Multidataset Independent Subspace Analysis

Rogers F. Silva, Ph.D.

Postdoctoral Fellow
The Mind Research Network

Data Scientist
Datalytic Solutions

with

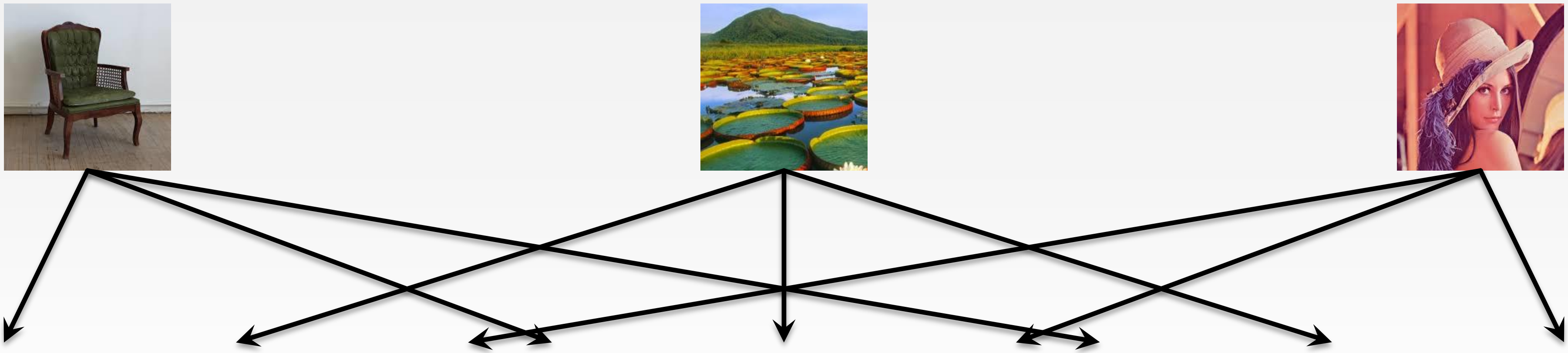
Sergey M. Plis, Ph.D. (MRN)
Tulay Adali, Ph.D. (UMBC)
Marios S. Pattichis, Ph.D. (UNM)
Vince D. Calhoun, Ph.D. (MRN/UNM)

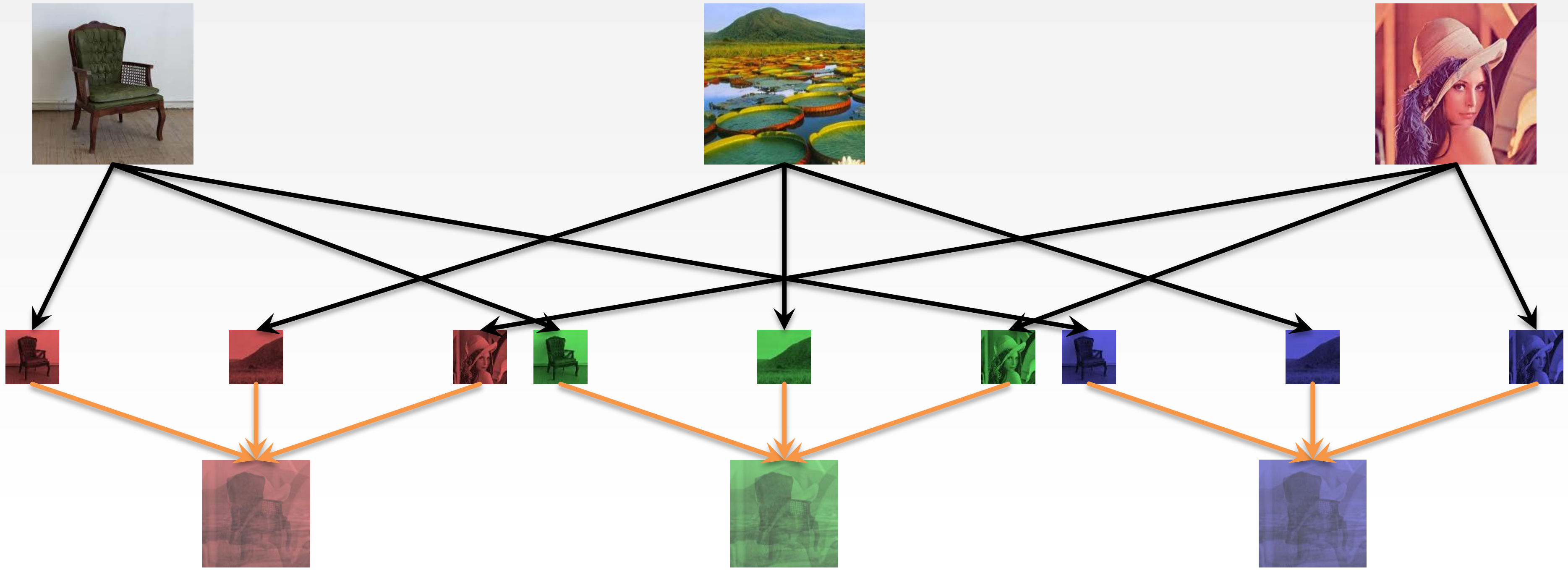
Jun/30/2017

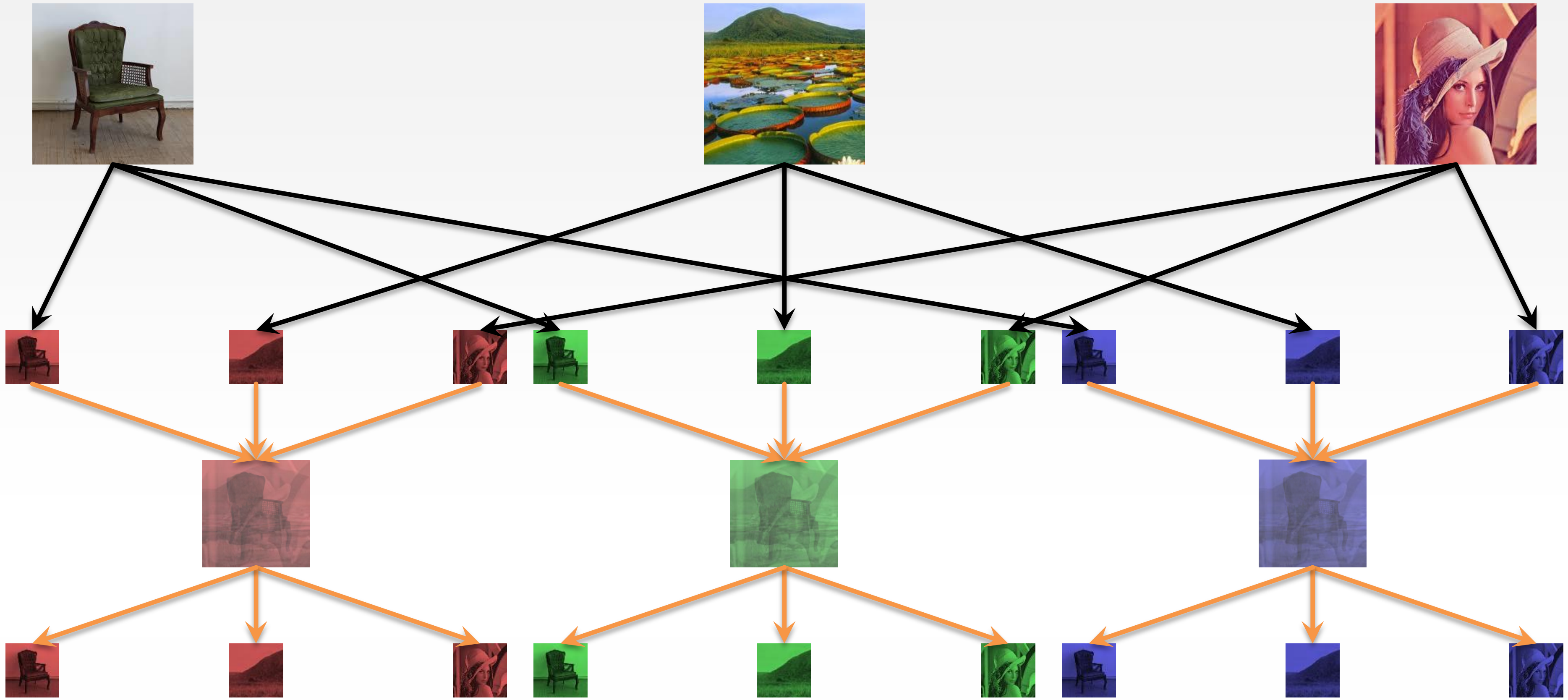
2017 Deep Learning Summer School

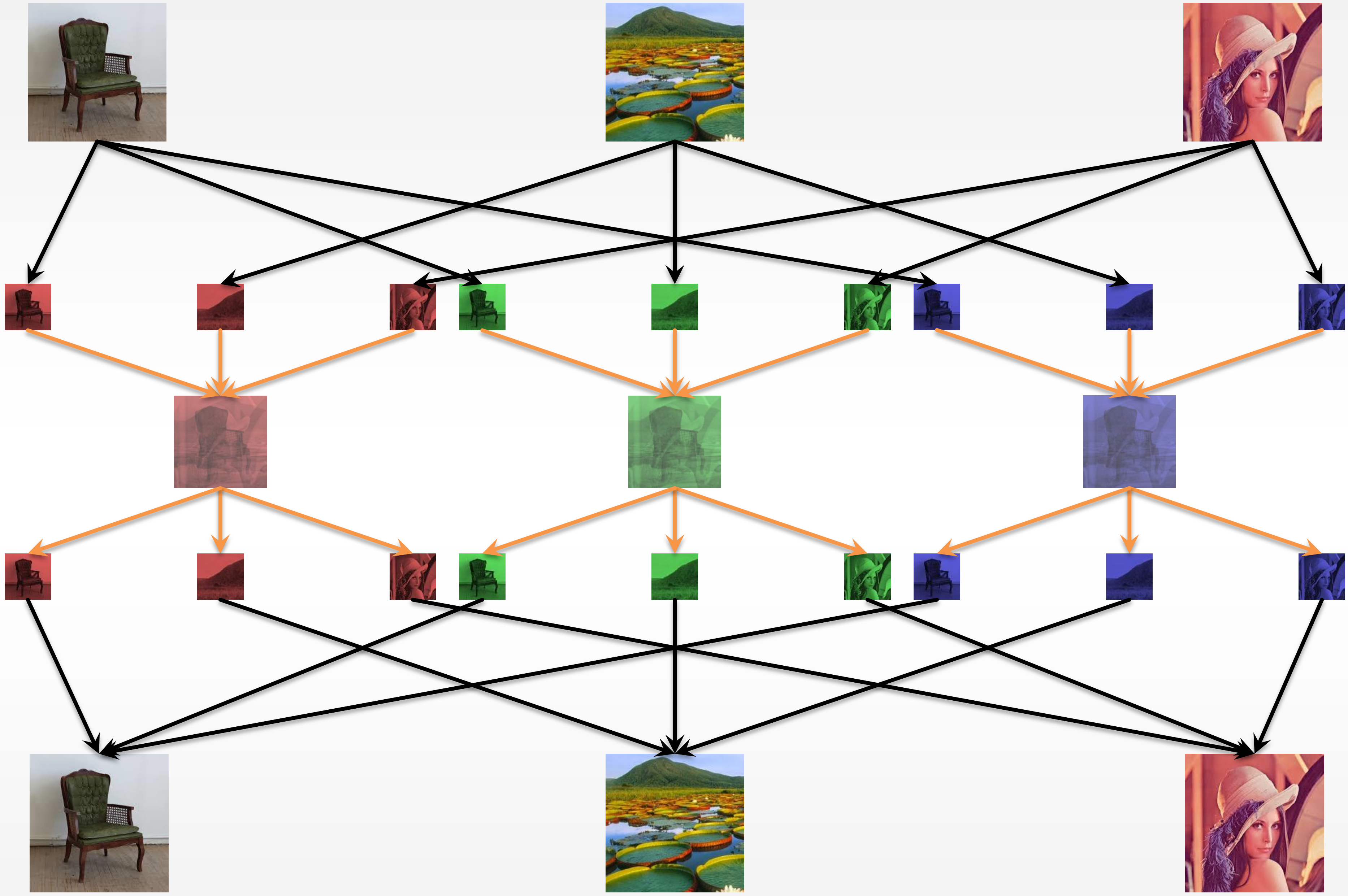
University of Montreal

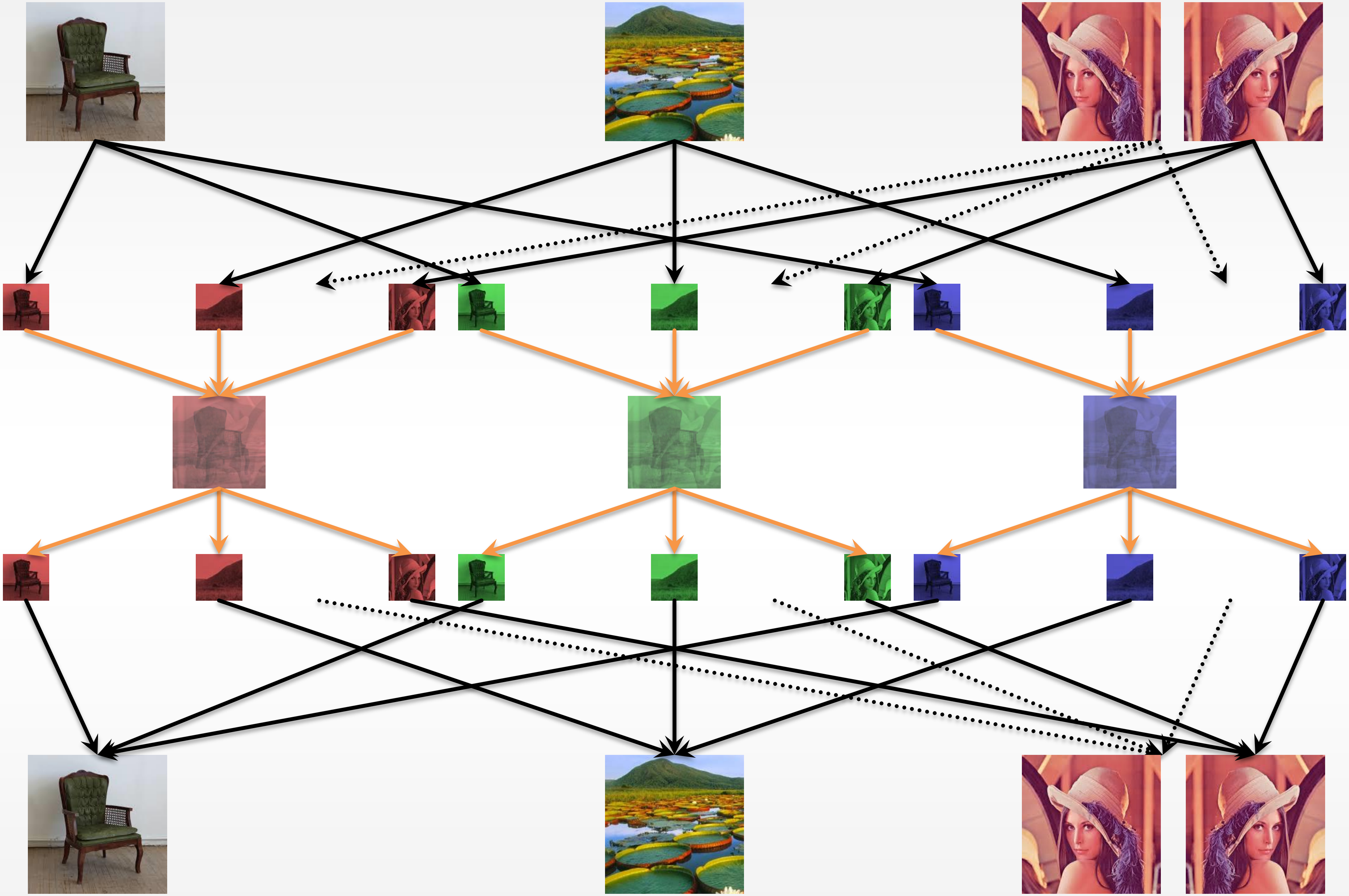












Multidataset Multidimensional Problems

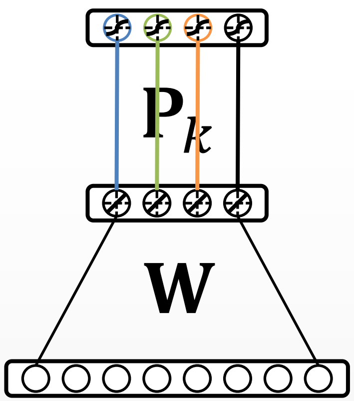
- A Hierarchy of Blind Source Separation Models [*Silva et al.*, *2016*]

Multidataset Multidimensional Problems

- A Hierarchy of Blind Source Separation Models [Silva et al., 2016]

$$J(\mathbf{W}) = KL\langle p(\mathbf{y}), \prod_k^K p_k(\mathbf{y}_k)\rangle,$$

$$= -\sum_i^{C=K} \log|\sigma_i| - \sum_k^K \mathbb{E}[\mathrm{h}_k(\mathrm{y}_k)],$$



$$\mathrm{h}_k(\mathrm{y}_k) = -\sqrt{\frac{y_k^2}{\sigma_k^2}} - \frac{1}{2}\log\det\sigma_k^2$$

$$\mathbf{y} = \mathbf{W}\mathbf{x}$$

$\mathbf{x}$

SDU

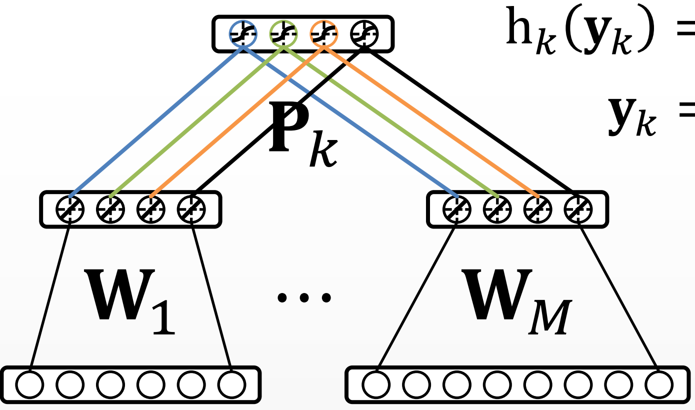
Single Dataset (**SD**)

Multidataset Multidimensional Problems

- A Hierarchy of Blind Source Separation Models *[Silva et al., 2016]*

$$J(\mathbf{W}) = KL\langle p(\mathbf{y}), \prod_k^K p_k(\mathbf{y}_k)\rangle,$$

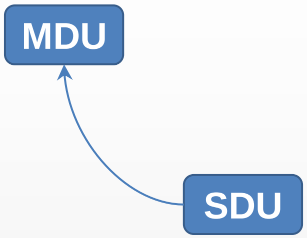
$$\quad = -\sum_m^M \sum_i^{C_m=K} \log|\sigma_{mi}| - \sum_k^K \mathbb{E}[\mathrm{h}_k(\mathbf{y}_k)],$$



$$\mathrm{h}_k(\mathbf{y}_k) = -\sqrt{\mathbf{y}_k^\top \mathbf{\Sigma}_k^{\mathrm{y}^{-1}} \mathbf{y}_k} - \frac{1}{2}\log\det \mathbf{\Sigma}_k^{\mathbf{y}}$$

$$\mathbf{y}_k = \mathbf{P}_k \mathbf{W} \mathbf{x}$$

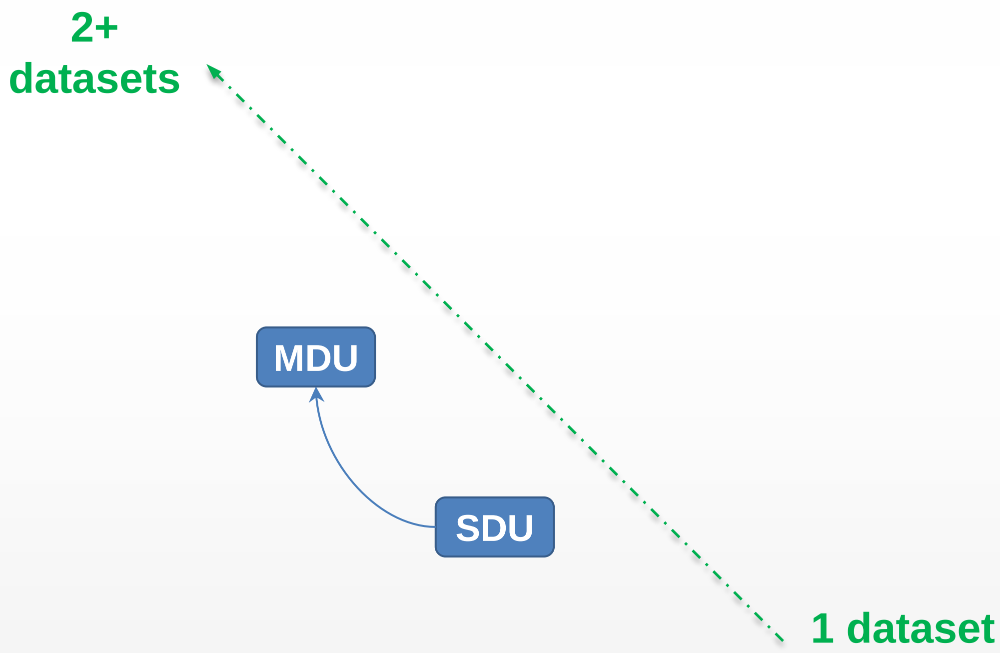
$$\mathbf{y}_m = \mathbf{W}_m \mathbf{x}_m$$



Single Dataset (**SD**) → Multidataset (**MD**)

Multidataset Multidimensional Problems

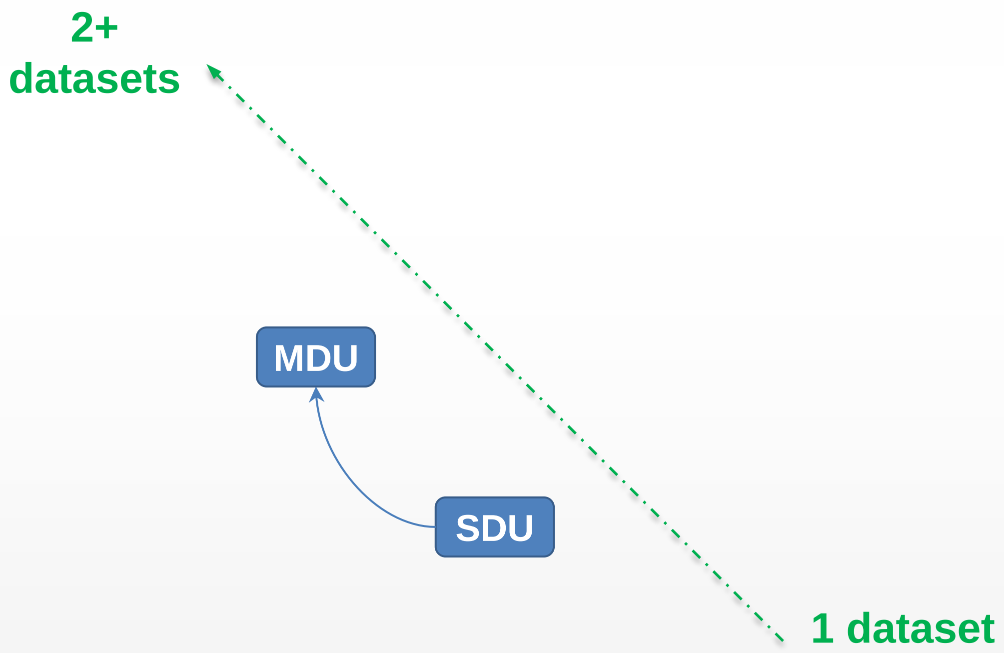
- A Hierarchy of Blind Source Separation Models *[Silva et al., 2016]*



Single Dataset (**SD**) → Multidataset (**MD**)

Multidataset Multidimensional Problems

- A Hierarchy of Blind Source Separation Models *[Silva et al., 2016]*

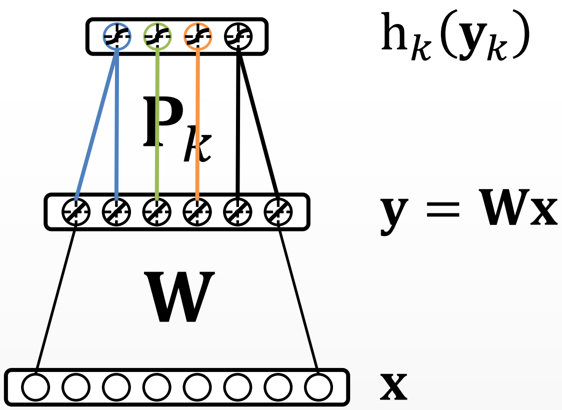


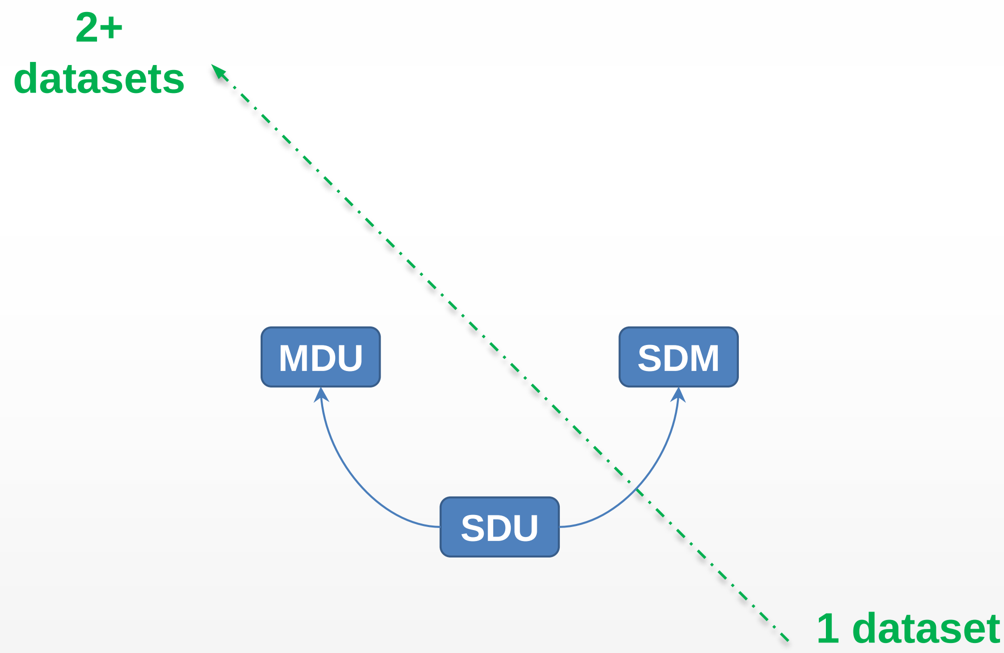
Single Dataset (**SD**) → Multidataset (**MD**)
Unidimensional (**U**)

Multidataset Multidimensional Problems

- A Hierarchy of Blind Source Separation Models *[Silva et al., 2016]*

$$J(\mathbf{W}) = KL\langle p(\mathbf{y}), \prod_k^K p_k(\mathbf{y}_k)\rangle,$$
$$= -\sum_i^C \log|\sigma_i| - \sum_k^K \mathbb{E}[\mathrm{h}_k(\mathbf{y}_k)],$$

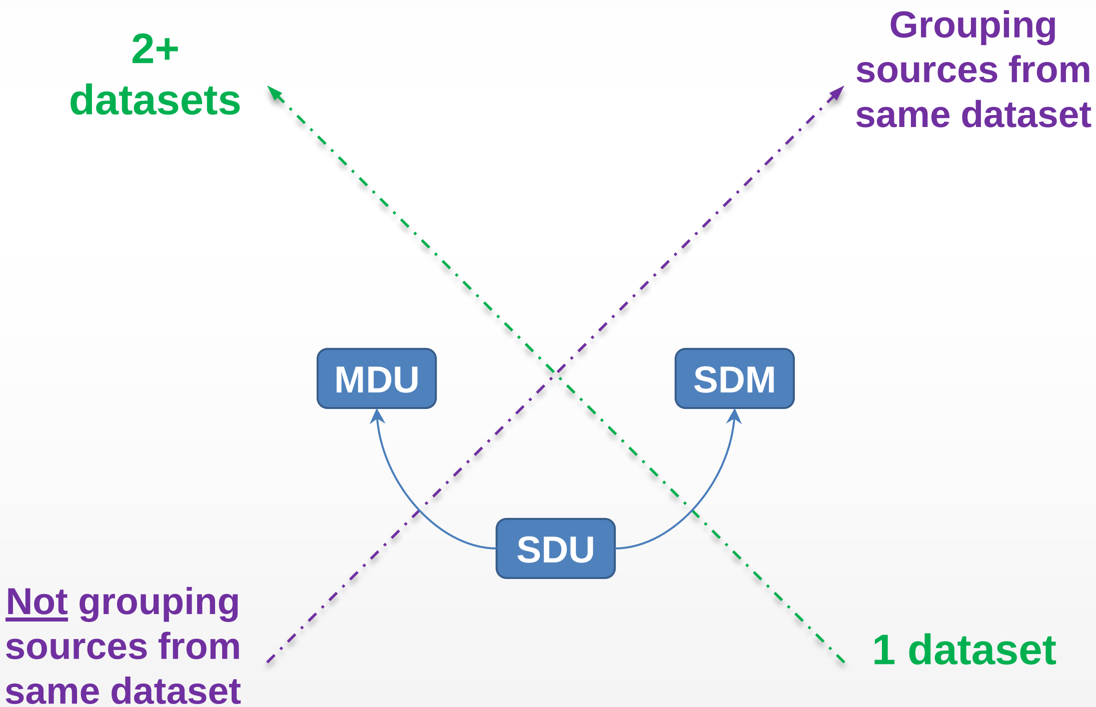




Single Dataset (**SD**) → Multidataset (**MD**)
Unidimensional (**U**) → Multidimensional (**M**)

Multidataset Multidimensional Problems

- A Hierarchy of Blind Source Separation Models *[Silva et al., 2016]*

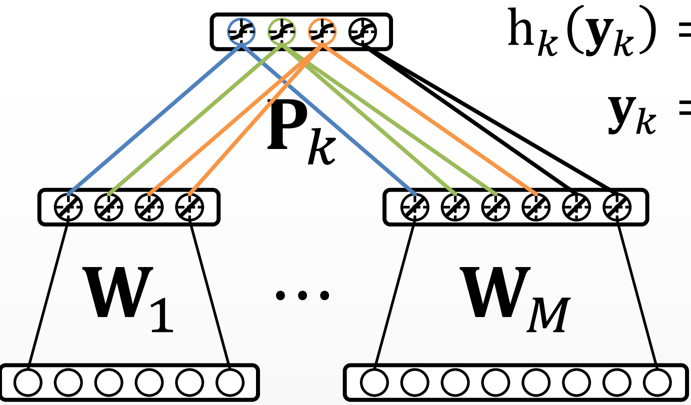


Single Dataset (**SD**) → Multidataset (**MD**)
Unidimensional (**U**) → Multidimensional (**M**)

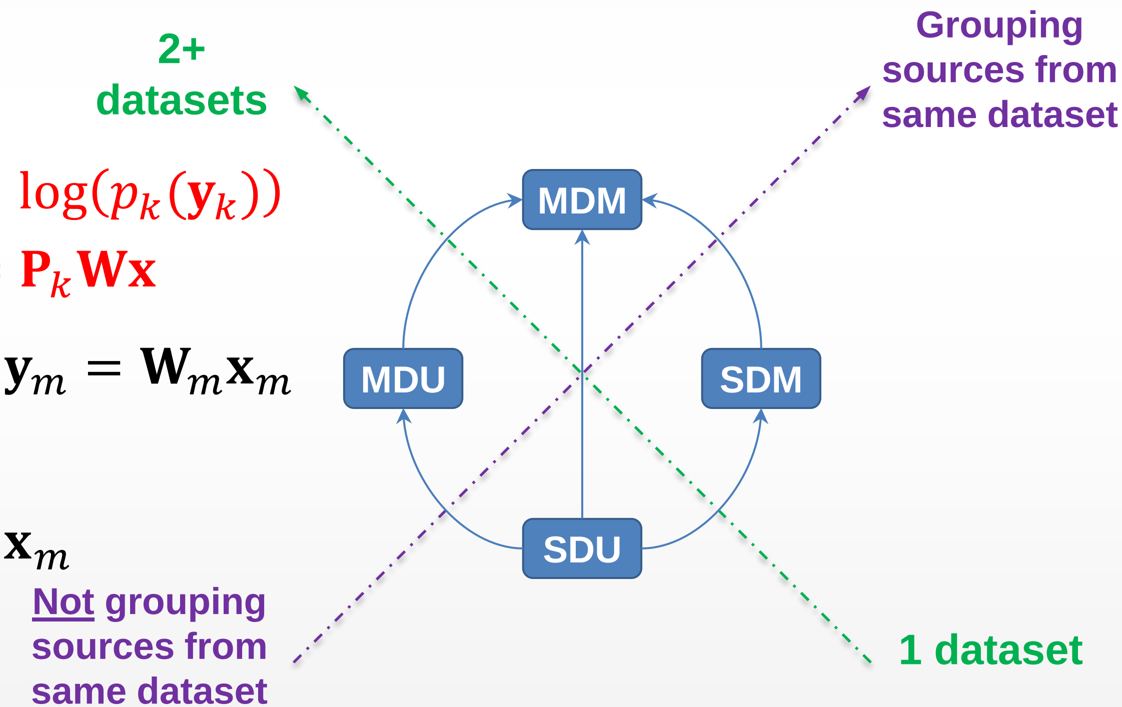
Multidataset Multidimensional Problems

- A Hierarchy of Blind Source Separation Models *[Silva et al., 2016]*

$$J(\mathbf{W}) = KL\langle p(\mathbf{y}), \prod_k^K p_k(\mathbf{y}_k)\rangle,$$
$$= -\sum_m^M \sum_i^{C_m} \log|\sigma_{mi}| - \sum_k^K \mathbb{E}[\mathrm{h}_k(\mathbf{y}_k)],$$



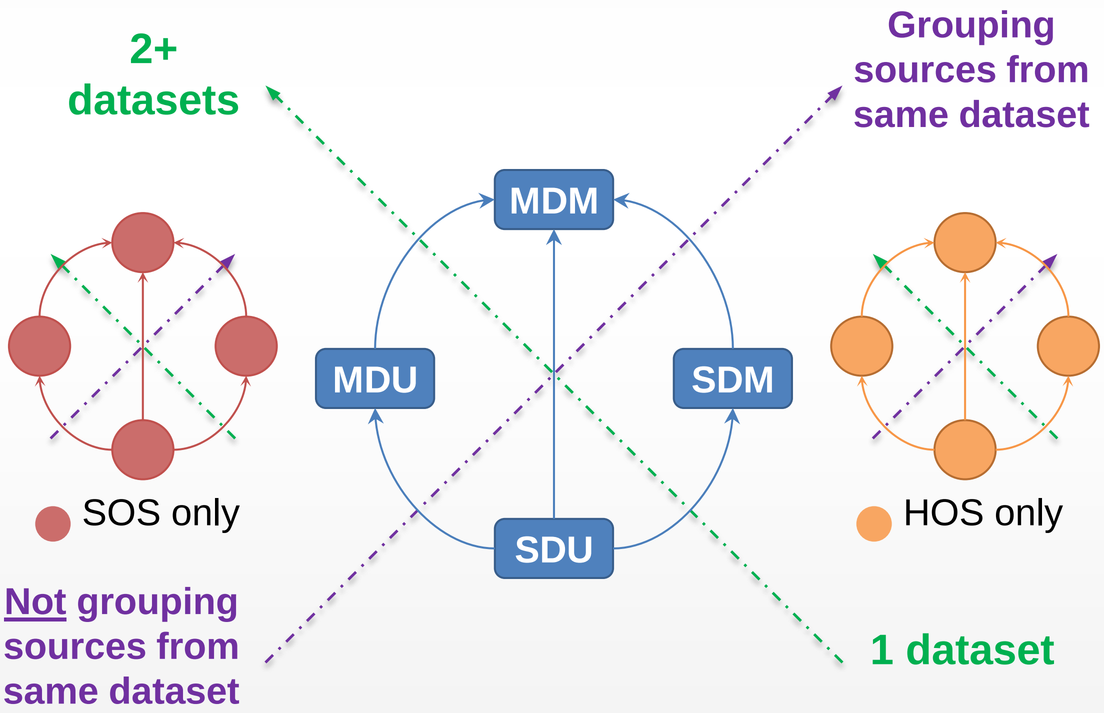
$$\mathrm{h}_k(\mathbf{y}_k) = \log(p_k(\mathbf{y}_k))$$
$$\mathbf{y}_k = \mathbf{P}_k\mathbf{W}\mathbf{x}$$
$$\mathbf{y}_m = \mathbf{W}_m\mathbf{x}_m$$
$$\mathbf{x}_m$$



Single Dataset (**SD**) → Multidataset (**MD**)
Unidimensional (**U**) → Multidimensional (**M**)

Multidataset Multidimensional Problems

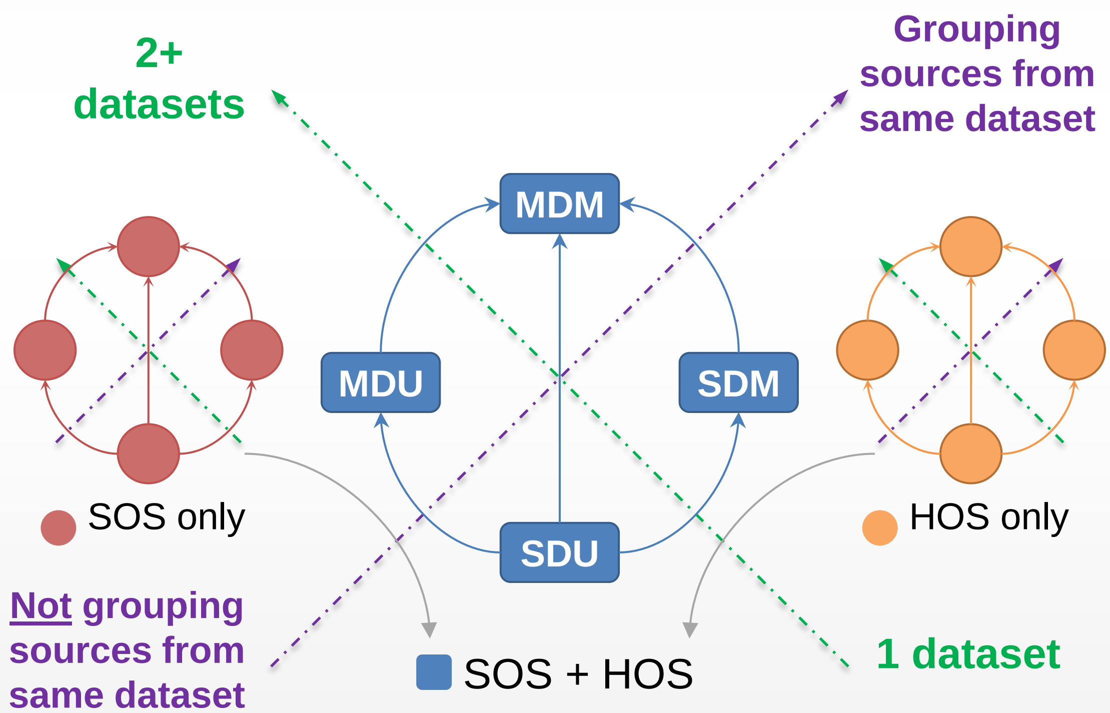
- A Hierarchy of Blind Source Separation Models *[Silva et al., 2016]*



Single Dataset (**SD**) → Multidataset (**MD**)
Unidimensional (**U**) → Multidimensional (**M**)

Multidataset Multidimensional Problems

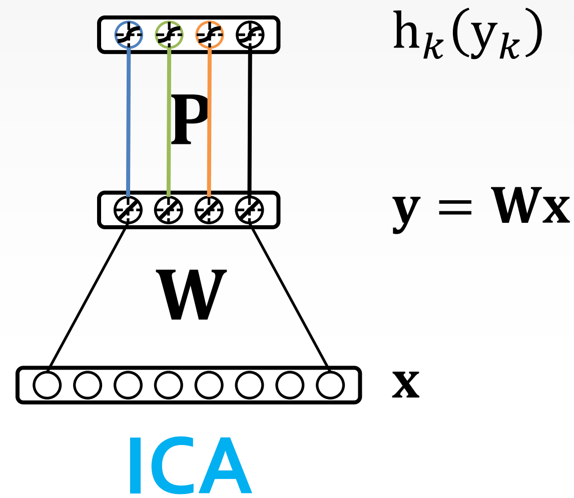
- A Hierarchy of Blind Source Separation Models *[Silva et al., 2016]*



Single Dataset (**SD**) → Multidataset (**MD**)
Unidimensional (**U**) → Multidimensional (**M**)

ICA Regularized with a Linear Autoencoder

$$J(\mathbf{W}) = -\sum_{i}^{C=K} \log|\sigma_i| - \sum_{k}^{K} \mathbb{E}[\mathrm{h}_k(\mathrm{y}_k)]$$

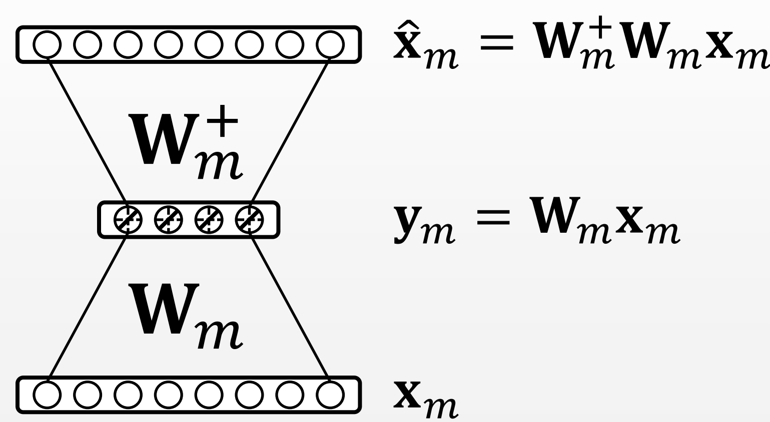


- Assuming white data $\mathbf{z}$, *Le et al., 2011* regularized ICA with a simple linear autoencoder $\hat{\mathbf{z}} = \mathbf{W}^{\top}\mathbf{W}\mathbf{z}$.

- Simultaneously estimate the reduced space while learning statistically independent hidden units.

- Modified linear autoencoder for non-white data: $\hat{\mathbf{x}}_m = \mathbf{W}_m^{+}\mathbf{W}_m\mathbf{x}_m$. Applied to <u>all</u> <u>modalities</u> as a **non-linear constraint**.

$$R(\mathbf{W}_m) = \|\mathbf{W}_m^{+}\mathbf{W}_m\mathbf{x}_m - \mathbf{x}_m\|_2^2$$

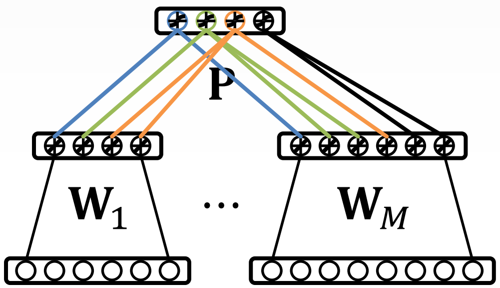


MISA: A Generalized Model for Independent Features

$$J(\mathbf{W}) = KL\langle p(\mathbf{y}), \textstyle\prod_k^K p_k(\mathbf{y}_k)\rangle,\ \mathbf{W} = \mathrm{blkdiag}(\mathbf{W}_m),\ \mathbf{x} = [\mathbf{x}_1^\top \ \cdots \ \mathbf{x}_M^\top]^\top$$

$$J(\mathbf{W}) = -\sum_m^M \sum_i^{C_m} \log|\sigma_{mi}| - \sum_k^K \mathbb{E}[\mathrm{h}_k(\mathbf{y}_k)]$$

$$J(\mathbf{W}) = -\sum_i^{C} \log|\sigma_{i}| - \sum_k^K \mathbb{E}[\mathrm{h}_k(\mathbf{y}_k)]$$



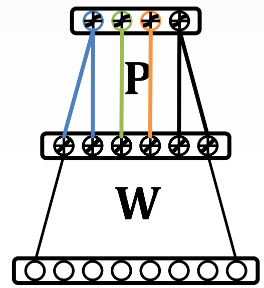
$\mathrm{h}_k(\mathbf{y}_k) = \log(p_k(\mathbf{y}_k|\mathbf{x}))$

$\mathbf{y}_k = \mathbf{P}_k\mathbf{W}\mathbf{x}$

$\mathbf{y}_m = \mathbf{W}_m\mathbf{x}_m$

$\mathbf{x}_m$

MISA



$\mathrm{h}_k(\mathbf{y}_k)$

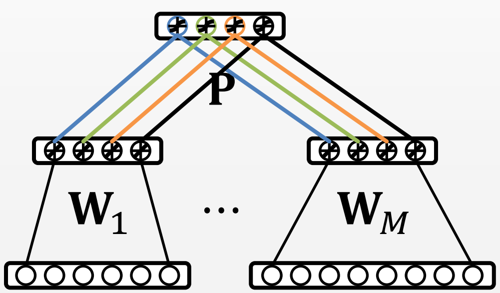
$\mathbf{y} = \mathbf{W}\mathbf{x}$

$\mathbf{x}$

ISA

$$J(\mathbf{W}) = -\sum_m^M \sum_i^{C_m=K} \log|\sigma_{mi}| - \sum_k^K \mathbb{E}[\mathrm{h}_k(\mathbf{y}_k)]$$

$$J(\mathbf{W}) = -\sum_i^{C=K} \log|\sigma_{i}| - \sum_k^K \mathbb{E}[\mathrm{h}_k(y_k)]$$

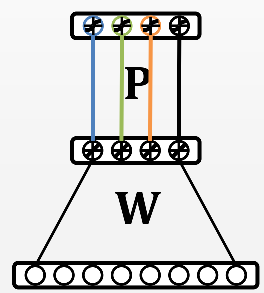


$\mathrm{h}_k(\mathbf{y}_k)$

$\mathbf{y}_m = \mathbf{W}_m\mathbf{x}_m$

$\mathbf{x}_m$

IVA



$\mathrm{h}_k(y_k)$

$\mathbf{y} = \mathbf{W}\mathbf{x}$

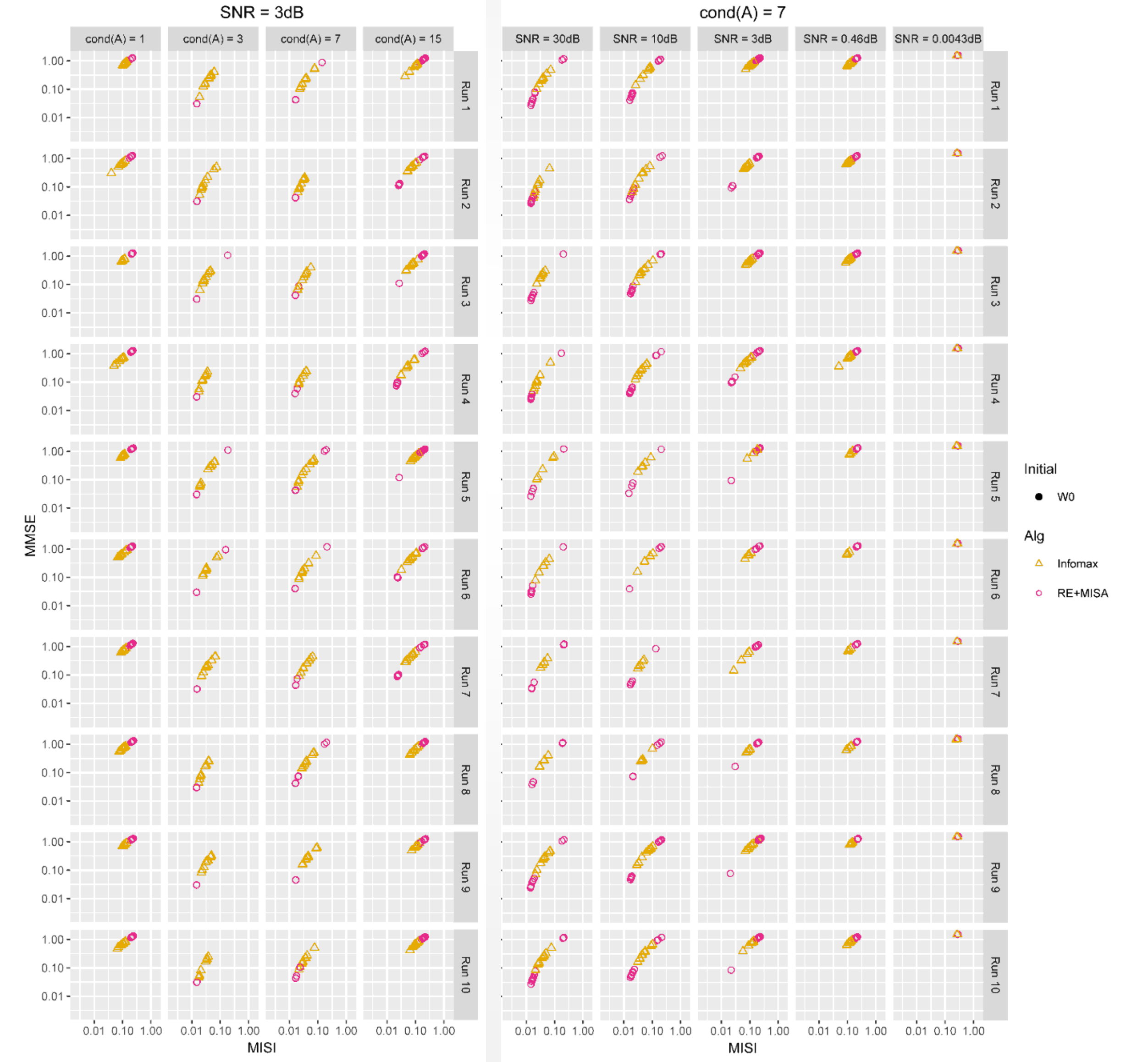
$\mathbf{x}$

ICA

Synthetic Data

ICA

- M = 1 dataset
- C = 75 hidden units into K = 75 one-dimensional subspaces
- N = 3500 examples sampled from a Laplace distribution
- Each run a new, unique (V × C) rectangular mixing matrix A (V = 8000)
- Initialized with ten different random row-orthogonal $\mathbf{W}_0$



Synthetic Data

ISA

– M = 1 Dataset

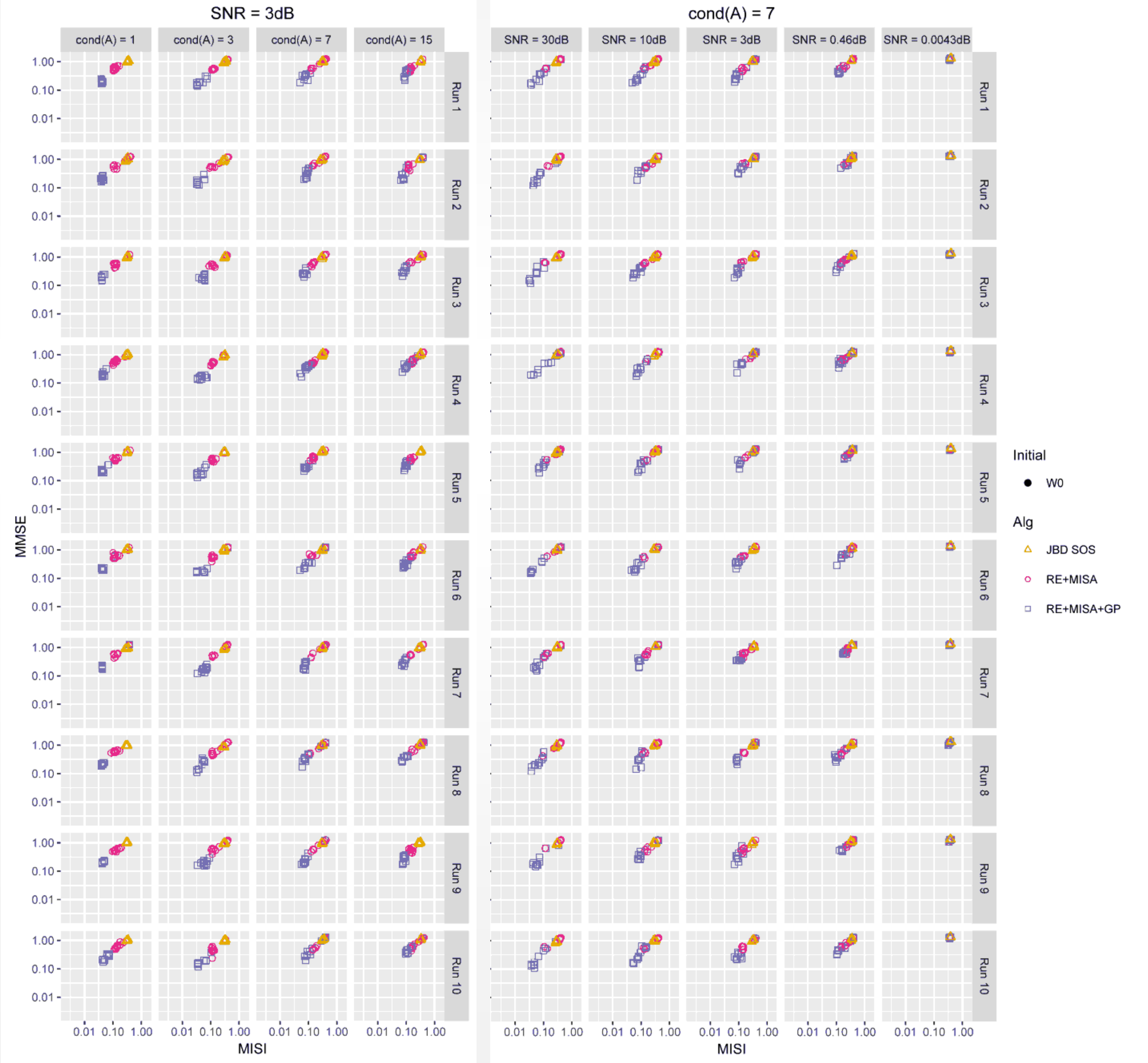
– C = 51 hidden units

– K = 18 d_k -dimensional subspaces, d_k = [1:5; 5:1; 1:5; 2; 2; 2]

– N = 5250 examples sampled from a Multivariate Laplace Distribution

– Each run a new, unique (V × C) rectangular mixing matrix A (V = 8000)

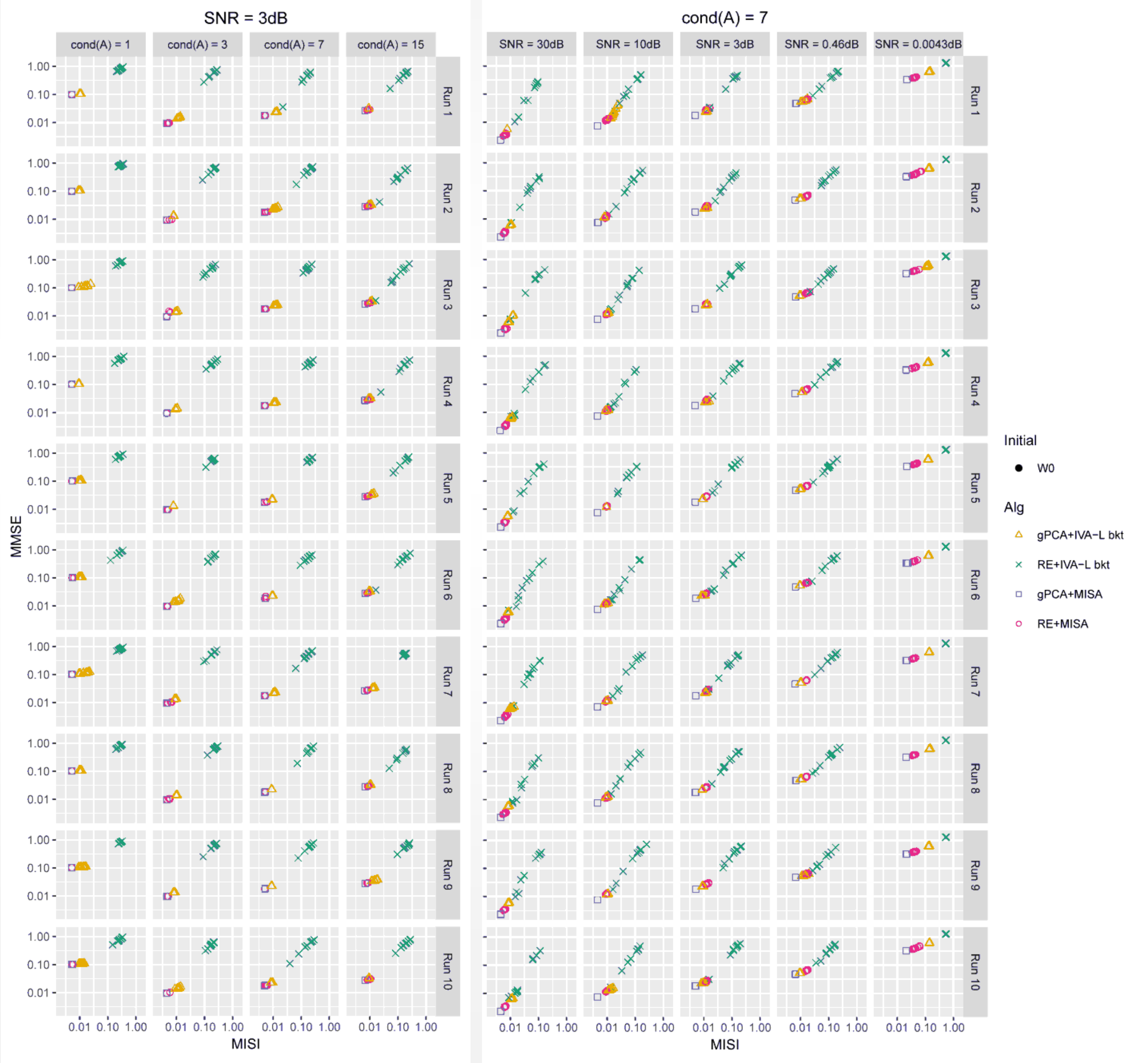
– Initialized with ten different random row-orthogonal $\mathbf{W}_0$



Synthetic Data

IVA

- M = 16 datasets
- C = 75 hidden units into K = 75 sixteen-dimensional subspaces
- N = 66000 examples sampled from a Multivariate Laplace Distribution
- Each run a new, unique (V × C) rectangular mixing matrix A (V = 250)
- Initialized with ten different random row-orthogonal $\mathbf{W}_0$



Hybrid Data

IVA

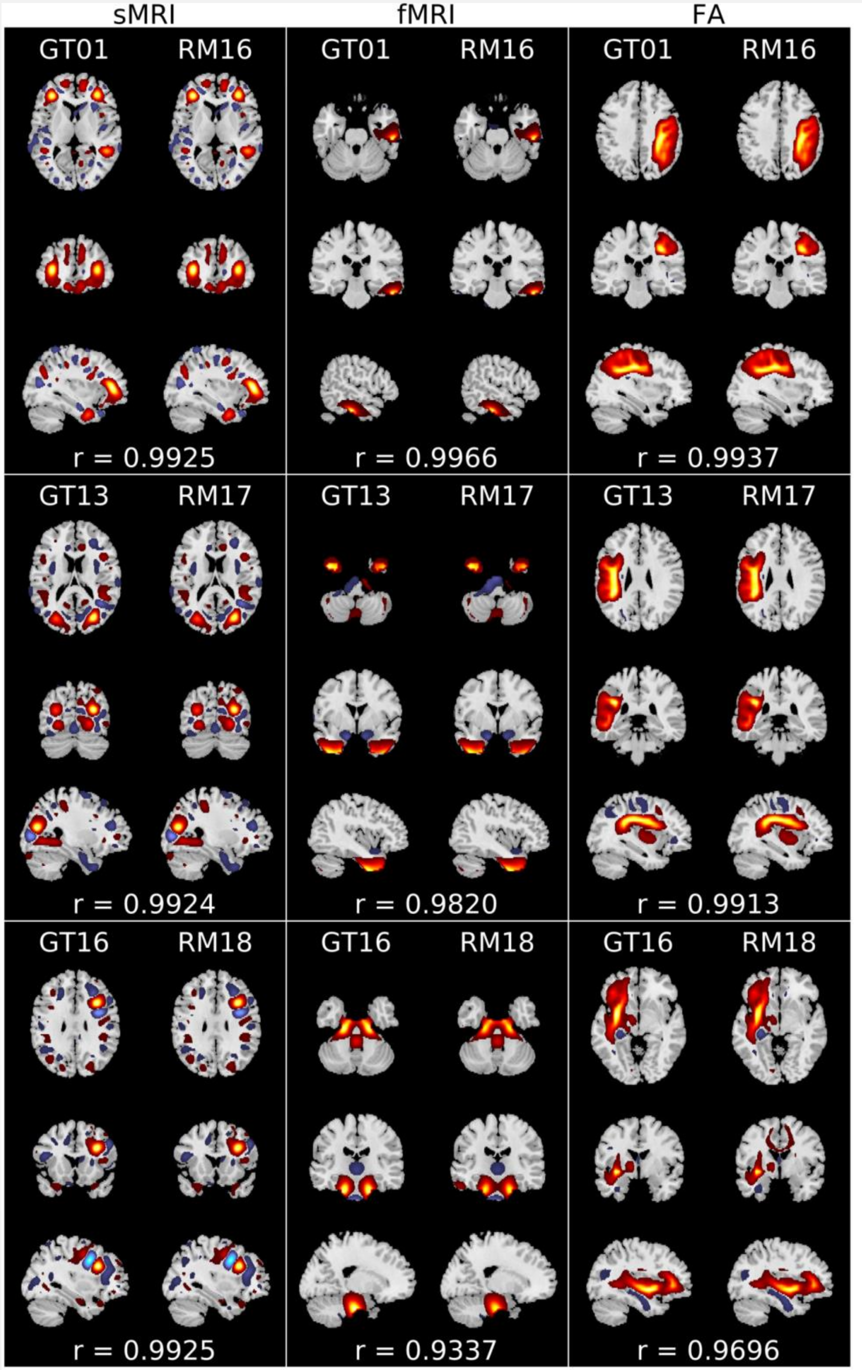
– M = 3 datasets
  – sMRI: V ≈ 300K voxels
  – fMRI: V ≈ 67K voxels
  – FA: V ≈ 15K voxels

– Mixing matrices $\mathbf{A}_m$ are the “real” part of the datasets
  – condition numbers: 1.52, 4.59, 1.63

– C = 20 hidden units into K = 20 three-dimensional subspaces

– N = 600 examples sampled from a Gaussian copula

– Additive Gaussian sensor noise: SNR = 3dB



What's Next?

- Explore deep non-linear models before subspace formation to identify optimal modality-specific depth.

- Learn the strictly sparse Subspace Assignment Matrix $\mathbf{P}$ automatically.

- Modality-specific architectures such as RNNs with hidden independent subspaces for modalities with sequential data