SKA machine learning perspectives for imaging, processing and analysis

Slava Voloshynovskiy
Stochastic Information Processing Group
University of Geneva
Switzerland

with contribution of:
D. Kostadinov, S. Ferdowsi, M. Diephuis, O. Taran and T. Holotyak

UNIVERSITÉ DE GENÈVE

SIP stochastic information processing

Outline

- Machine learning challenges in SKA
- Challenge 1 – imaging-reconstruction
- Challenge 2 – data compression for transfer and storage
- Challenge 3 – analytics and processing of big data
- Conclusions

Remark

- **Compilation of expertise from**

UNIVERSITÉ DE GENÈVE
FACULTÉ DES SCIENCES

Computer Science department
Section of mathematics
Observatory

n|w University of Applied Sciences Northwestern Switzerland

Hes·so
Haute Ecole Spécialisée de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences Western Switzerland

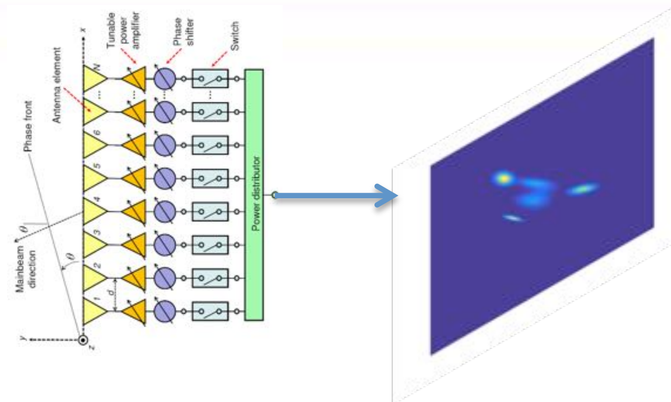
Machine learning realities and SKA

New perspectives of machine learning based image processing due to:

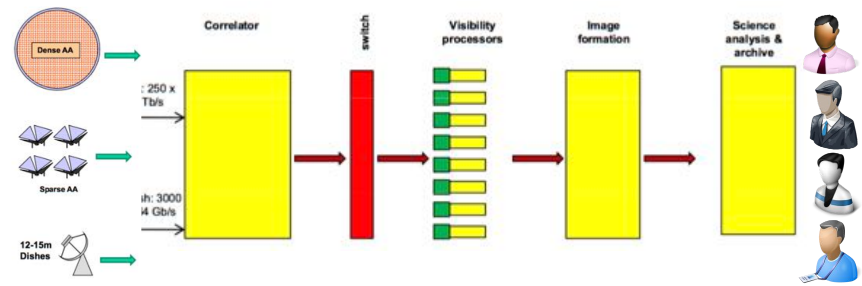
- large amount of collected observations (training data)
- new powerful computational facilities
- modern phased antenna arrays
- optimisation algorithms

Main SKA challenges

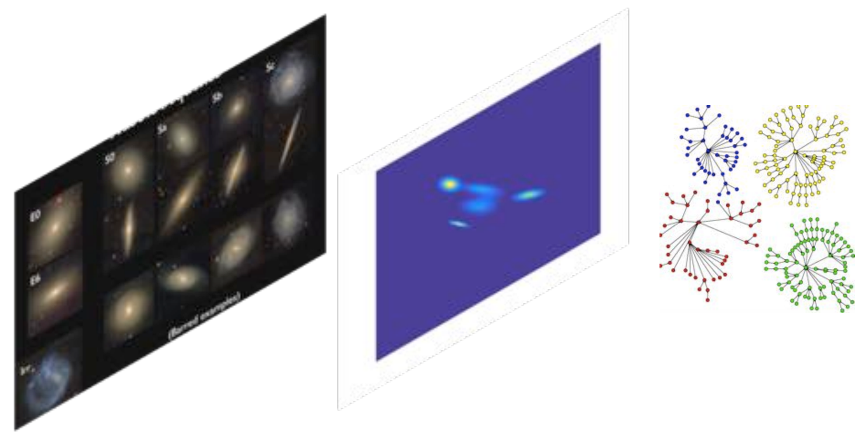
- **Challenge 1: Imaging-reconstruction**
  - Huge amount of computation for pair-wise correlations, calibration, reconstruction



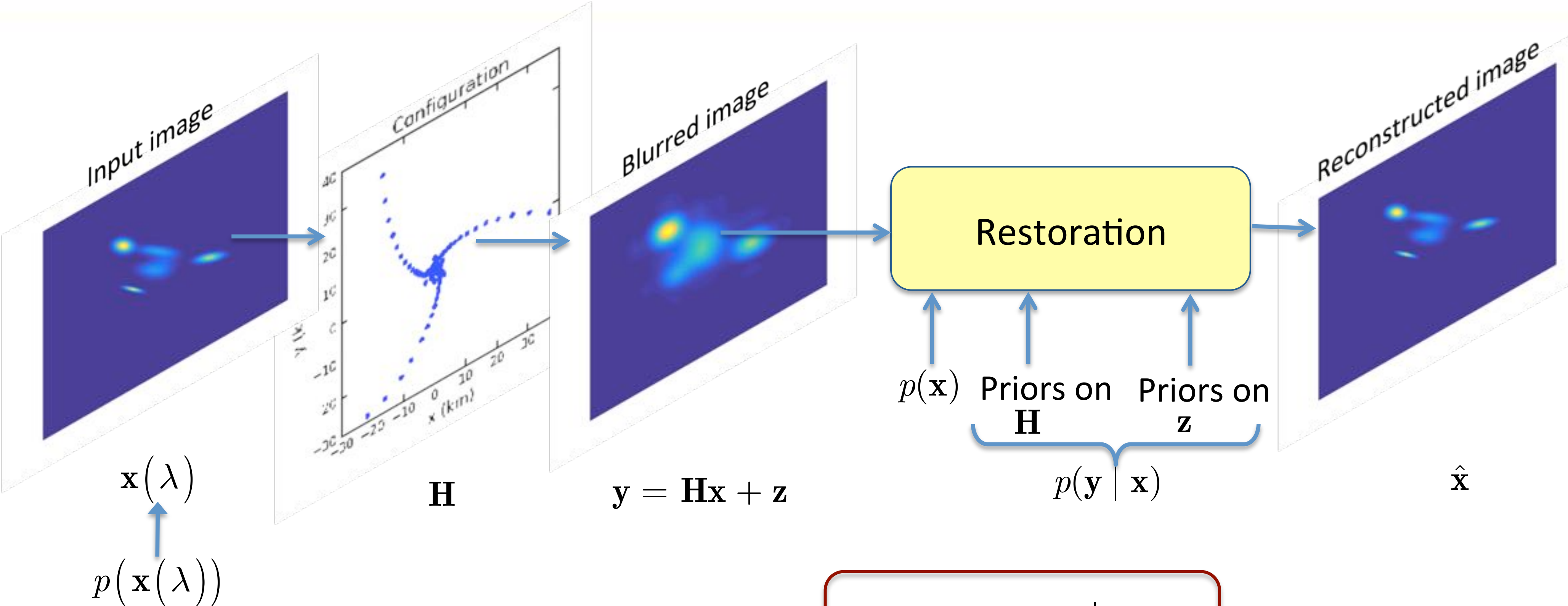
- **Challenge 2: Data transfer and storage**
  - Data transfer from correlators to reconstruction servers, data centers, SDP and end users



- **Challenge 3: Analytics**
  - Automatic processing of produced data (recognition, mining, search, tracking,…)



Challenge 1: Imaging – generic approach



- MAP $\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$
- Main issue:
  How to model $p(\mathbf{x})$ to obtain **accurate**, **tractable** and **low-complexity** solution?

Challenge 1: Imaging – "hand-crafted" approach

Traditional approaches to definition of $p(\mathbf{x})$

Statistical/deterministic approaches

Direct domain

Smoothness of solution, local correlations…..

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{a}} \left\| \mathbf{y} - \mathbf{H}\mathbf{x} \right\|_2^2 + \lambda\Omega(\mathbf{x})$$

$$\Omega(\mathbf{x}) = -\ln p(\mathbf{x})$$

Transform domain

(decorrelation, energy compaction, directivity, …)

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \left\| \mathbf{y} - \mathbf{H}\mathbf{D}\mathbf{a} \right\|_2^2 + \lambda\Omega(\mathbf{a})$$

- $\hat{\mathbf{x}} = \mathbf{D}\hat{\mathbf{a}}$
  fixed, signal independent (DCT, DWT….)
- $\Omega(\mathbf{a}) = -\ln p(\mathbf{a})$
  i.i.d. GGD
  i.i.d. Student
  i.i.d. Mixture of Gaussians
  etc

Sparsity-based approach

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \left\| \mathbf{y} - \mathbf{H}\Psi\mathbf{a} \right\|_2^2 + \lambda\Omega(\mathbf{a}) \quad \Rightarrow \quad \hat{\mathbf{x}} = \Psi\hat{\mathbf{a}}$$

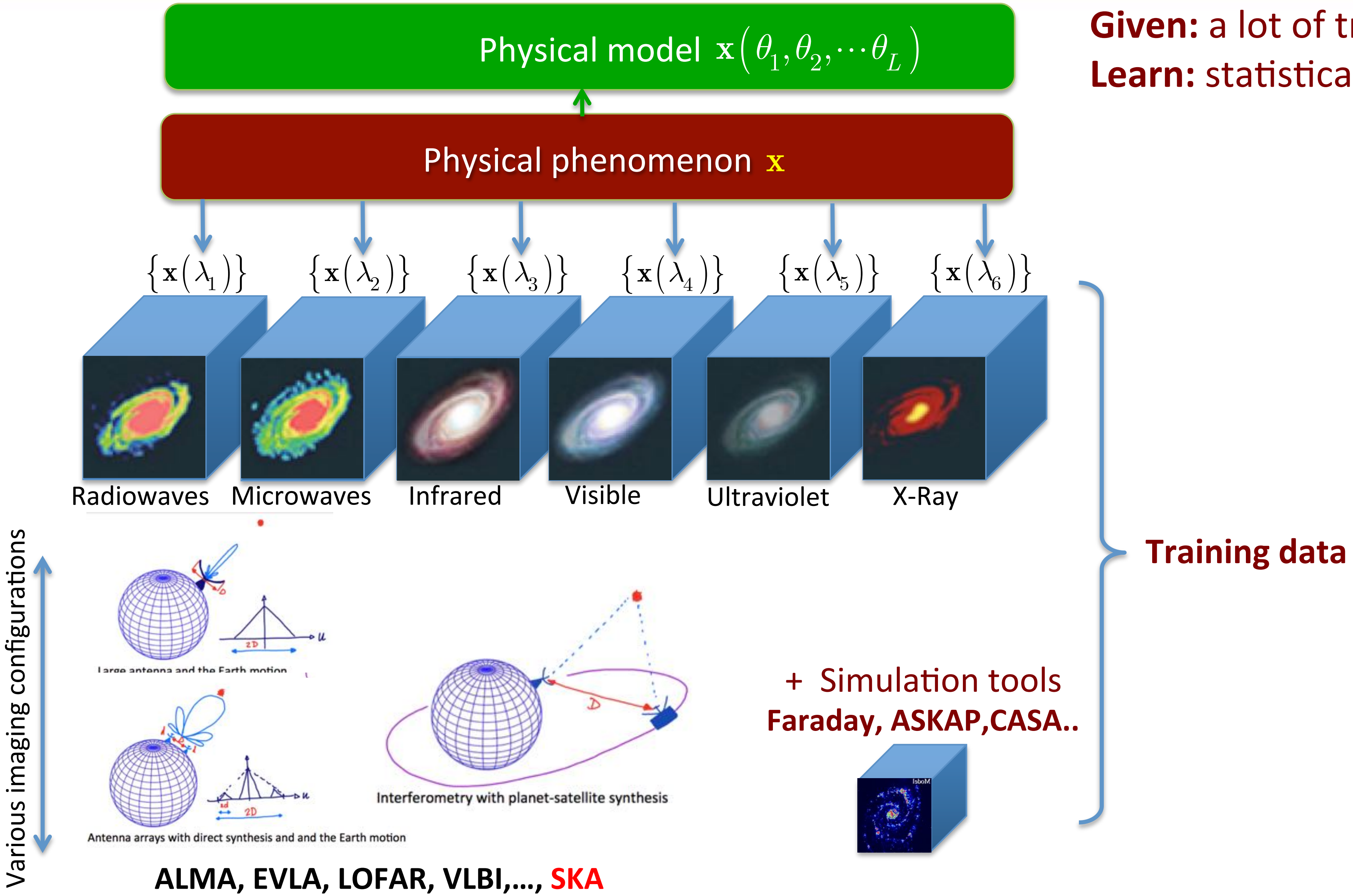
Overcomplete and can be learned

$\Omega(\mathbf{a}) = \left\| \mathbf{a} \right\|_0$

- Nonconvex and NP-hard problem: relaxatiotion/greedy approaches $\Omega(\mathbf{a}) = \left\| \mathbf{a} \right\|_1$
- Complex optimization tools:
  - proximal algorithms
  - prime-dual methods
  - augmented Langrangian ……leading to parallel and distributed solutions

Challenge 1: Imaging – "machine learning" approach

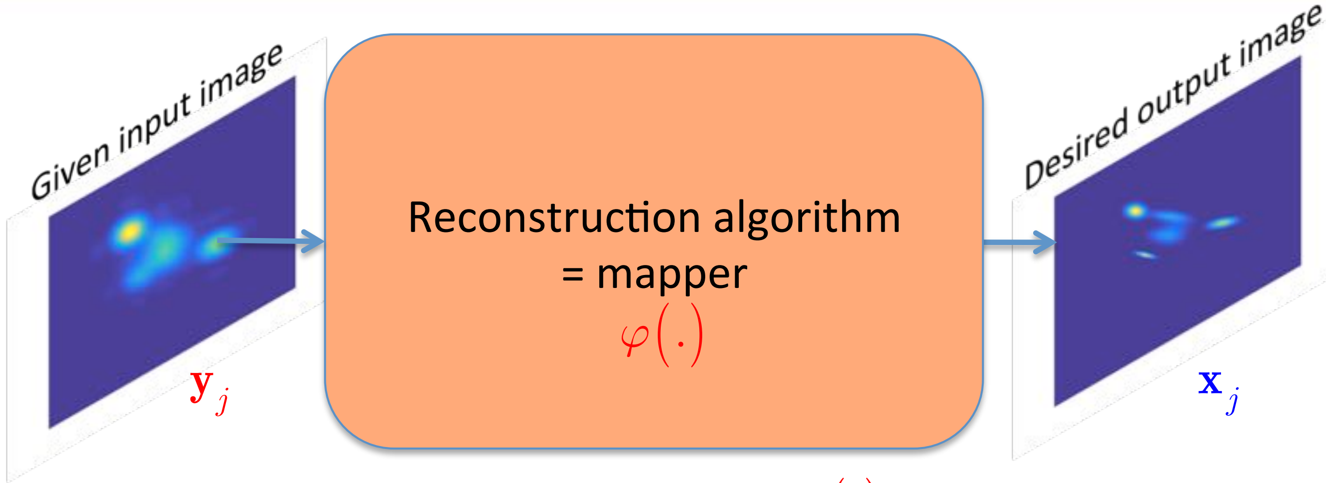
Given: a lot of training data
Learn: statistical model $p(\mathbf{x})$



Challenge 1: Imaging – as learning problem

Given: training data

$\left(\mathbf{y}_j, \mathbf{x}_j\right), 1 \leq j \leq K$

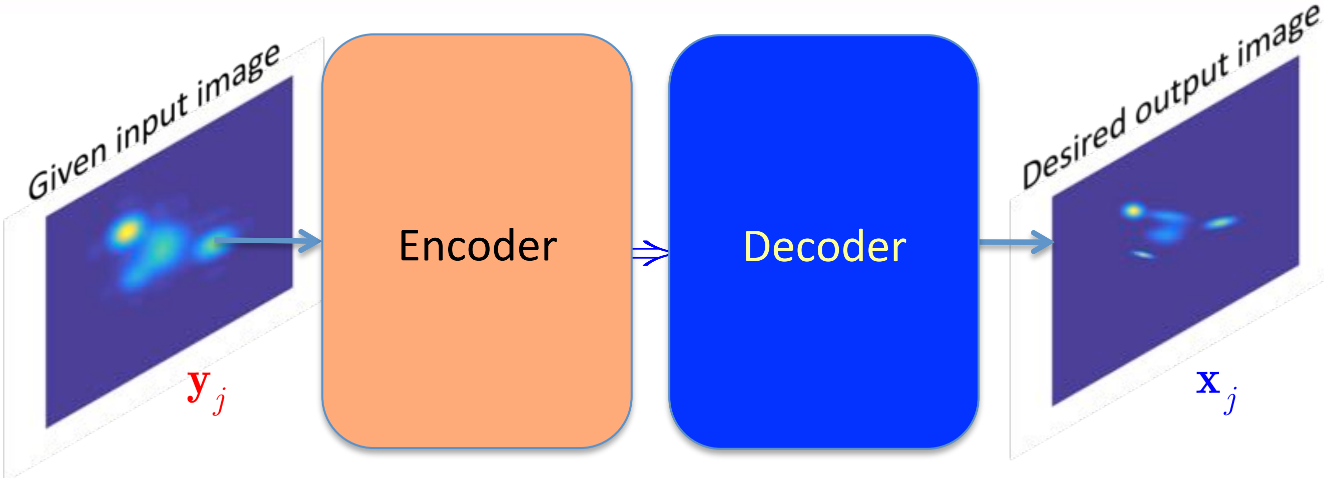


Learn a mapper $\varphi(.)$

Challenge 1: Imaging – learning as encoding-decoding

Given: training data

$\left(\mathbf{y}_j, \mathbf{x}_j\right), 1 \leq j \leq K$

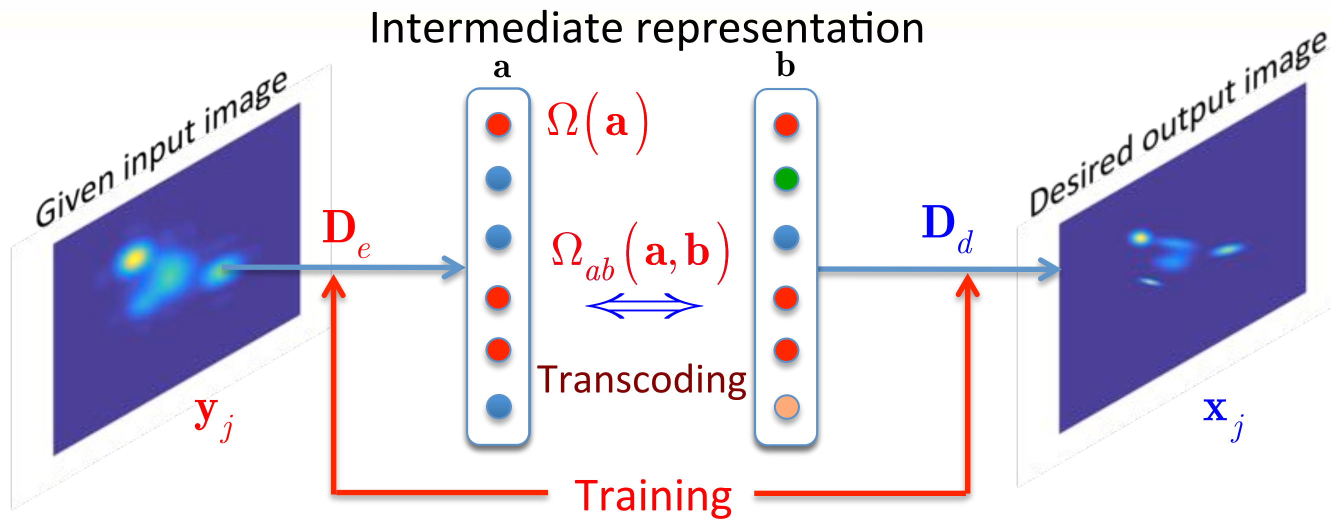


Split in two parts

Challenge 1: Imaging – learning as encoding-decoding

Given: training data

$\left(\mathbf{y}_j, \mathbf{x}_j\right), 1 \le j \le K$



Encoder-decoder training

$$\left(\hat{\mathbf{D}}_e, \hat{\mathbf{D}}_d\right) = \underset{\hat{\mathbf{D}}_e, \hat{\mathbf{D}}_d}{\arg\min} \sum_{j=1}^{K} \underbrace{\left\|\mathbf{a}_j - \mathbf{D}_e \mathbf{y}_j\right\|_2^2 + \lambda_1 \Omega_a\left(\mathbf{a}_j\right) + \lambda_2 \Omega_{ab}\left(\mathbf{a}_j, \mathbf{b}_j\right)}_{\text{Encoder}} + \lambda_3 \left\|\mathbf{x}_j - \mathbf{D}_d \mathbf{b}_j\right\|_2^2 \quad \Upsilon\left(\mathbf{D}_e, \mathbf{D}_d\right)$$

Decoder: $\lambda_2 \Omega_{ab}\left(\mathbf{a}_j, \mathbf{b}_j\right) + \lambda_3 \left\|\mathbf{x}_j - \mathbf{D}_d \mathbf{b}_j\right\|_2^2$

Projection problem

$$\hat{\mathbf{a}} = \underset{\mathbf{a}}{\arg\min} \left\|\mathbf{a} - \mathbf{D}_e \mathbf{y}\right\|_2^2 + \gamma_1 \Omega_a\left(\mathbf{a}\right) + \gamma_2 \Omega_{ab}\left(\mathbf{a}, \hat{\mathbf{b}}\right)$$

$$\hat{\mathbf{a}} = \psi\left(\mathbf{D}_e \mathbf{y}, \hat{\mathbf{b}}\right)$$

$\rightleftarrows$

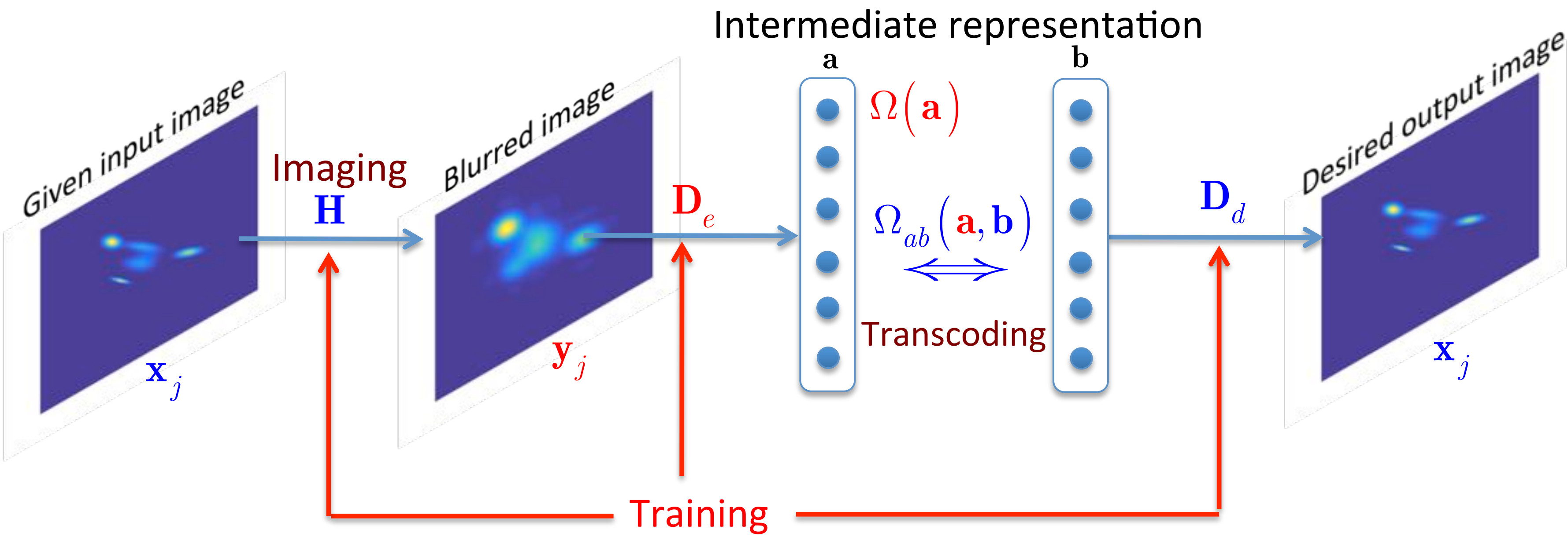
Reconstruction problem

$$\hat{\mathbf{b}} = \underset{\mathbf{a}}{\arg\min} \left\|\mathbf{y} - \mathbf{D}_d \mathbf{b}\right\|_2^2 + \lambda \Omega_{ab}\left(\hat{\mathbf{a}}, \mathbf{b}\right)$$

$$\hat{\mathbf{x}} = \mathbf{D}_d \hat{\mathbf{b}}$$

Challenge 1: Imaging – learning optimal imaging configurations

Objective: joint optimization of reconstruction and imaging (CS – random sampling)



Imaging system-encoder-decoder training

Encoder

$$\left(\hat{\mathbf{H}},\hat{\mathbf{D}}_e,\hat{\mathbf{D}}_d\right)=\arg\min_{\mathbf{H},\hat{\mathbf{D}}_e,\hat{\mathbf{D}}_d}\sum_{j=1}^{K}\left\|\mathbf{a}_j-\mathbf{D}_e\mathbf{H}\mathbf{x}_j\right\|_2^2+\lambda_1\Omega_a\left(\mathbf{a}_j\right)+\lambda_2\Omega_{ab}\left(\mathbf{a}_j,\mathbf{b}_j\right)+\lambda_3\left\|\mathbf{x}_j-\mathbf{D}_d\mathbf{b}_j\right\|_2^2$$

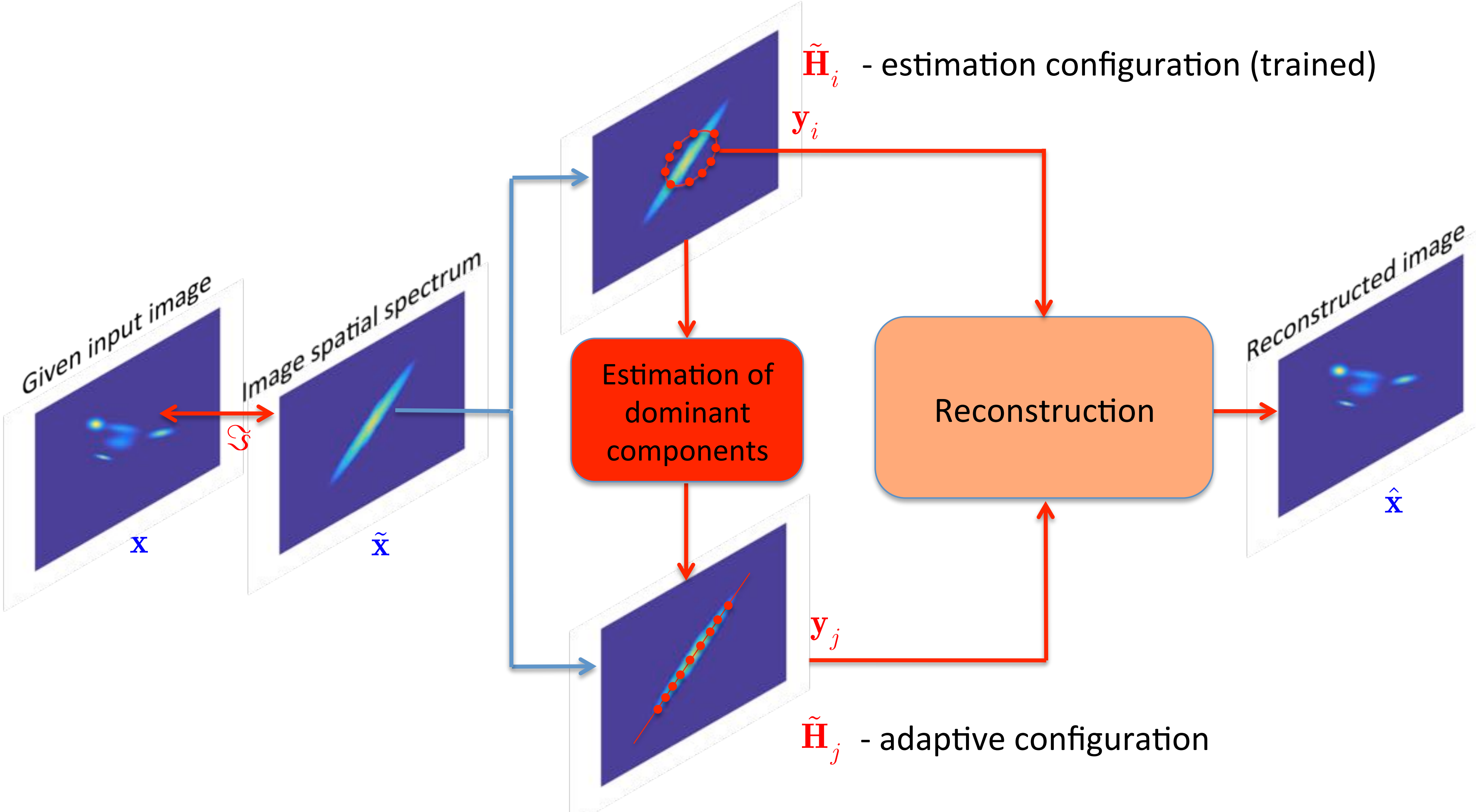
Decoder

$\Upsilon\left(\mathbf{D}_e,\mathbf{D}_d\right)$

$\Phi\left(\mathbf{H}\right)$ **constraints on number and possible geometry of synthesized arrays**

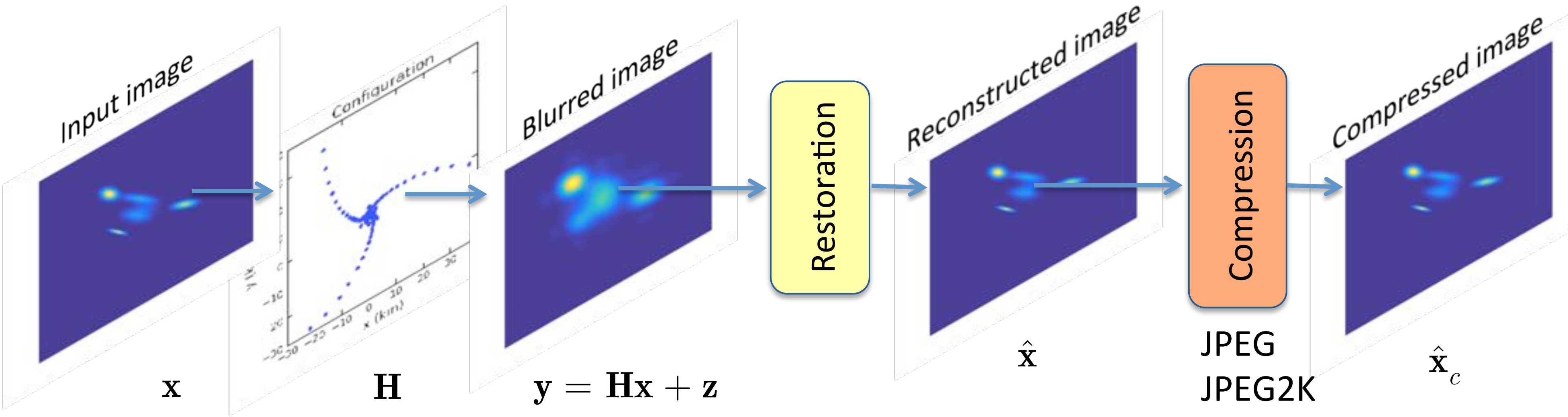
Challenge 1: Imaging – learning for "adaptive" imaging

Objective: minimize the load on correlators ⇒ adaptive "light-weight" imaging



Challenge 2: Compression for transfer, storage and distribution

- **Traditional approach to compression**



Generic and "image independent"

- **Alternative - compressive sensing $\Rightarrow$ <u>quantized</u> compressive sensing**

$$\mathbf{y} = Q\left(\mathbf{H}\mathbf{x} + \mathbf{z}\right)$$

observations are quantized (even to several bits)

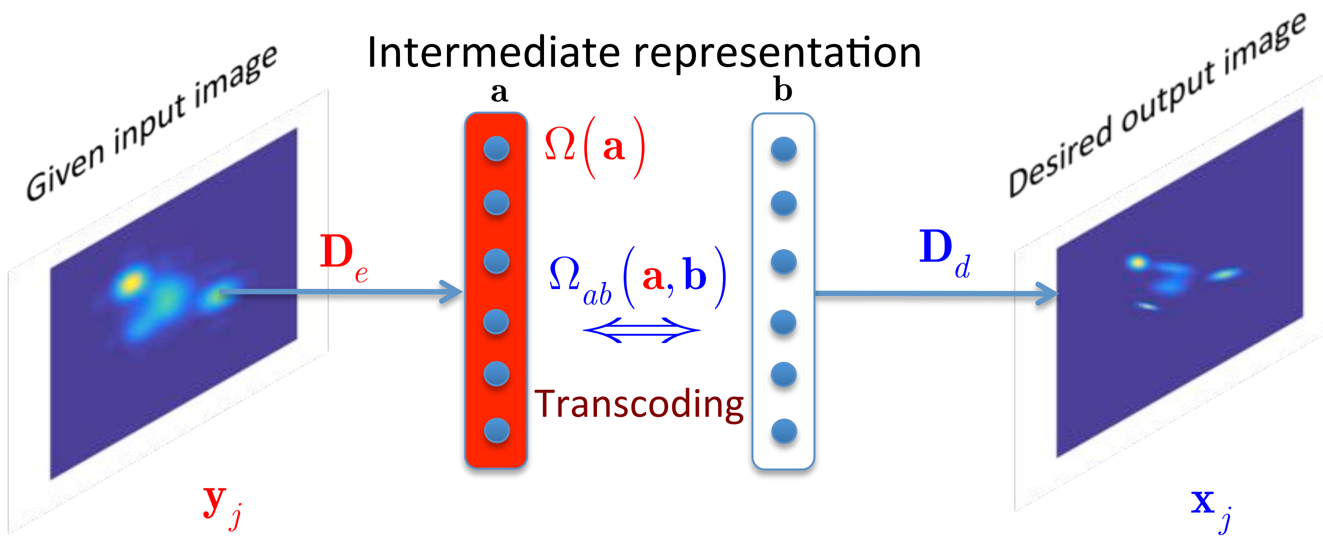
$$\hat{\mathbf{x}} = \arg\min_{\mathbf{a}} \left\| \mathbf{y} - Q\left(\mathbf{H}\mathbf{x}\right) \right\|_2^2 + \lambda\Omega\left(\mathbf{x}\right)$$

- inverse problem

Challenge 2: Compression – machine learning approach

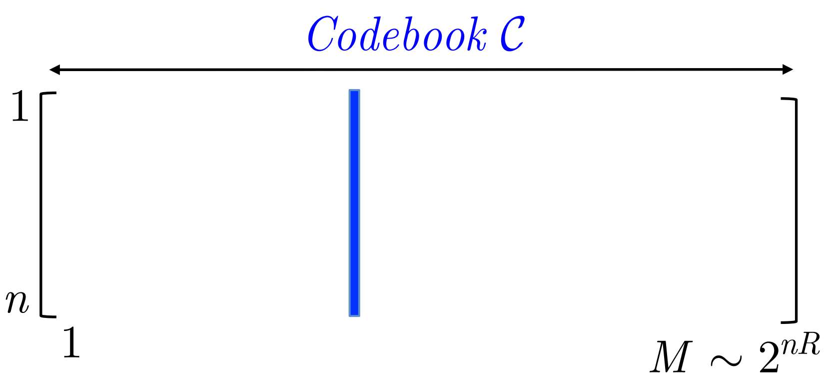
- **Option 1: Replace generic JPEG/JPEG2K codecs by special algorithms trained on RI images**
  - We suppose that the image is already reconstructed and the problem is how to deliver it to the end users

- **Option 2: Sparse code representation between the Encoder-Decoder**
  - Encoder-decoder pairs in reconstruction are trained with the entropy constrained sparse code
  - Efficient coding based on **structured** codebooks vs **random** ones

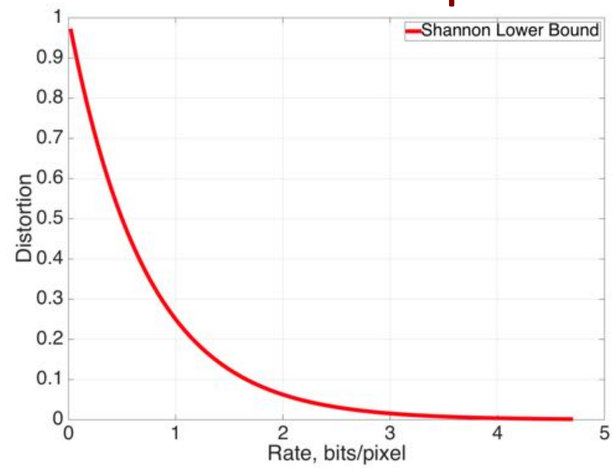


Challenge 2: Compression – information-theoretic approach

Codebook training (like Shannon R(D) theory) from $p(\mathbf{x})$



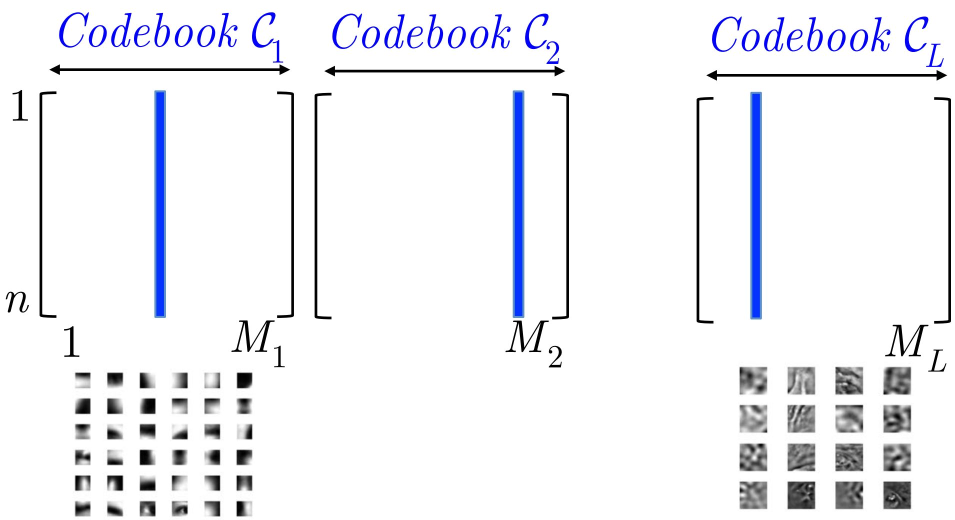
Rate-distortion plot



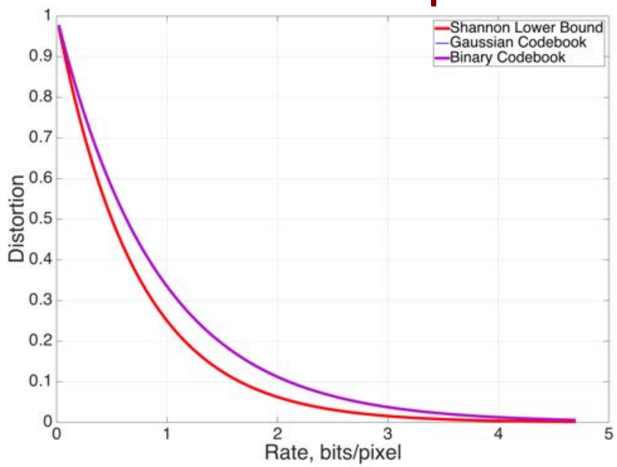
- Information-theoretic approach (limits)
- $p(\mathbf{x})$ is not known
- Codebook size is exponential in n
- Codebook is unstructured
- High compression complexity and memory

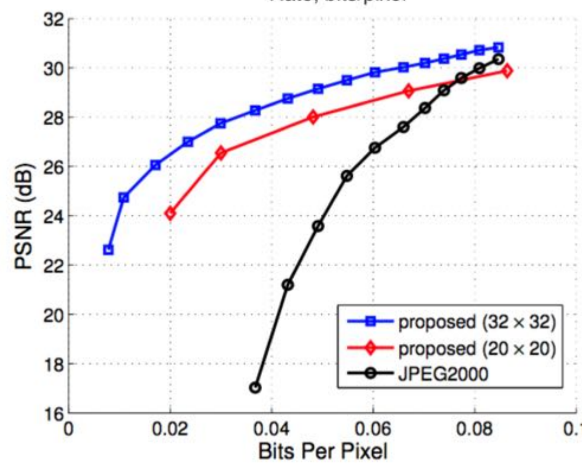
Challenge 2: Compression – machine learning approach

Structured codebooks with successive refinement



Rate-distortion plot

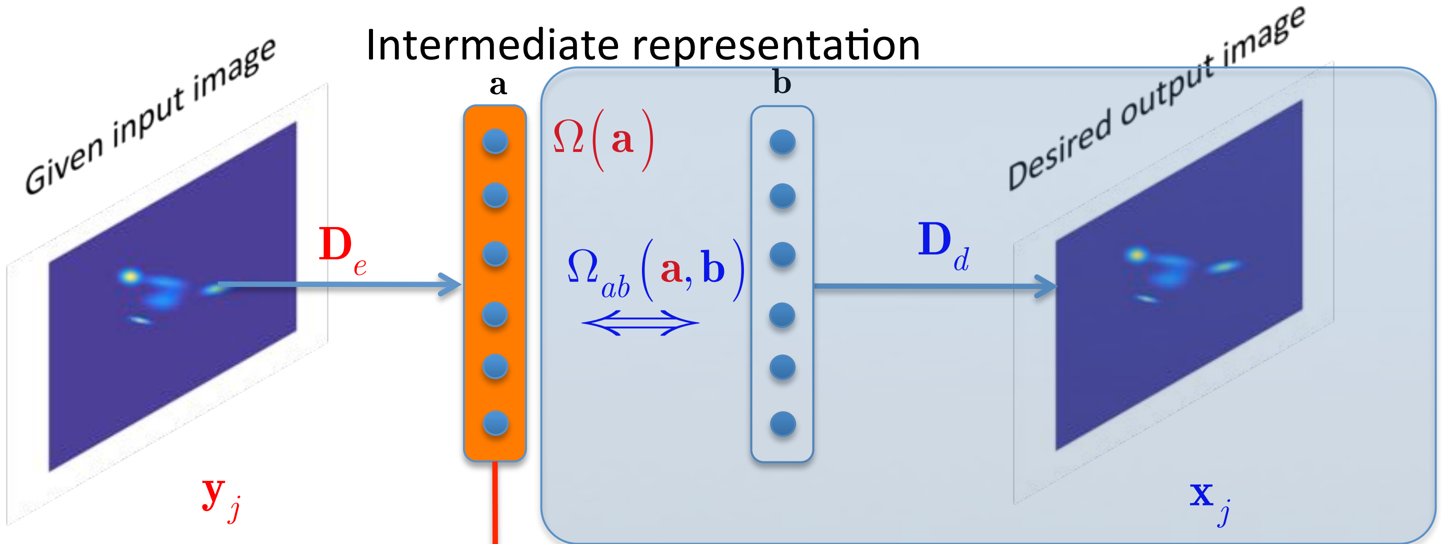


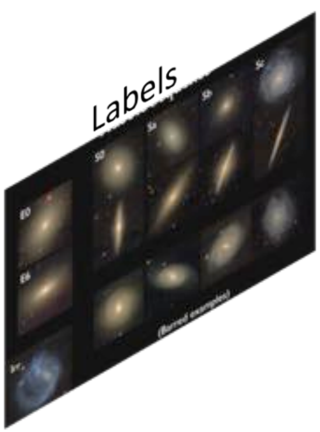


- L **trained** codebooks (structured)
- Polynomial complexity
- Approaching Shannon lower bound
- Outperforming JPEG/JPEG2K for very low bit rates

Challenge 3: Analytics – automatic processing of Big Data

Main issue: dimensionality and amount of images for automatic processing





Feature vector = compact and discriminative image representation

Supervised "hashing"
010011110010101

- Recognition
- Data mining
- Search
- Indexing

Automatic tools for Big Data analysis

Conclusions

- **SKA is a great tool for research**
- …. and not only in astronomy and astro-physics

- It is a "test" platform for many ideas in **machine learning based image processing** thanks to **big data** and **flexible imaging**

- In turns, it will lead to new alternative approaches to three SKA challenges covering:
  - imaging and reconstruction algorithms
  - compression algorithms
  - automatic processing of big data