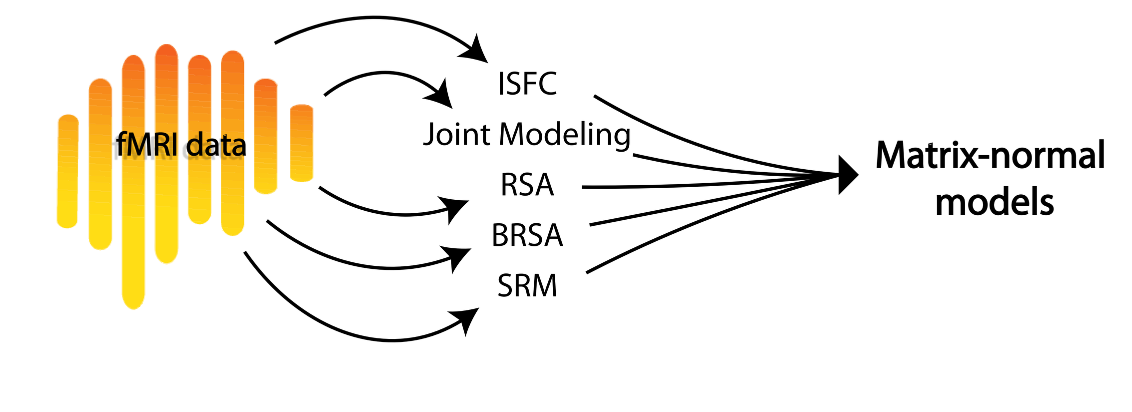
# Matrix-normal models for fMRI analysis

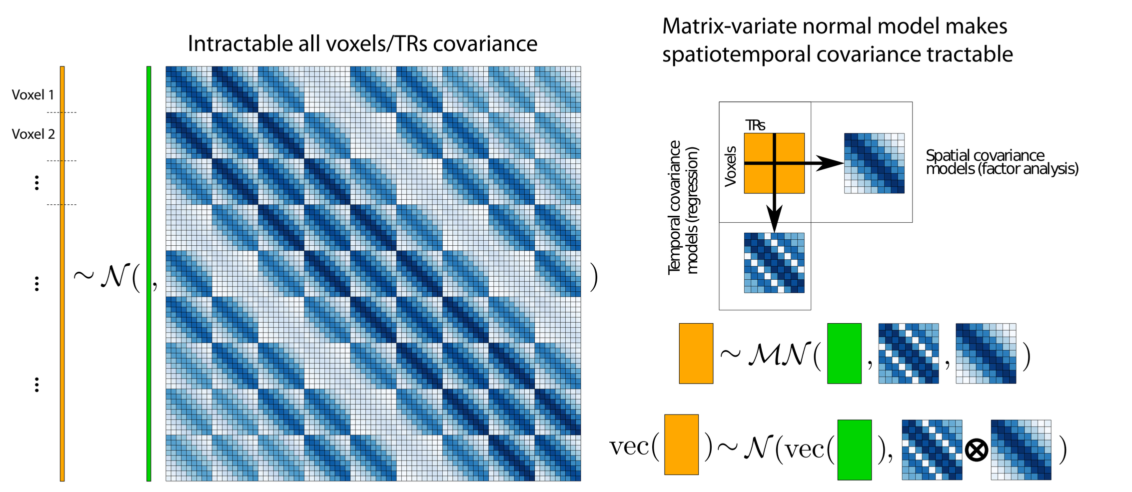
Michael Shvartsman[1], Narayanan Sundaram[2], Mikio C. Aoi[1], Adam Charles[1], Theodore L. Willke[2], Jonathan D. Cohen[1]

Princeton University    Parallel Computing Lab, Intel Corporation

## Our main contribution




## The matrix-variate normal distribution




## One generative model captures many existing analyses [1]

$$\begin{aligned}
&\mathbf{Y}_i \mid \mathbf{F}_i, \mathbf{B}_i, \mathbf{Z}, \mathbf{W}_i, \Sigma_i, \Omega \sim \\
&\quad \mathcal{MN}(\mathbf{F}_i\mathbf{Z} + \mathbf{B}_i\mathbf{X} + \mathbf{J}\mathbf{W}, \rho_i^2\Sigma, \Omega) \\
&\mathbf{F}_i \mid \mathbf{C}, \mathbf{U} \sim \mathcal{MN}(0, \mathbf{C}, \mathbf{U}) \\
&\mathbf{Z} \mid \mathbf{D}, \mathbf{V} \sim \mathcal{MN}(0, \mathbf{D}, \mathbf{V}) \\
&\mathbf{B}_i \mid \mathbf{G}, \mathbf{K} \sim \mathcal{MN}(\beta_0, \mathbf{G}, \mathbf{K}) \\
&\mathbf{W}_i \mid \mathbf{H}, \mathbf{R} \sim \mathcal{MN}(\mathbf{W}_0, \mathbf{H}, \mathbf{R}).
\end{aligned}$$

- $\mathbf{Y}_i$: data for subject $i$. $\mathbf{X}/\mathbf{J}$ are temporal/spatial design matrices.
- This poster: temporal cov. $\Omega$ is AR(1), spatial $\Sigma, \mathbf{C}$ are diagonal.

## brainiak.matnormal: a prototyping tool for matrix-normal models

**MN-RSA implemented in <60 lines of code!**

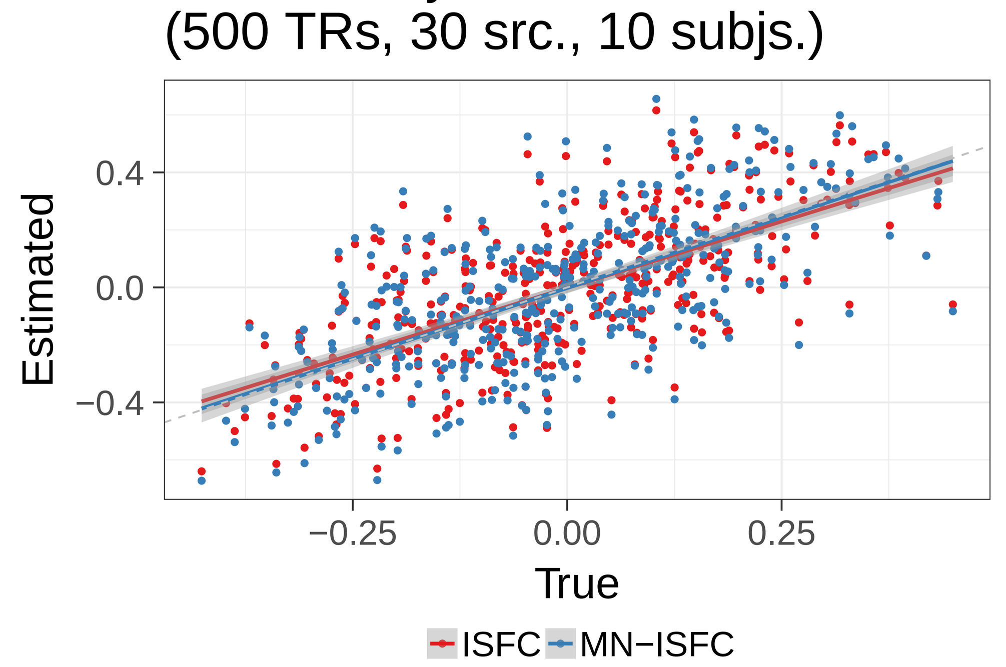
```
rsa_cov = CovFullRankCholesky(size=k)
space_noise_cov = CovDiagonal(size=v)
time_noise_cov = CovAR1(size=t)
params = [rsa_cov.get_optimize_vars(),
          time_noise_cov.get_optimize_vars(),
          space_noise_cov.get_optimize_vars()]
loss = -(time_noise_cov.logp +
         space_noise_cov.logp +
         rsa_cov.logp +
         matnorm_logp_marginal_row(Y,
            row_cov=time_noise_cov,
            col_cov=space_noise_cov,
            marg=X, marg_cov=rsa_cov))
optimizer.minimize(loss)
U = rsa_cov.Sigma
C = cov2corr(U)
```

**Automatic marginalization and covariance structure selection.**

## MN-ISFC: new MLE estimator

- **Guaranteed to return valid covariance**, comparable RMSE to original method[4].

Est. Corr, synth. data
(500 TRs, 30 src., 10 subjs.)



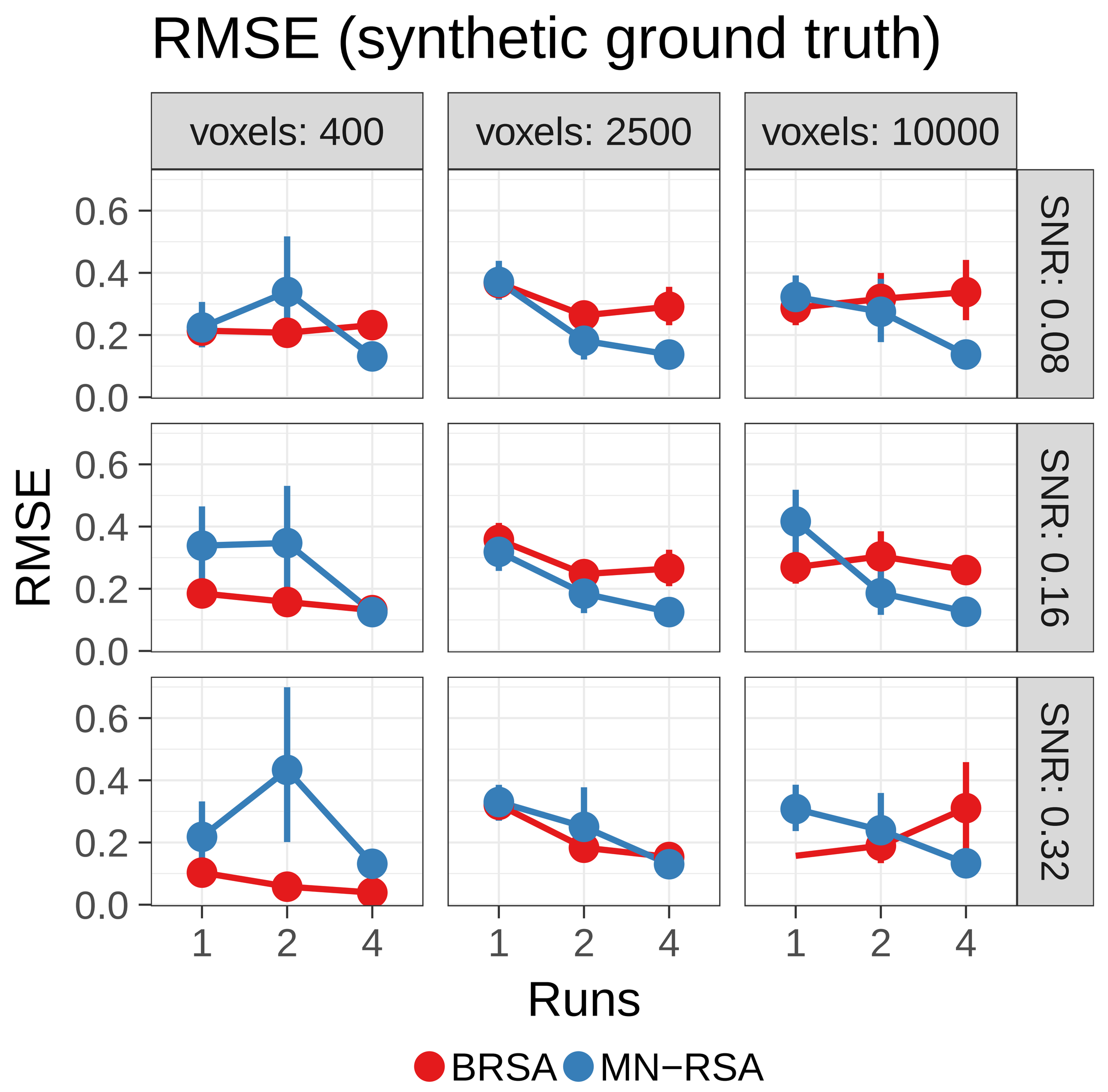

## MN-RSA: faster and more accurate at large data, unbiased

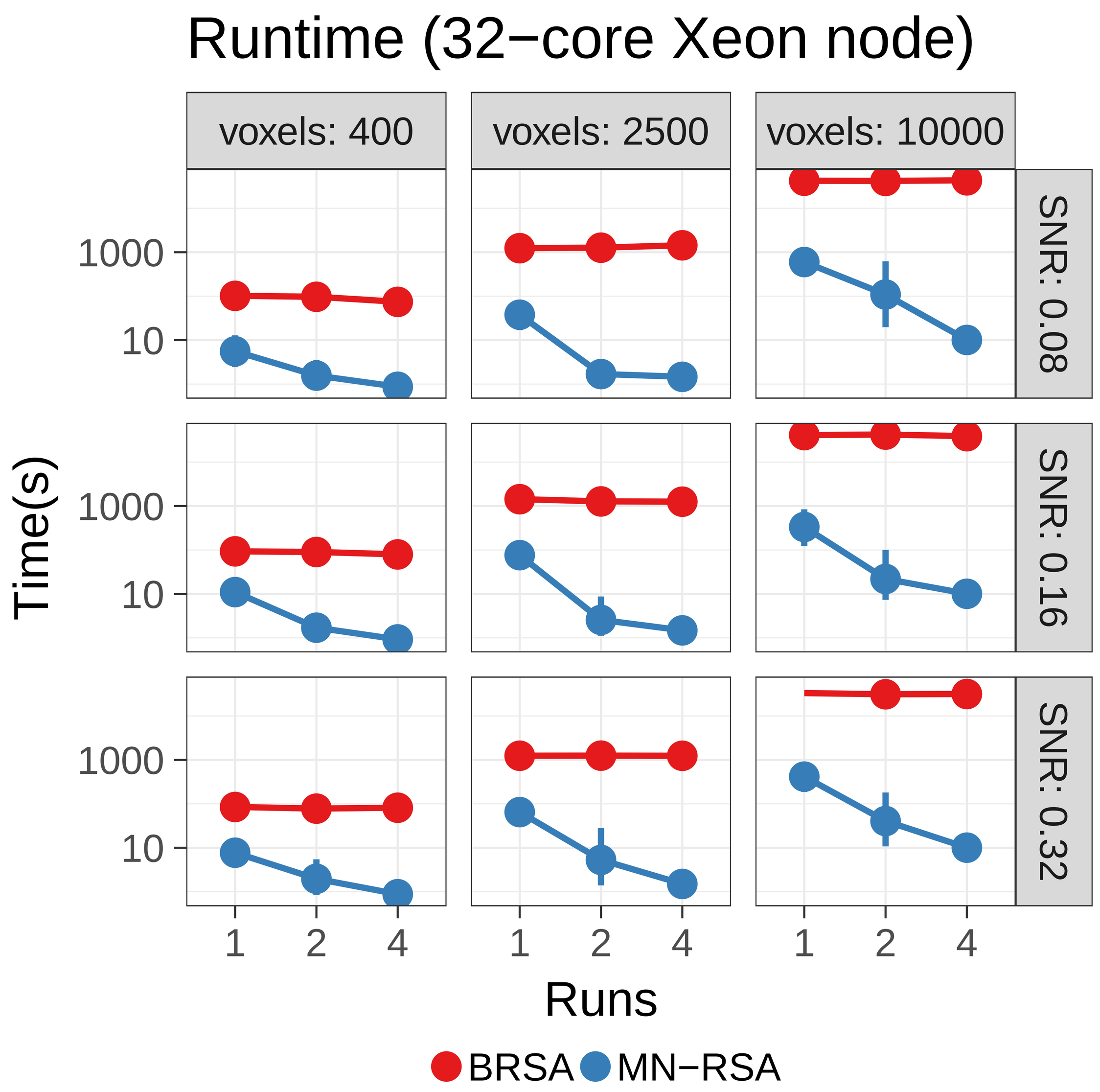
$$\begin{aligned}
&\mathbf{Y}_i \mid \mathbf{F}_i, \mathbf{B}_i, \mathbf{Z}, \mathbf{W}_i, \Sigma_i, \Omega \sim \\
&\quad \mathcal{MN}(\mathbf{F}_i\mathbf{Z} + \mathbf{B}_i\mathbf{X} + \mathbf{J}\mathbf{W}, \rho_i^2\Sigma, \Omega) \\
&\mathbf{F}_i \mid \mathbf{C}, \mathbf{U} \sim \mathcal{MN}(0, \Sigma, \mathbf{U}) \\
&\mathbf{Z} \mid \mathbf{D}, \mathbf{V} \sim \mathcal{MN}(0, \mathbf{D}, \mathbf{V}) \\
&\mathbf{B}_i \mid \mathbf{G}, \mathbf{K} \sim \mathcal{MN}(\beta_0, \Sigma, \mathbf{K}) \\
&\mathbf{W}_i \mid \mathbf{H}, \mathbf{R} \sim \mathcal{MN}(\mathbf{W}_0, \mathbf{H}, \mathbf{R}).
\end{aligned}$$

- **Mitigates bias like BRSA** [2] by marginalizing **B**.
- **Fewer parameters** (different noise model).
- **More conservative under null.**

**Closed form, unbiased estimator:**
Assuming $\mathbf{X} \neq 0, \mathbf{K} \neq 0, \Omega^{-1} \neq 0$:

$$\begin{aligned}
\widehat{\mathbf{K}} &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top(\frac{1}{v}\mathbf{Y}\Sigma^{-1}\mathbf{Y}^\top - \Omega)\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\
\mathbb{E}[\widehat{\mathbf{K}}] &= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\left(\frac{1}{v}\mathbb{E}[\mathbf{Y}\Sigma^{-1}\mathbf{Y}^\top] - \Omega\right) \\
&\quad \cdot \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\
&= (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\left((\Omega + \mathbf{X}\mathbf{K}\mathbf{X}^\top)\frac{1}{v}\mathrm{Tr}[\Sigma^{-1}\Sigma] - \Omega\right) \\
&\quad \cdot \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1} \\
&= \mathbf{K}
\end{aligned}$$

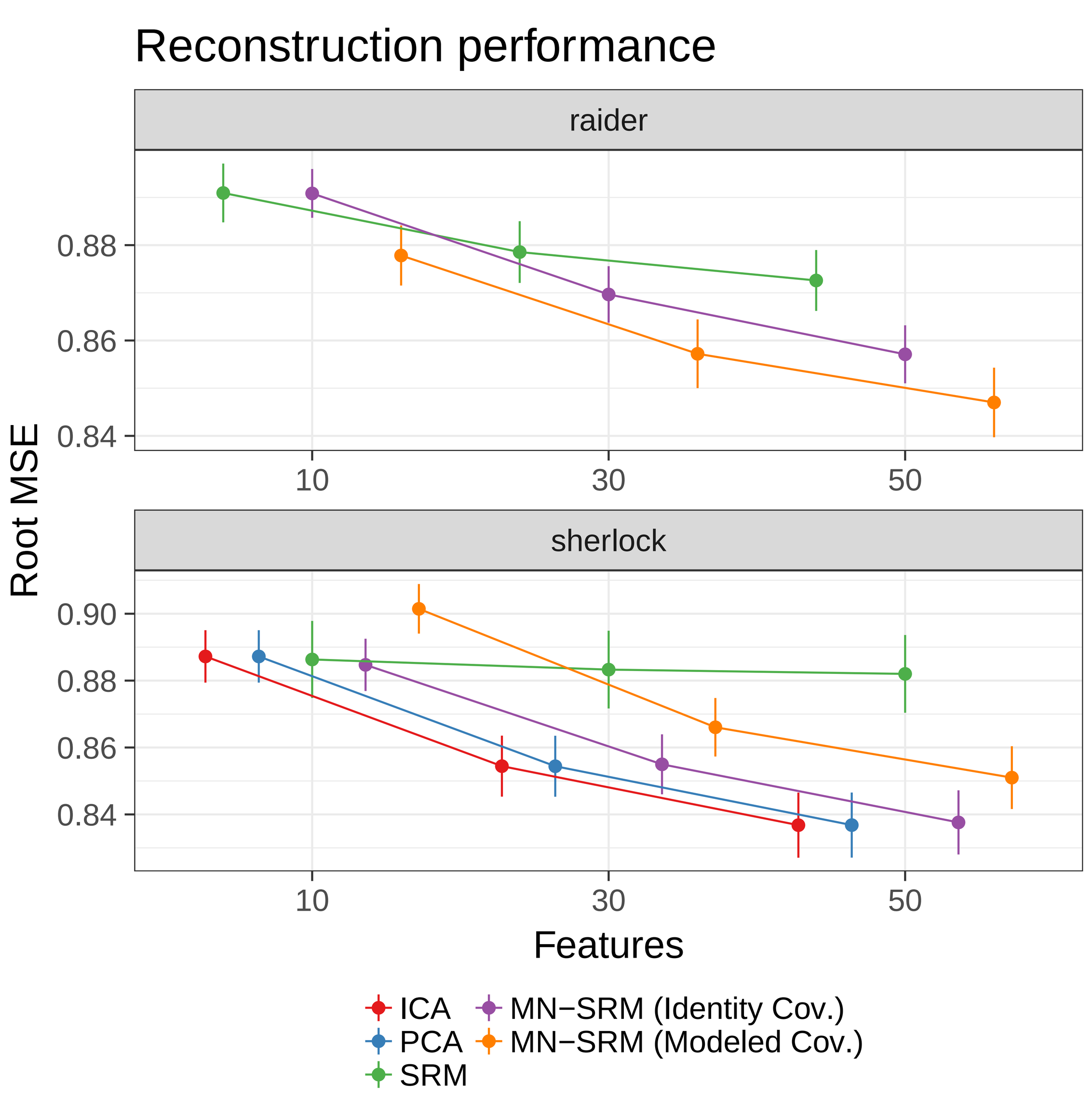


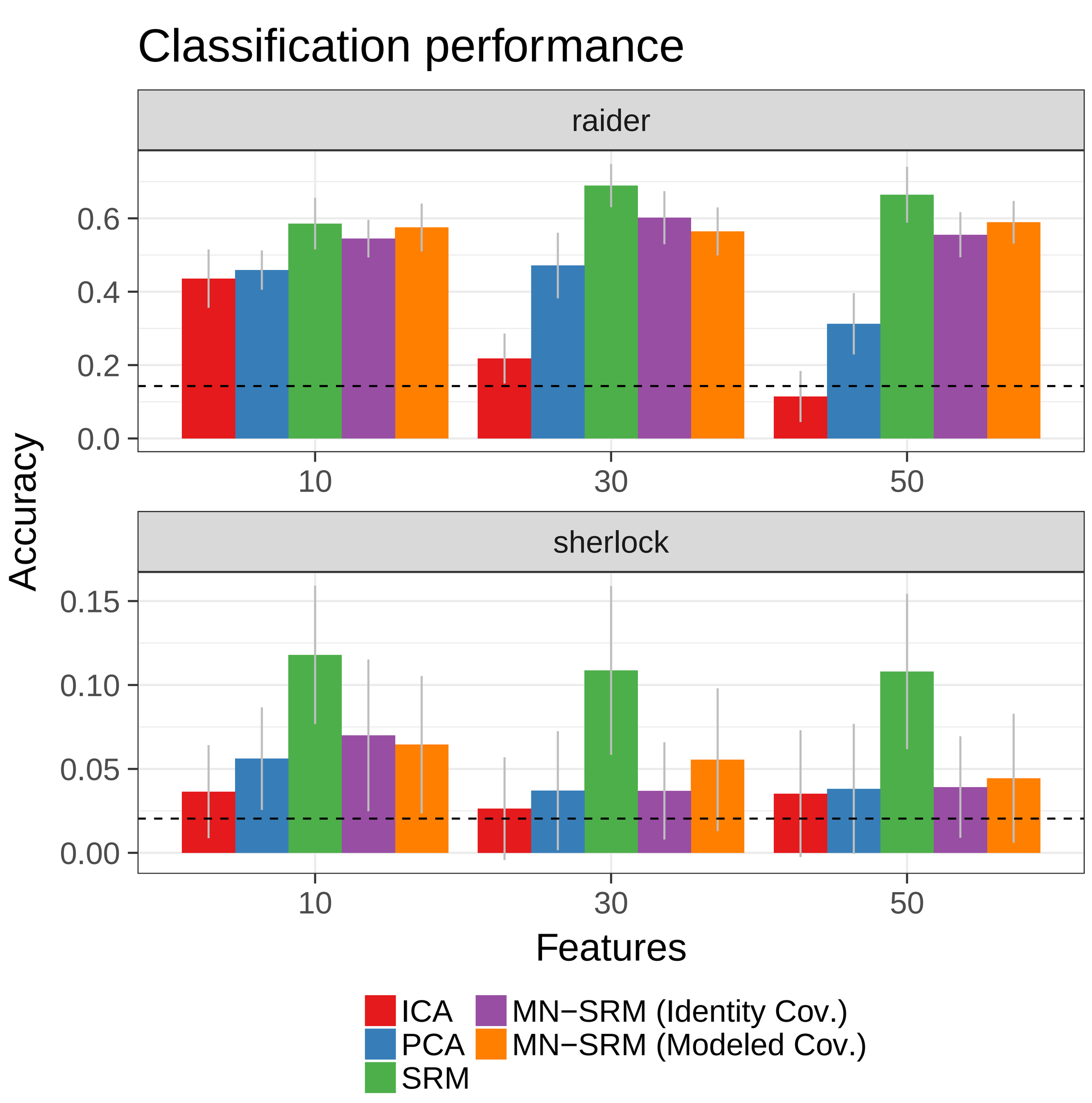




## MN-SRM: improved reconstruction, fewer parameters

- **ECM algorithm for fast estimation.**
- **Fewer parameters than original SRM** [3] by marginalizing $\mathbf{F}_i$ instead of $\mathbf{Z}$.
- **Better out-of-sample reconstruction** (but worse feature selection).

$$\begin{aligned}
&\mathbf{Y}_i \mid \mathbf{F}_i, \mathbf{B}_i, \mathbf{Z}, \mathbf{W}_i, \Sigma_i, \Omega \sim \\
&\quad \mathcal{MN}(\mathbf{F}_i\mathbf{Z} + \mathbf{B}_i\mathbf{X} + \mathbf{J}\mathbf{W}, \rho_i^2\Sigma, \Omega) \\
&\mathbf{F}_i \mid \mathbf{C}, \mathbf{U} \sim \mathcal{MN}(0, \mathbf{C}, \mathbf{U}) \\
&\mathbf{Z} \mid \mathbf{D}, \mathbf{V} \sim \mathcal{MN}(0, \mathbf{D}, \mathbf{V}) \\
&\mathbf{B}_i \mid \mathbf{G}, \mathbf{K} \sim \mathcal{MN}(\beta_0, \mathbf{G}, \mathbf{K}) \\
&\mathbf{W}_i \mid \mathbf{H}, \mathbf{R} \sim \mathcal{MN}(\mathbf{W}_0, \mathbf{H}, \mathbf{R}).
\end{aligned}$$




## Paper and code

**Paper (AISTATS 2018):** `http://arxiv.org/abs/1711.03058`
**Code:** Under review for inclusion in BrainIAK (`brainiak.org`).

## References

[1] Shvartsman, M, Sundaram, N, Aoi, M. C., Charles, A. S., Wilke, T. C., & Cohen J. D. AISTATS 2018; [2] Cai, M. B., Schuck, N. W., Pillow, J. W., & Niv, Y. NIPS 2016; [3] Chen, P.-H., Chen, J., Yeshurun, Y., Hasson, U., Haxby, J., & Ramadge, P. J. NIPS 2015; [4] Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., & Hasson, U. Nat. Comms. 7:12141 (2016).