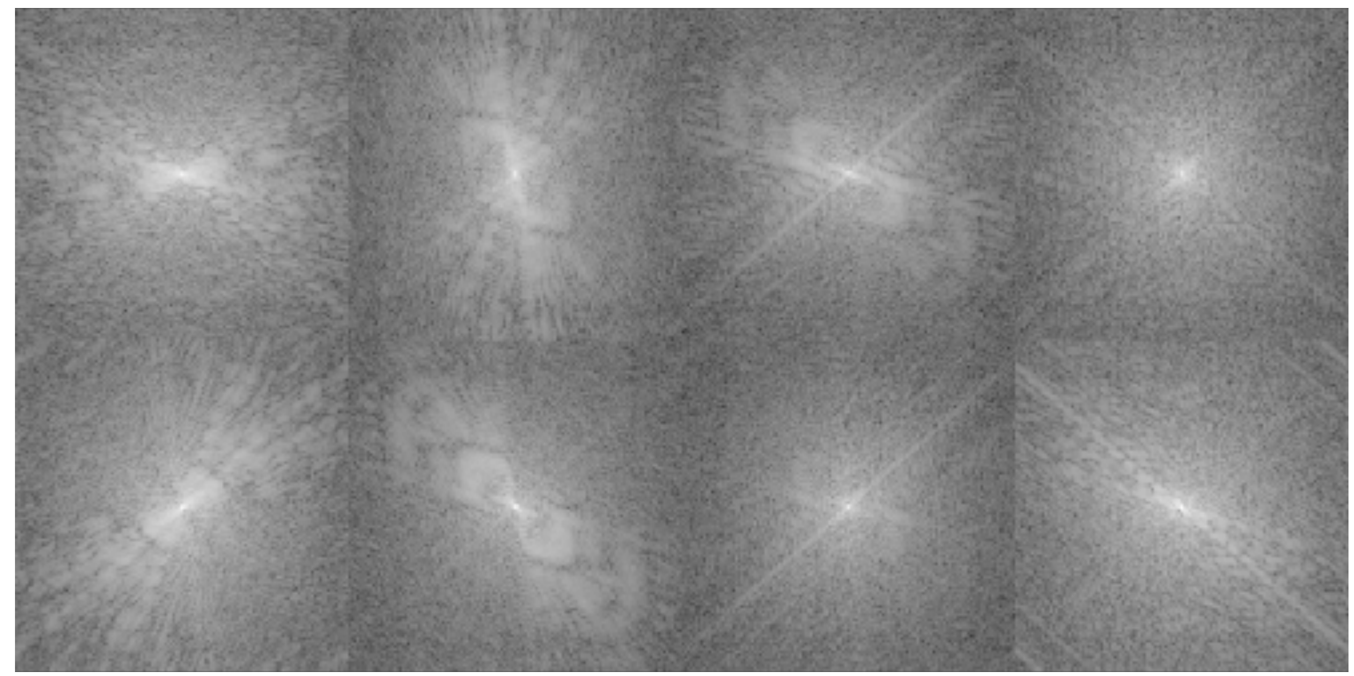


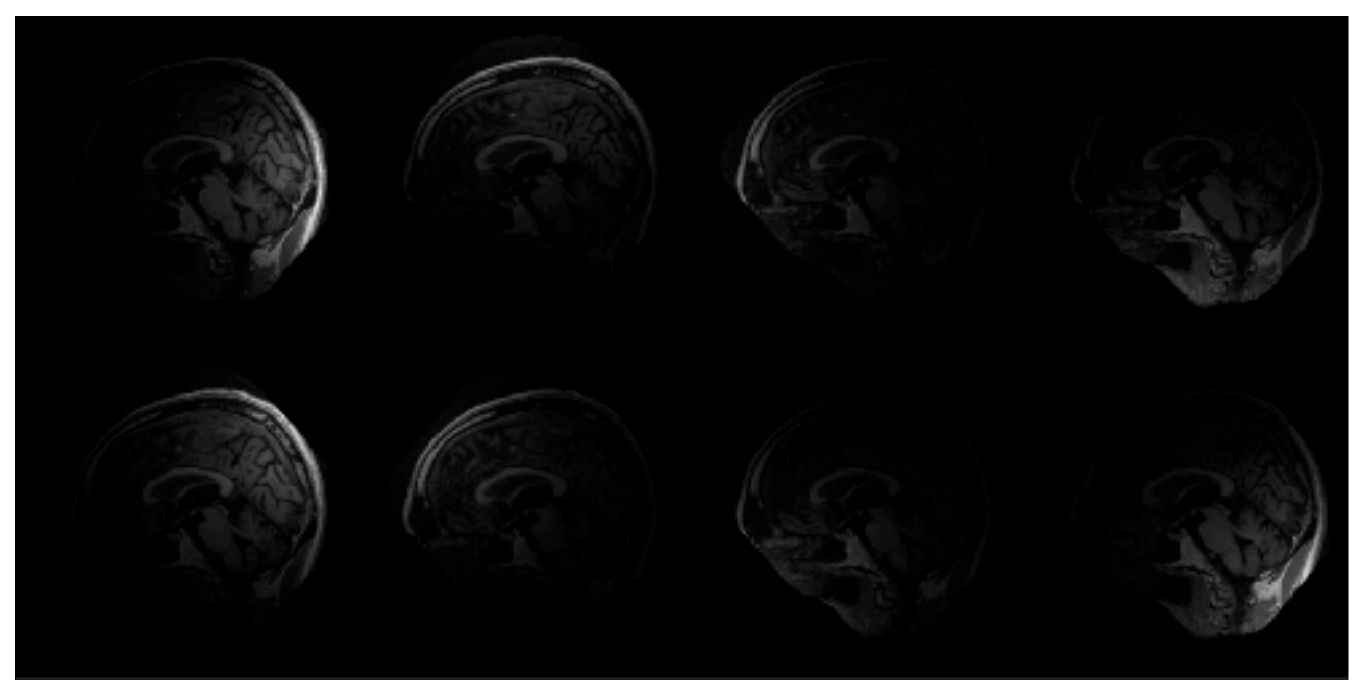


MRI - multi coil (4xaccel)

VD

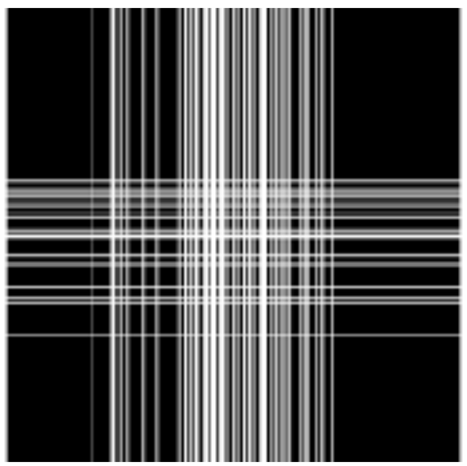


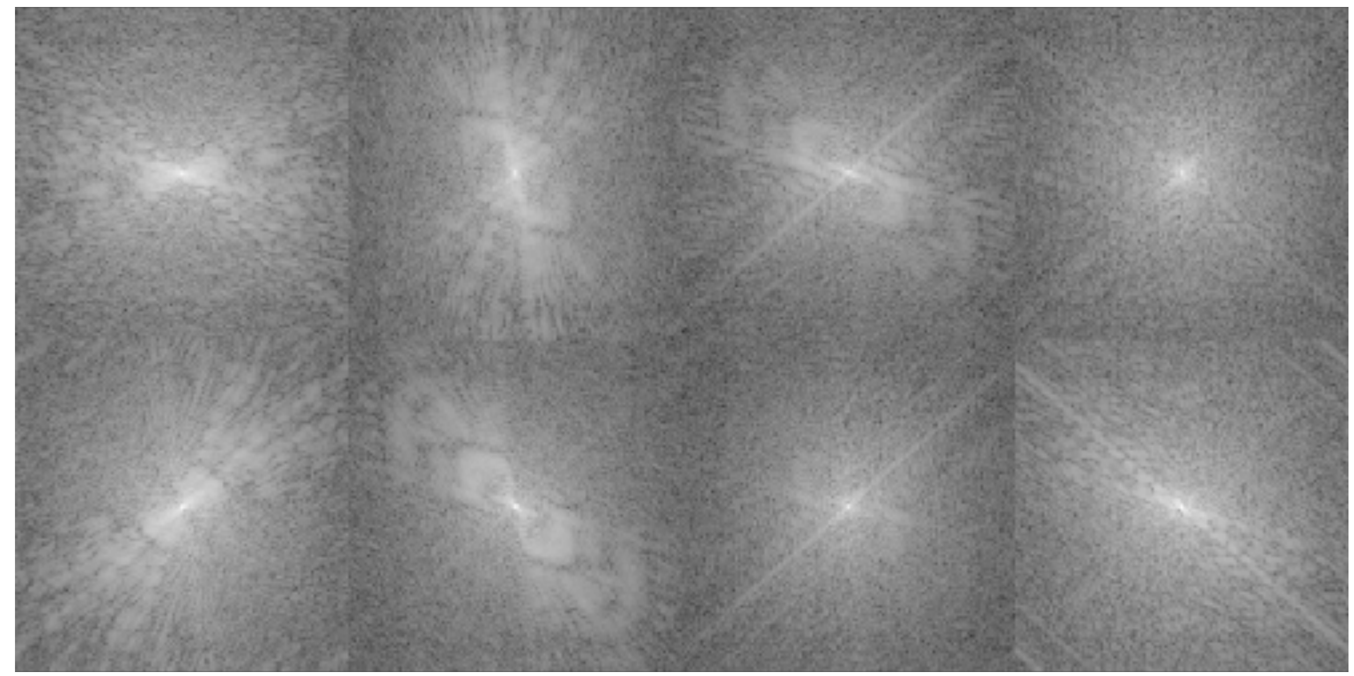




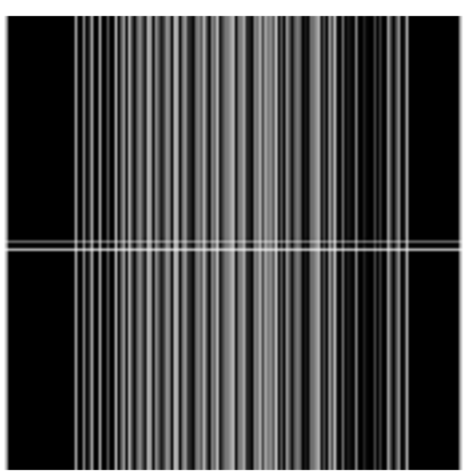
MRI - multi coil (4xaccel)

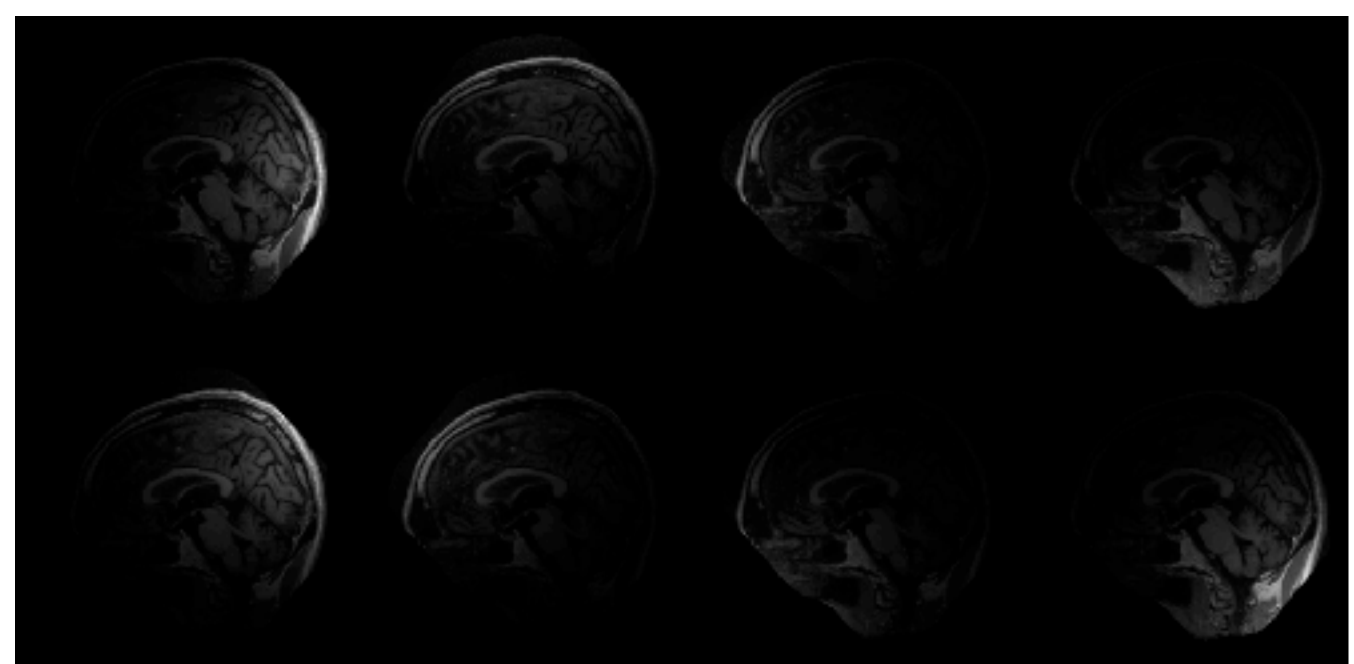
VD





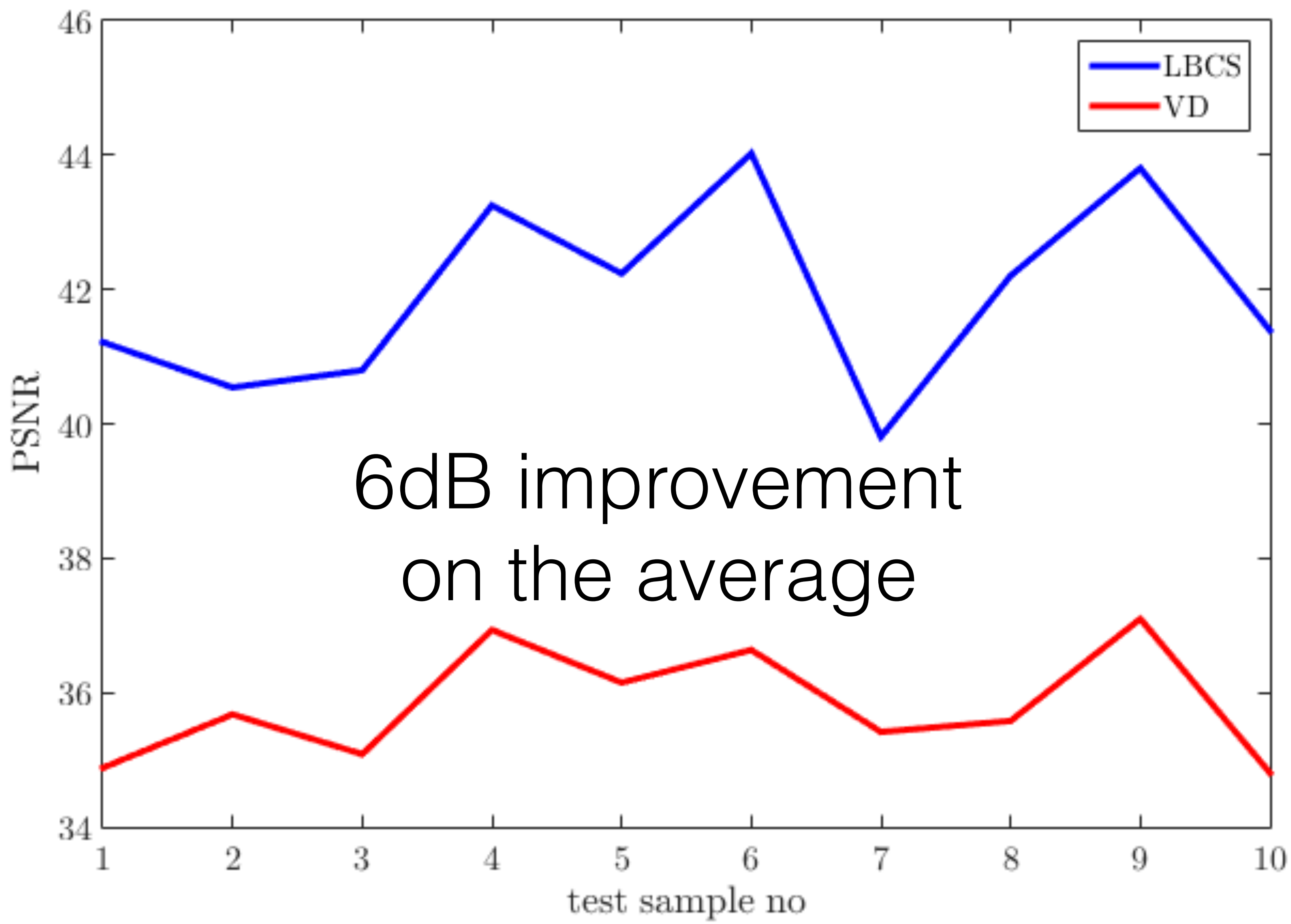
LBCS





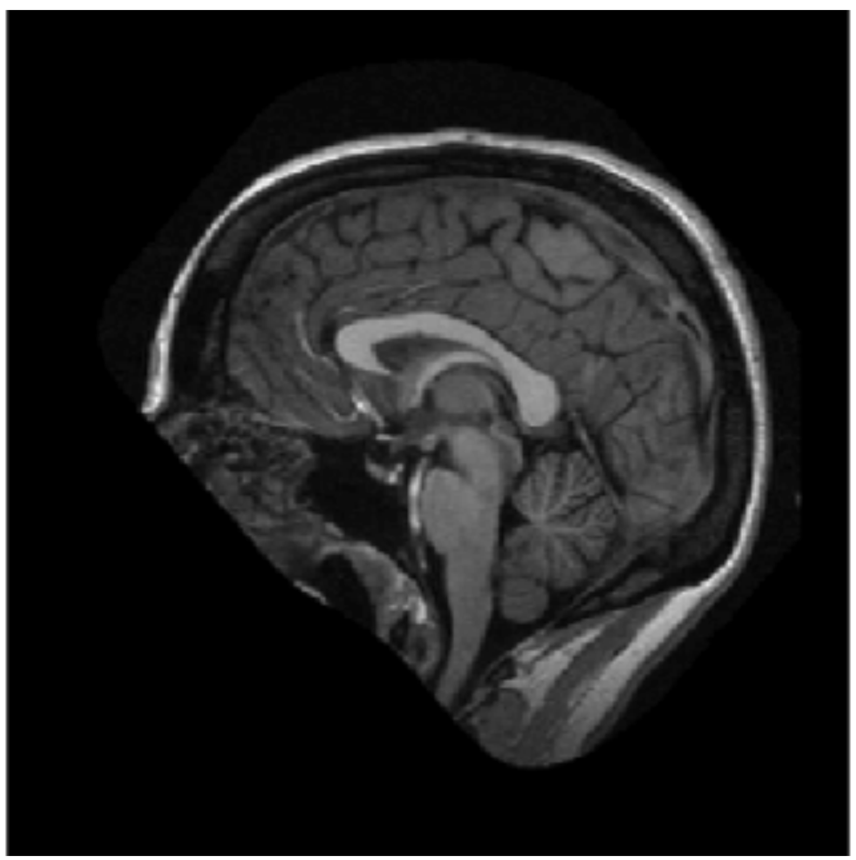
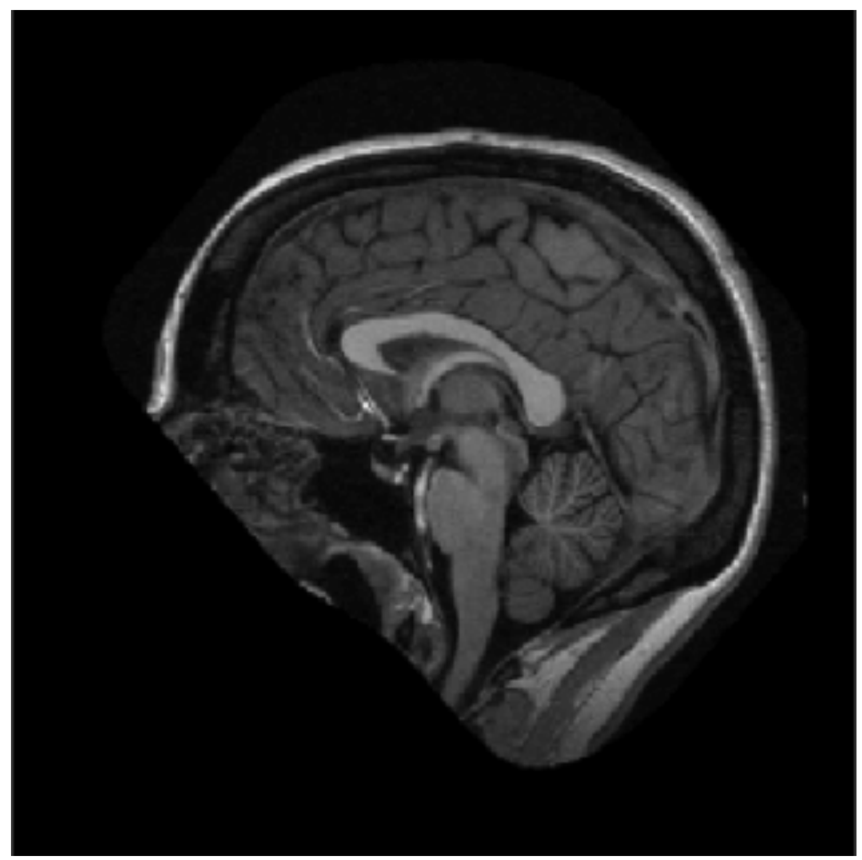
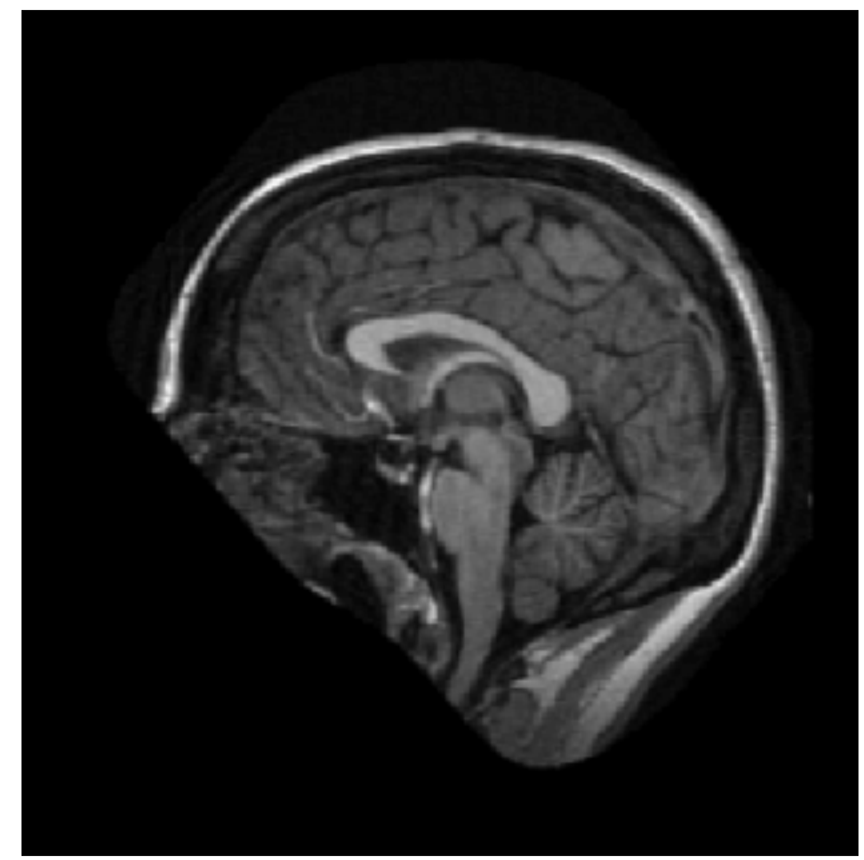
Decoder: BP with shearlets

MRI - multi coil (4xaccel)



MRI - multi coil (4xaccel)

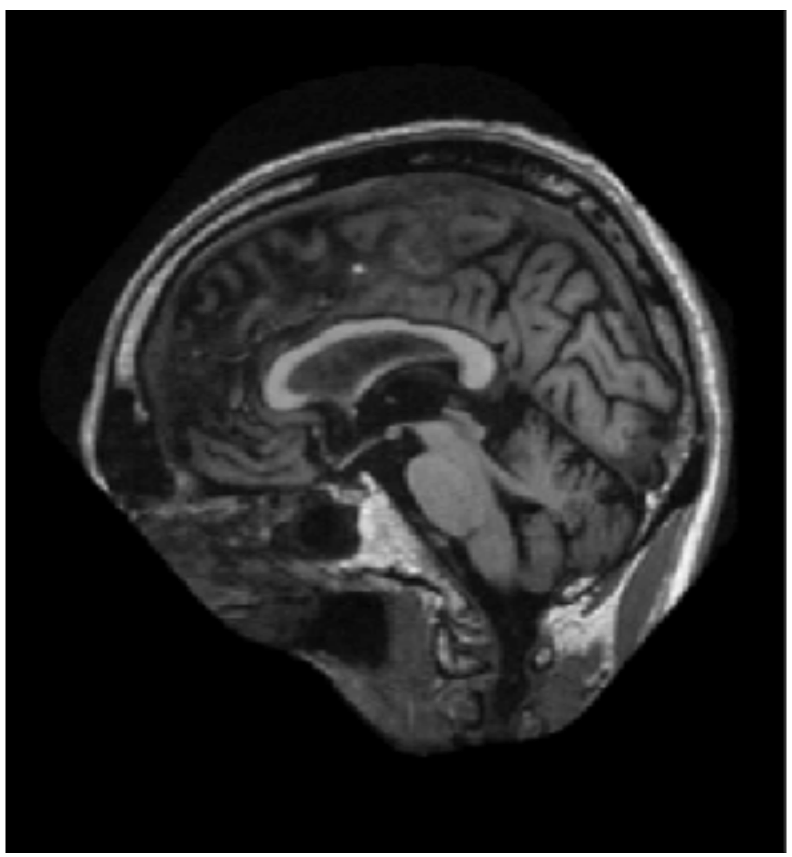
LBCS Patient #22 VD

40.80dB 35.08dB

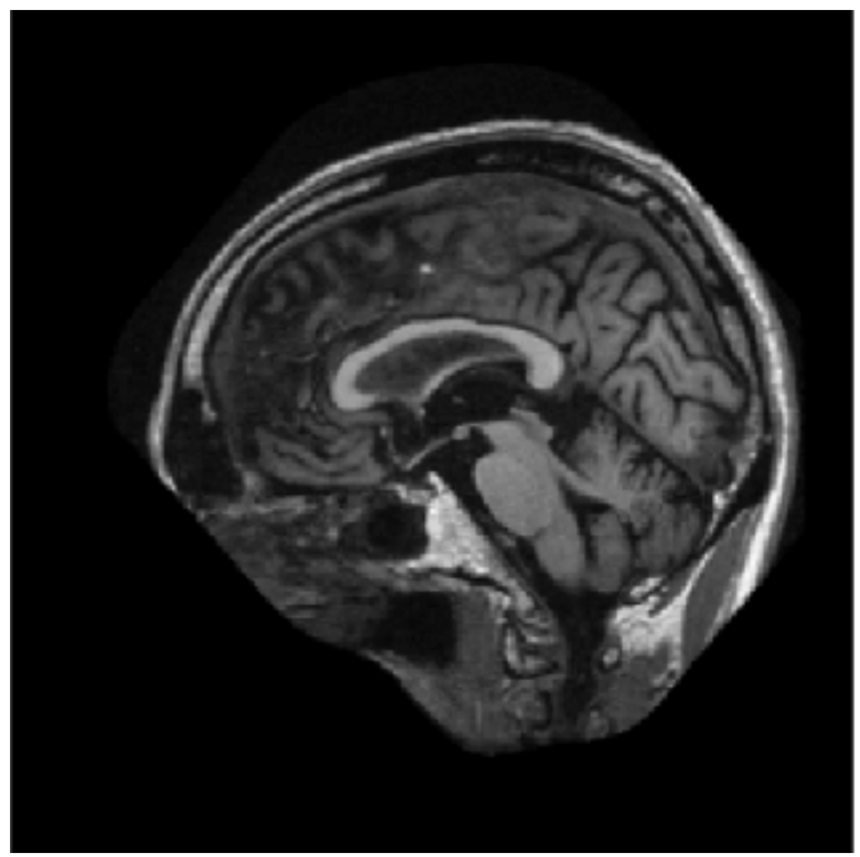
MRI - multi coil (4xaccel)

LBCS

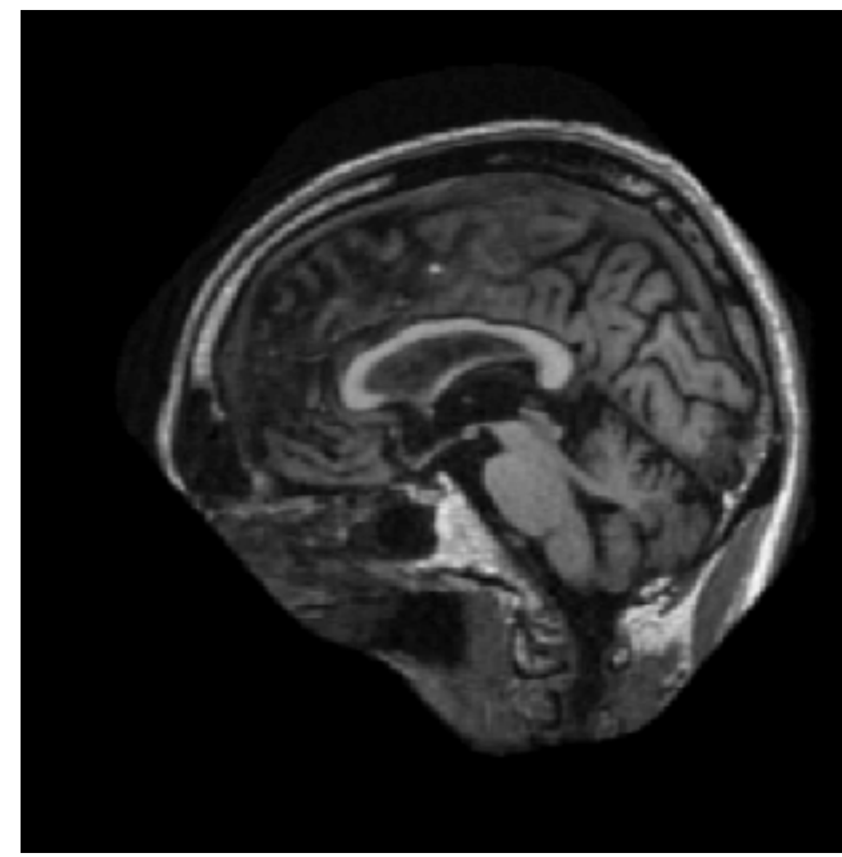


41.39dB

Patient #29



VD



34.80dB

Learning-based CS

Middle-out compression unleashed
with machine learning

http://lions.epfl.ch/publications