Scaling up Cognitive Neuroscience with HTC

Chris Cox

July 25, 2018

Cognitive Neuroscience in 2018

- Huge online (OpenNeuro, Human Connectome Project, etc.) and offline (lab/department/collaborator archives) data repositories.
- Non-linear optimization routines to align anatomical structures.
- Probabilistic routines for image segmentation.
- High-dimensional signal reconstruction/feature extraction.
- Whole-brain function and connectivity analyses w/ machine learning.
- Modeling and simulating spatiotemporal signal.
- Real-time neurofeedback.

The science continues to scale up ...

The science continues to scale up …

- More data.

The science continues to scale up …

- More data.
- More intensive pre-processing.

The science continues to scale up …

- More data.
- More intensive pre-processing.
- More sophisticated models …

The science continues to scale up …

- More data.
- More intensive pre-processing.
- More sophisticated models …
- … leveraging More structure.

The science continues to scale up …

- More data.
- More intensive pre-processing.
- More sophisticated models …
- … leveraging More structure.
- More responsive and personalized experimental tools.

The science continues to scale up …

- More data.
- More intensive pre-processing.
- More sophisticated models …
- … leveraging More structure.
- More responsive and personalized experimental tools.

The science continues to scale up …

- More data.
- More intensive pre-processing.
- More sophisticated models …
- … leveraging More structure.
- More responsive and personalized experimental tools.

- More computational demand, meaning …

The science continues to scale up …

- More data.
- More intensive pre-processing.
- More sophisticated models …
- … leveraging More structure.
- More responsive and personalized experimental tools.

- More computational demand, meaning …
- ***<u>More time between having an idea and obtaining a result.</u>***

High Throughput Computing

- Highly effective if a task can be split into independent pieces.
- Scaling up the science can be accommodated by recruiting more machines.
- HTC is widely available through the Open Science Grid.

HTC is widely applicable to cognitive neuroscience

- Many conventional and state of the art procedures can be naturally divided into pieces that can be run in parallel.

Image coregistration, normalization, and segmentation

- Each subject can be allocated to a different machine.
- Accommodate more complex image processing, and keep processing time under control.

- *Scale up to datasets with hundreds or thousands of participants.*

Univariate Analyses and Searchlight Analyses

- Voxels and searchlights are independent.
- Sophisticated modeling can be done at each voxel/searchlight by distributing load over multiple machines.

- *Scale up to datasets with hundreds or thousands of participants.*

Relating brain and behavior with Machine Learning

- Cross validation.
- Hyper-parameter selection.
- Permutation testing.

- A full analysis can involve hundreds of thousands of models, ***but each is independent***.

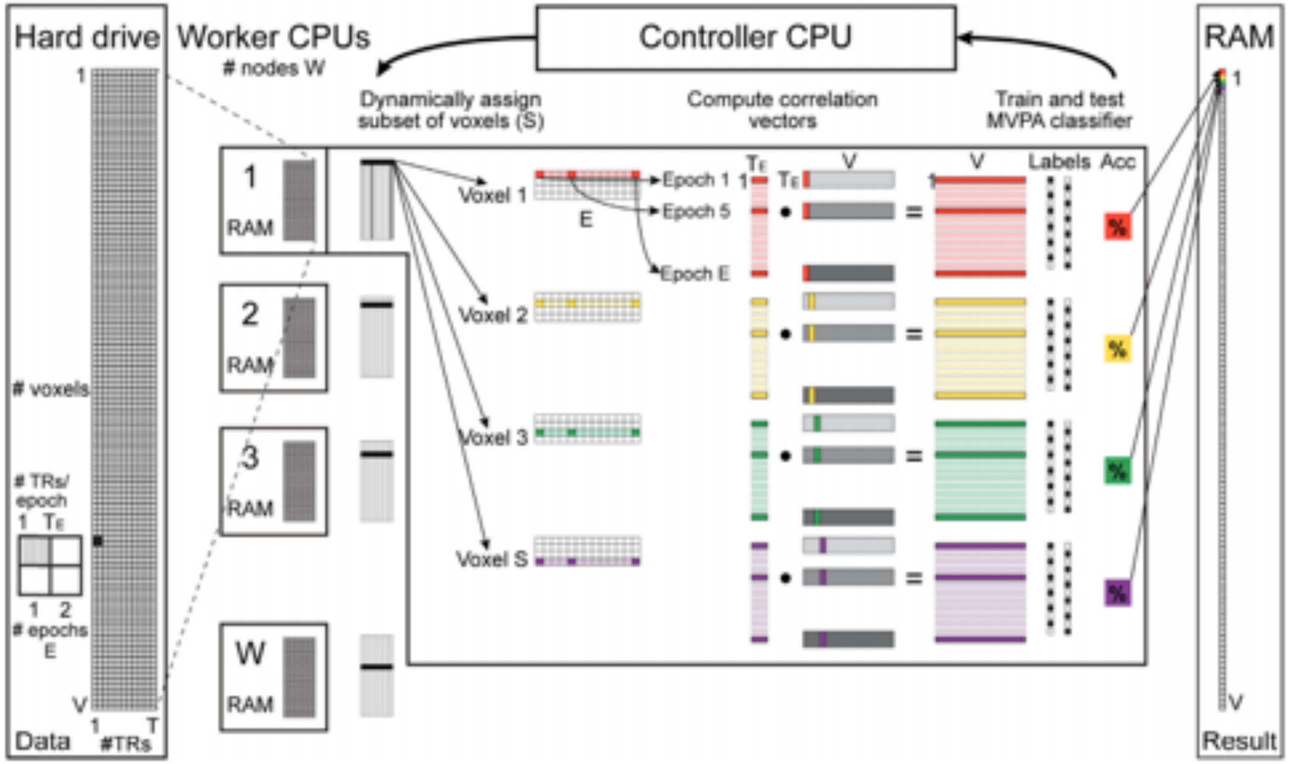
Network Discovery through Whole-brain modeling

Consider “full correlation matrix analysis” in fMRI (Wang et al. 2015)

Network Discovery through Whole-brain modeling

Consider "full correlation matrix analysis" in fMRI (Wang et al. 2015)

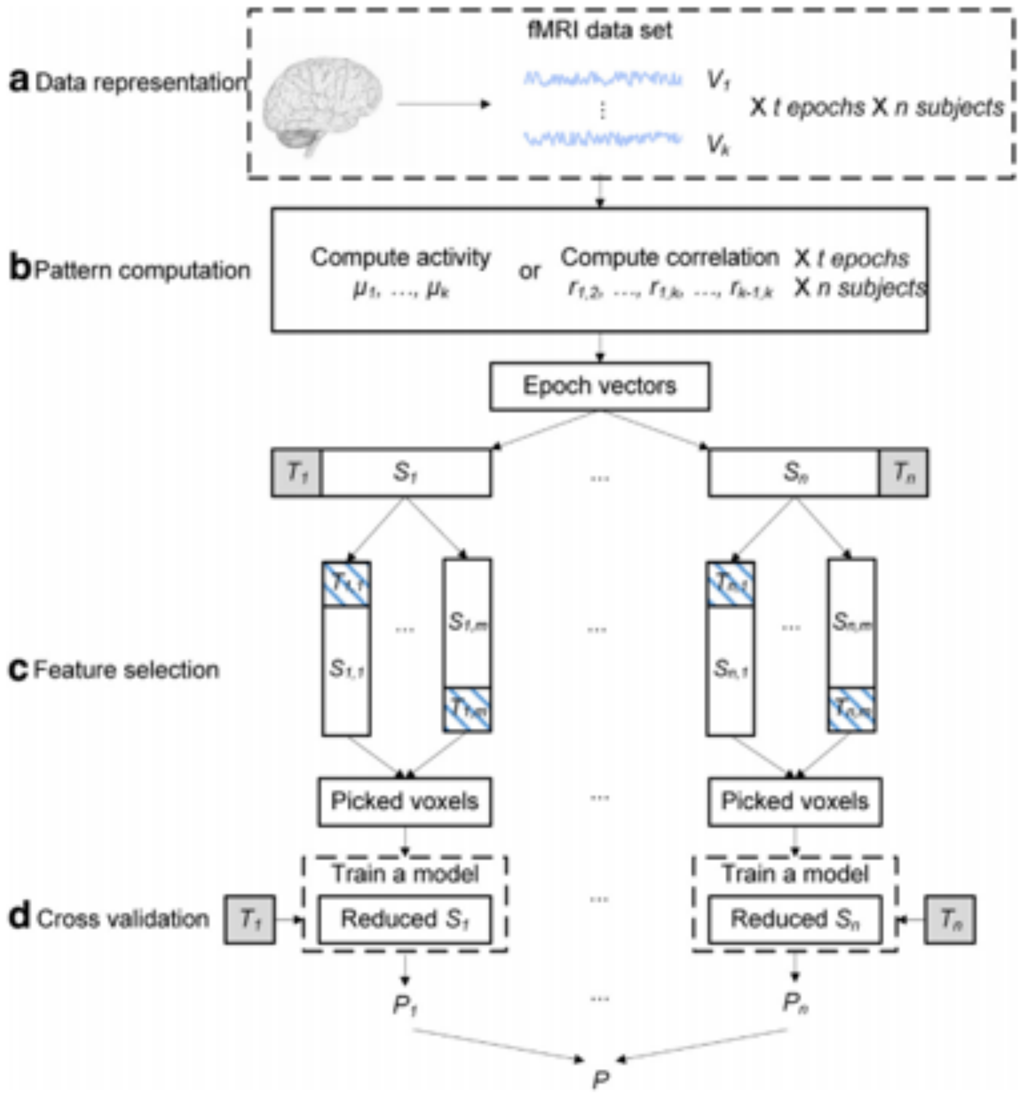
Y. Wang et al. / Journal of Neuroscience Methods 251 (2015) 108–119



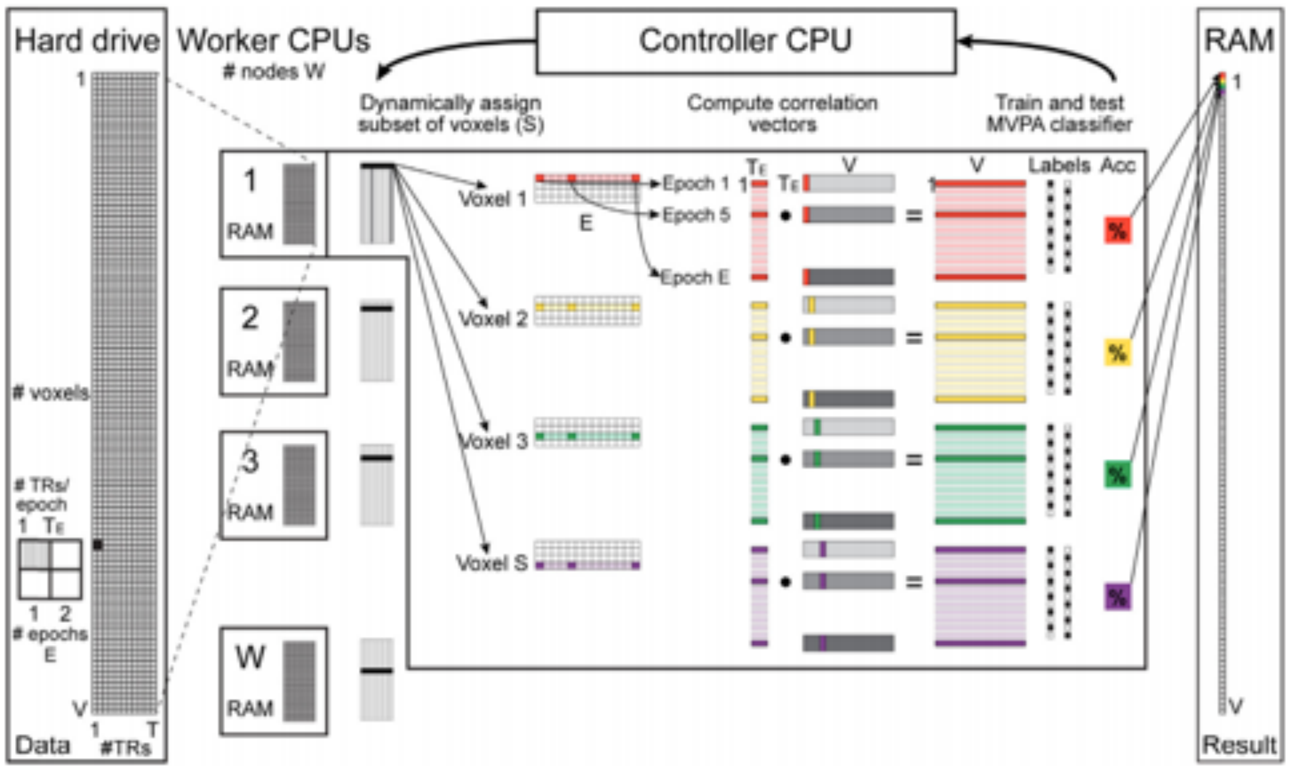
Network Discovery through Whole-brain modeling

Consider “full correlation matrix analysis” in fMRI (Wang et al. 2015)

Y. Wang et al. / Journal of Neuroscience Methods 251 (2015) 108–119



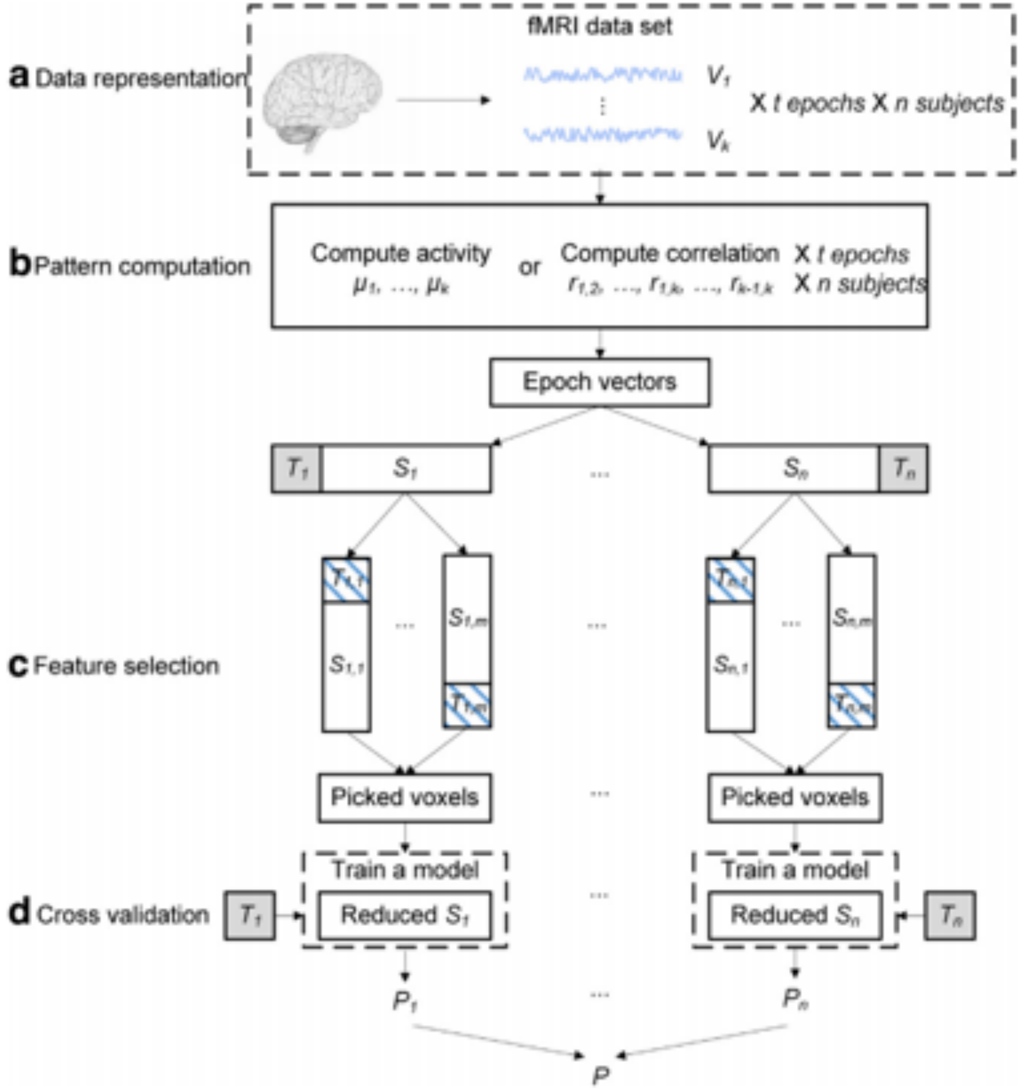
Y. Wang et al. / Journal of Neuroscience Methods 251 (2015) 108–119



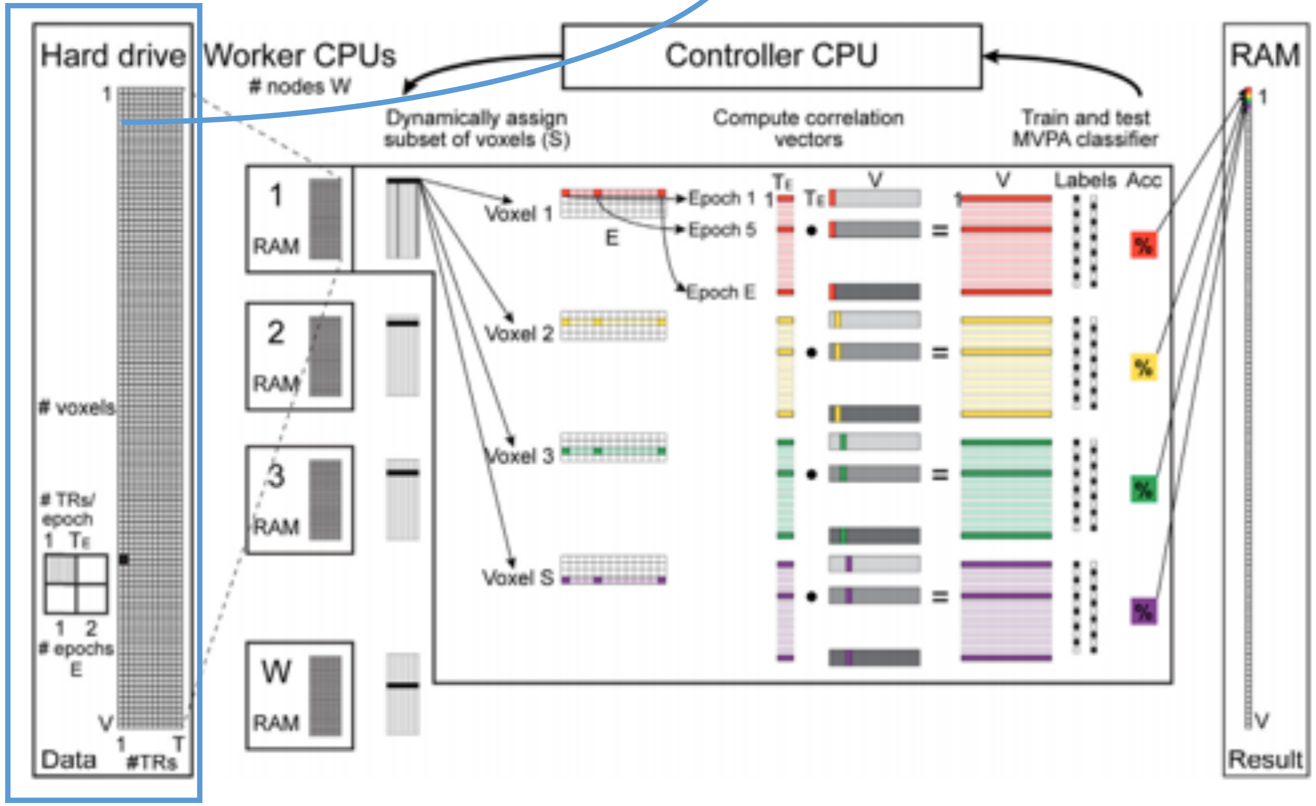
Network Discovery through Whole-brain modeling

Consider "full correlation matrix analysis" in fMRI (Wang et al. 2015)

Y. Wang et al. / Journal of Neuroscience Methods 251 (2015) 108–119

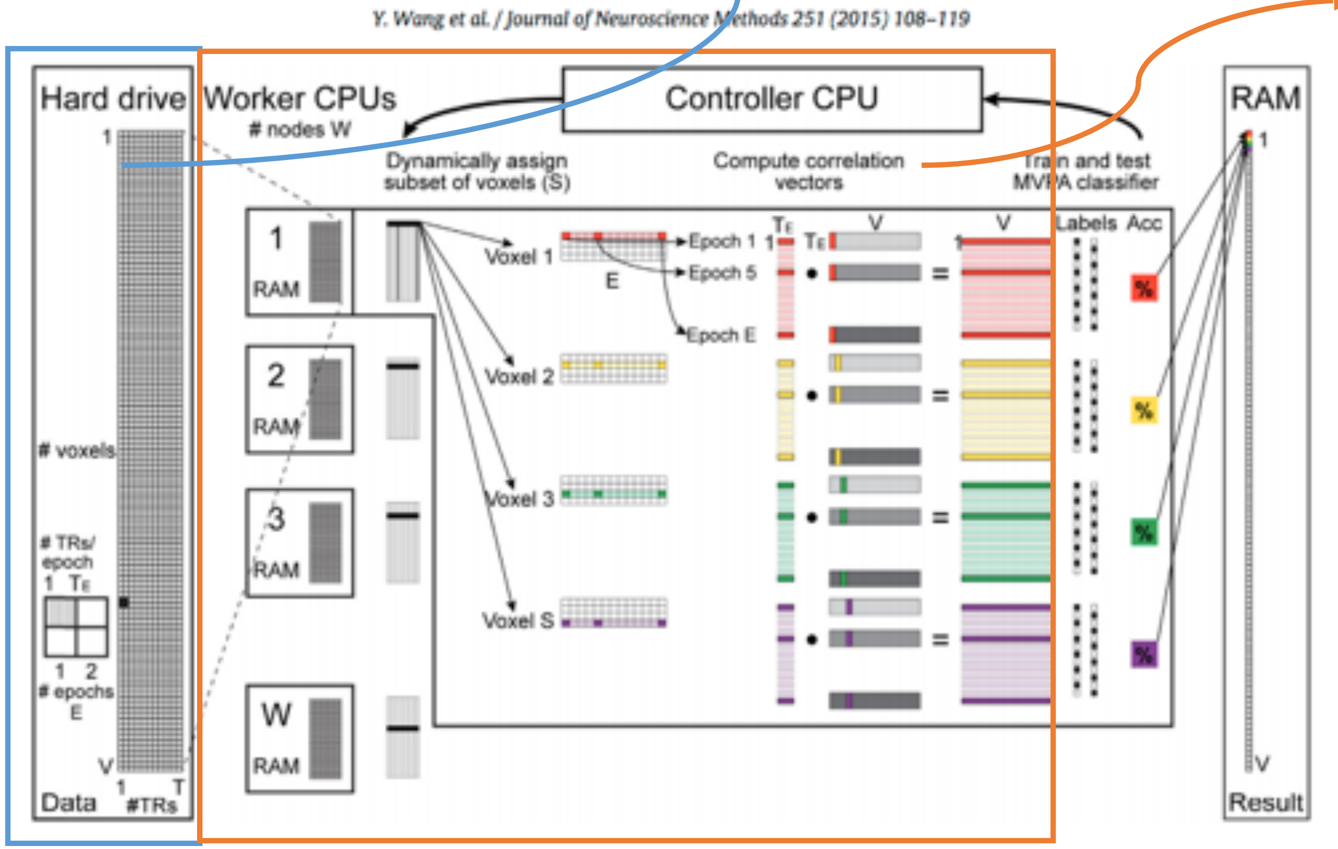


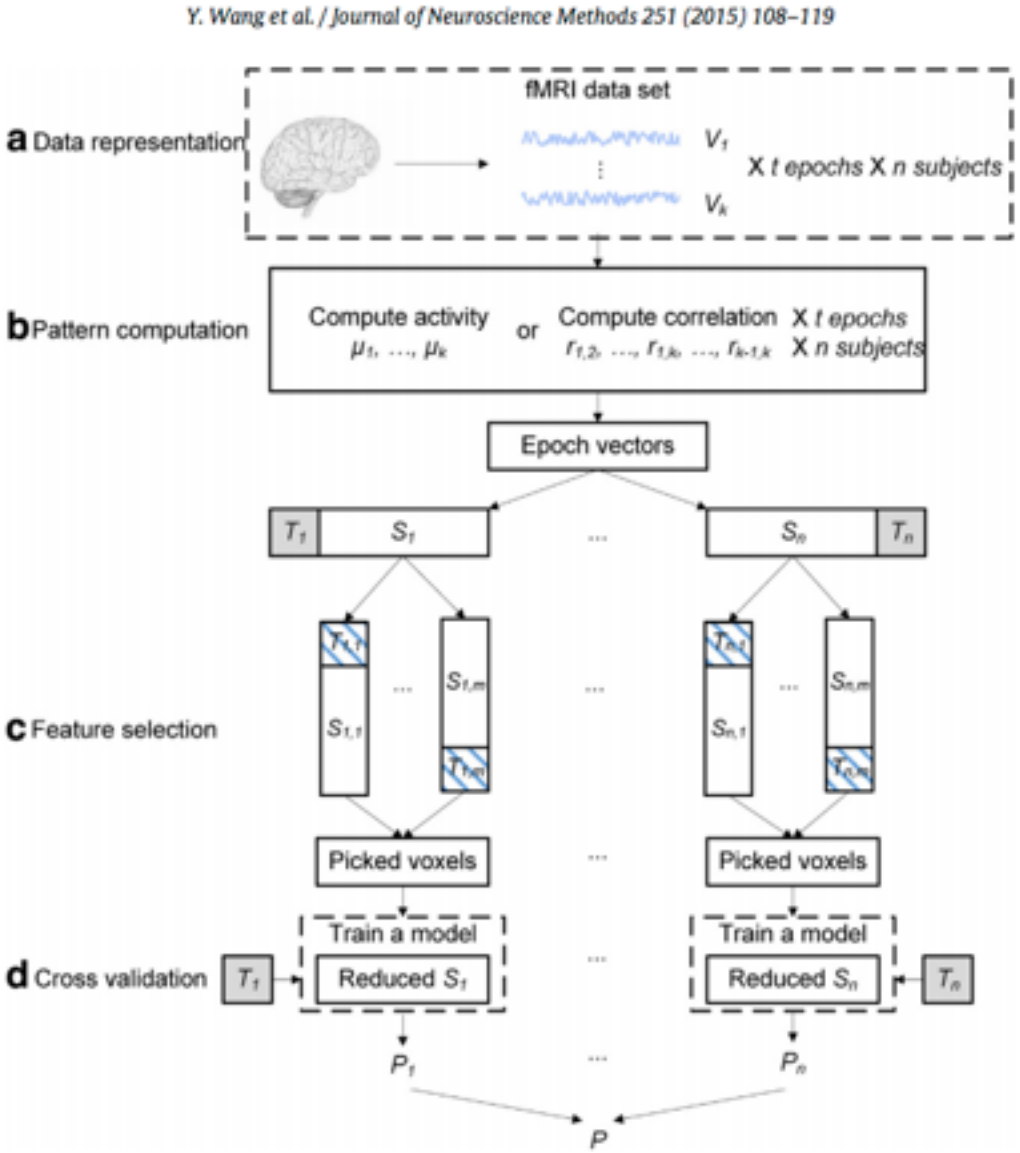
Y. Wang et al. / Journal of Neuroscience Methods 251 (2015) 108–119



Network Discovery through Whole-brain modeling

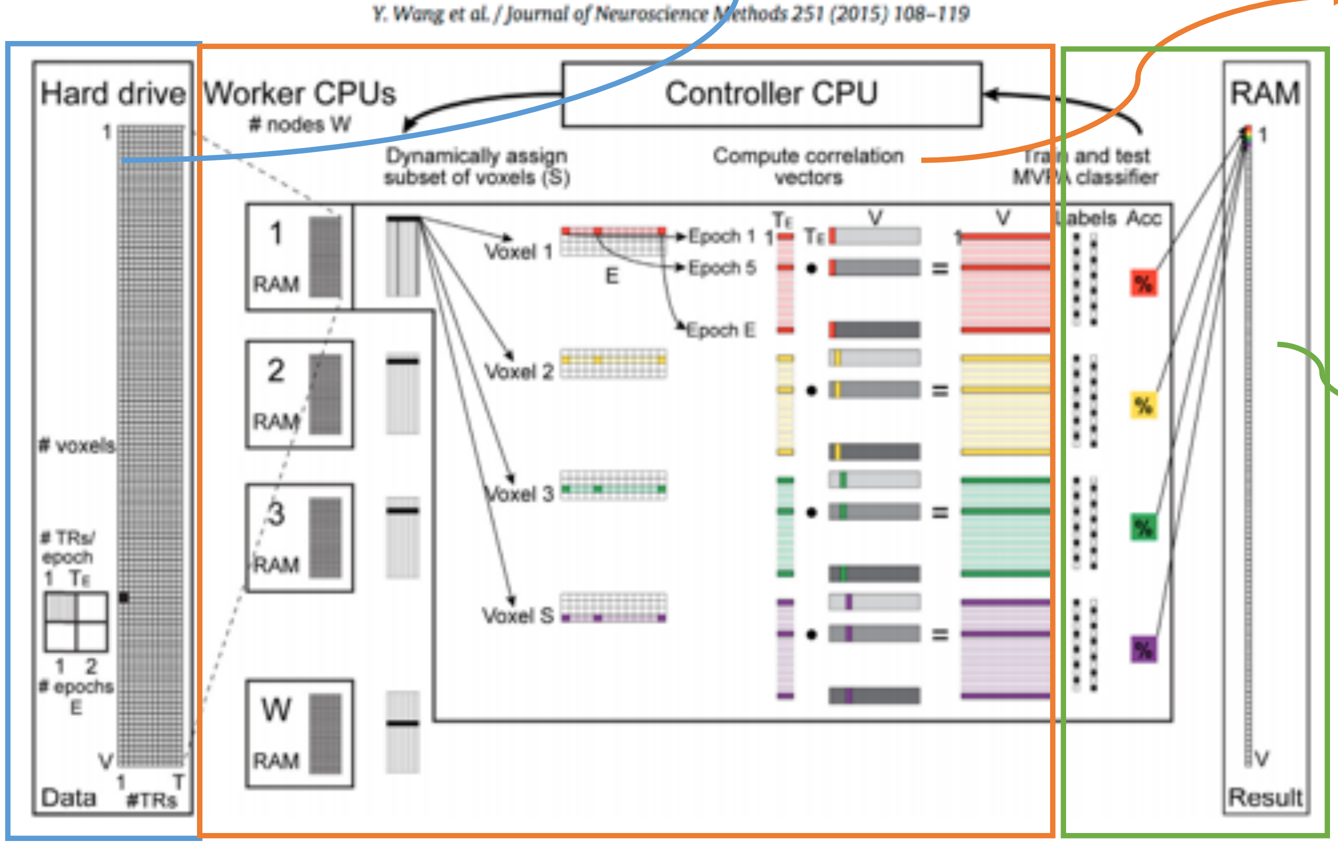
Consider "full correlation matrix analysis" in fMRI (Wang et al. 2015)

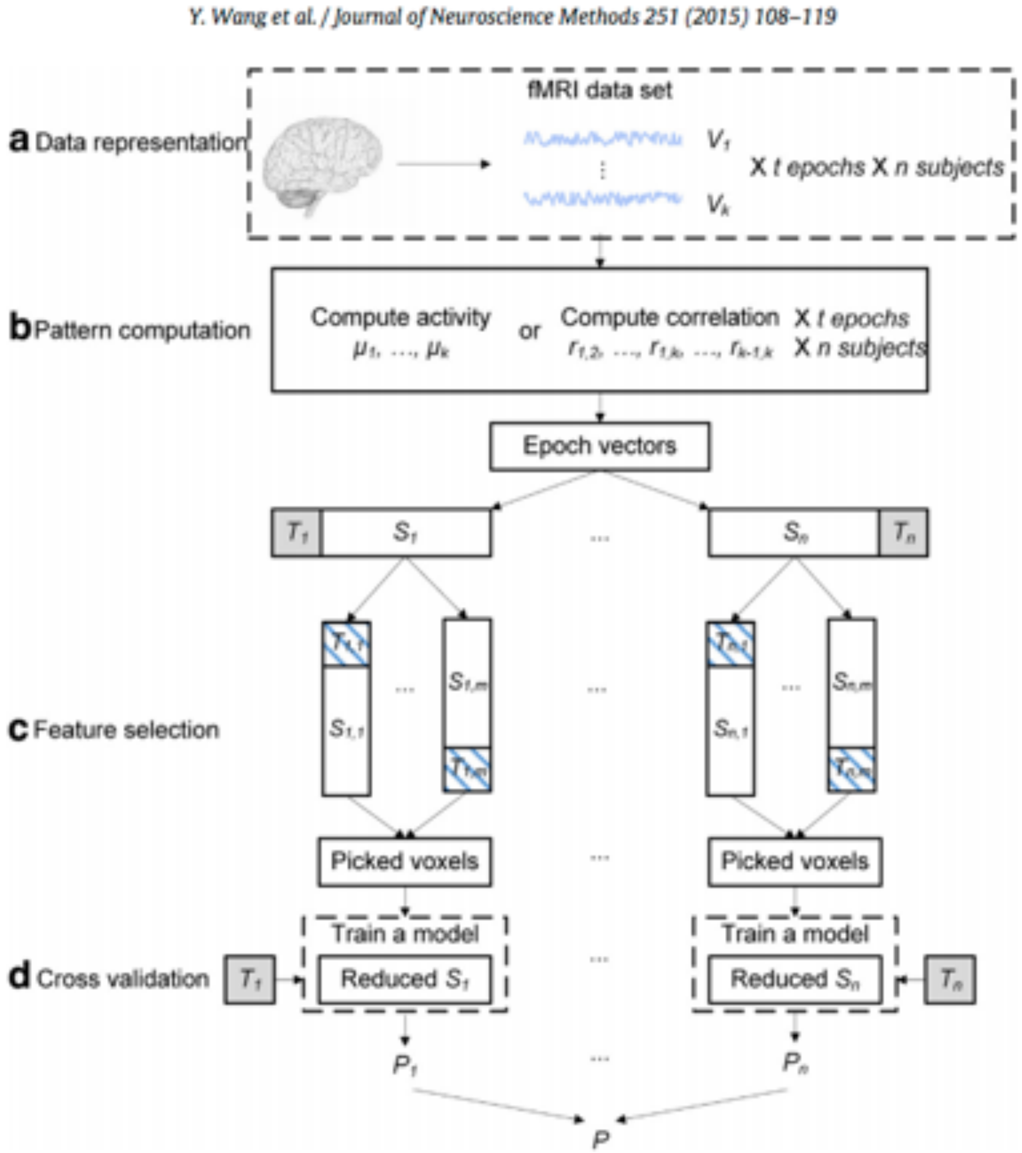




Network Discovery through Whole-brain modeling

Consider "full correlation matrix analysis" in fMRI (Wang et al. 2015)



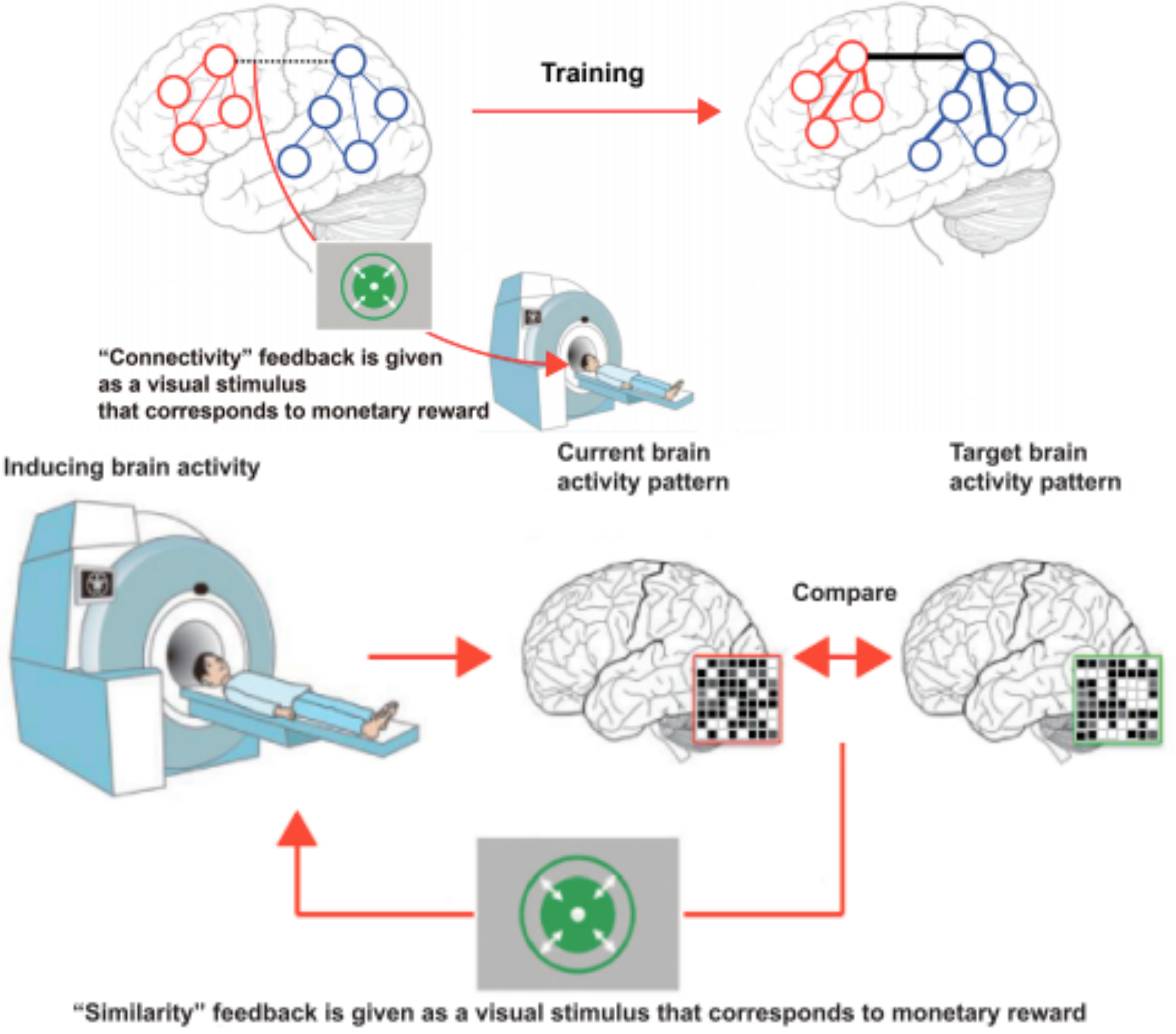


Real-time Neurofeedback

- Brain may be segmented and processed in parallel.
- Cross validation can be performed in parallel if necessary.

Yamada et al.

2017



Theoretically Motivated Brain Modeling

Theoretically Motivated Brain Modeling

- What if the brain encodes semantic knowledge with distributed representation?

Theoretically Motivated Brain Modeling

- What if the brain encodes semantic knowledge with distributed representation?

1. Multivariate signal.

Theoretically Motivated Brain Modeling

- What if the brain encodes semantic knowledge with distributed representation?

1. Multivariate signal.
2. Same region encodes multiple things.

Theoretically Motivated Brain Modeling

- What if the brain encodes semantic knowledge with distributed representation?

1. Multivariate signal.
2. Same region encodes multiple things.
3. Same region can encode the same content with different patterns in different people.

Theoretically Motivated Brain Modeling

- What if the brain encodes semantic knowledge with distributed representation?

1. Multivariate signal.
2. Same region encodes multiple things.
3. Same region can encode the same content with different patterns in different people.
4. Patterns need not be localized.

Theoretically Motivated Brain Modeling

Theoretically Motivated Brain Modeling

- What methods exist?

Theoretically Motivated Brain Modeling

- What methods exist?

1. Univariate analysis assumes a very different kind of activation.

Theoretically Motivated Brain Modeling

- What methods exist?

1. Univariate analysis assumes a very different kind of activation.
2. Searchlight assumes information is local.

Theoretically Motivated Brain Modeling

- What methods exist?

1. Univariate analysis assumes a very different kind of activation.
2. Searchlight assumes information is local.
3. Sparse regression (e.g. LASSO) ignores all spatial information within and across subjects.

Theoretically Motivated Brain Modeling

- What methods exist?

1. Univariate analysis assumes a very different kind of activation.
2. Searchlight assumes information is local.
3. Sparse regression (e.g. LASSO) ignores all spatial information within and across subjects.

Theoretically Motivated Brain Modeling

- What methods exist?

1. Univariate analysis assumes a very different kind of activation.
2. Searchlight assumes information is local.
3. Sparse regression (e.g. LASSO) ignores all spatial information within and across subjects.

- Need a whole-brain model that is flexibly sensitive to localization, while still being sensitive to complex patterns that vary across people.

Sparse Overlapping Sets Lasso

Sparse Overlapping Sets Lasso

- Rao, Cox, Rogers, and Nowak (2013; 2015) introduced SOS Lasso, which is a complex optimization with multiple hyperparameters.
- No way to estimate them without searching the parameter space.
- No way to evaluate performance without cross validation.
- No way to understand voxel significance without permutation testing.
- Models are fit to all voxels in cortex for all participants in an experiment simultaneously.

- An exciting method, with intense computational needs.

HTC is widely applicable to cognitive neuroscience