On the move: Dynamical systems for modeling, measurement and inference in compressed sensing

Christopher J. Rozell

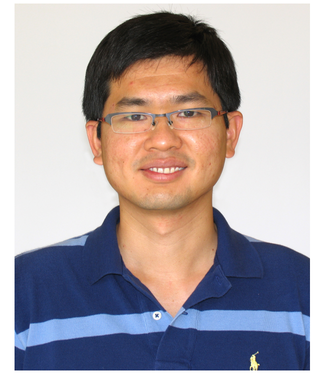
Georgia Institute of Technology

Georgia Tech College of Engineering

CSIP Center for Signal & Information Processing

Acknowledgments

- Aurèle Balavoine
- Rich Baraniuk
- Adam Charles
- Armin Eftekhari
- Pierre Garrigues
- Jennifer Hasler
- Don Johnson
- Bruno Olshausen
- Justin Romberg
- Sam Shapero
- Mike Wakin
- Han Lun Yap
- Mengchen Zhu

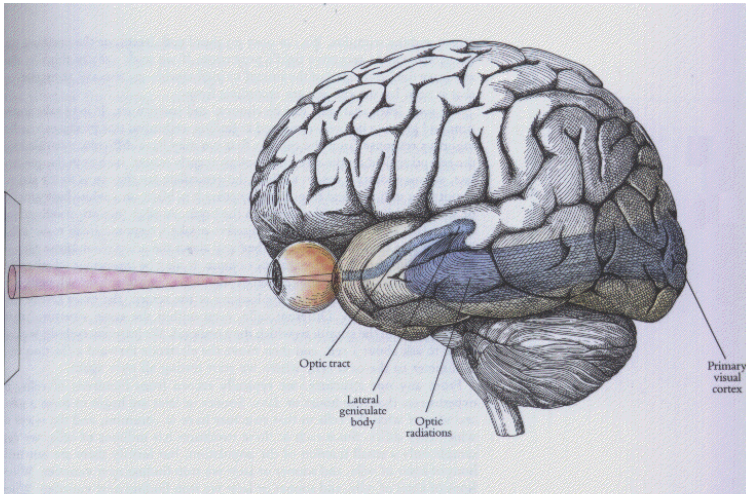


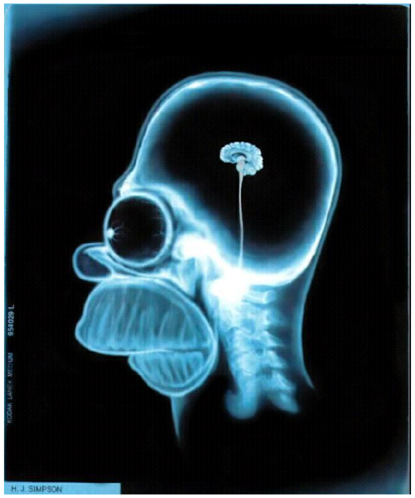
"I not only use all the brains I have, but all I can borrow."

-Woodrow Wilson

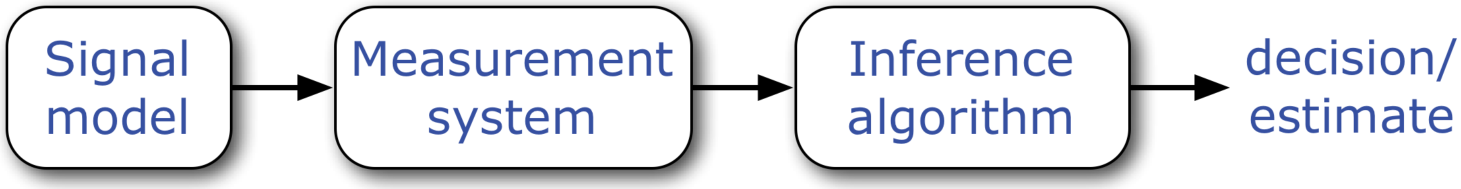
Today's plan: dynamics in the pipeline

Signal model → Measurement system → Inference algorithm → decision/ estimate



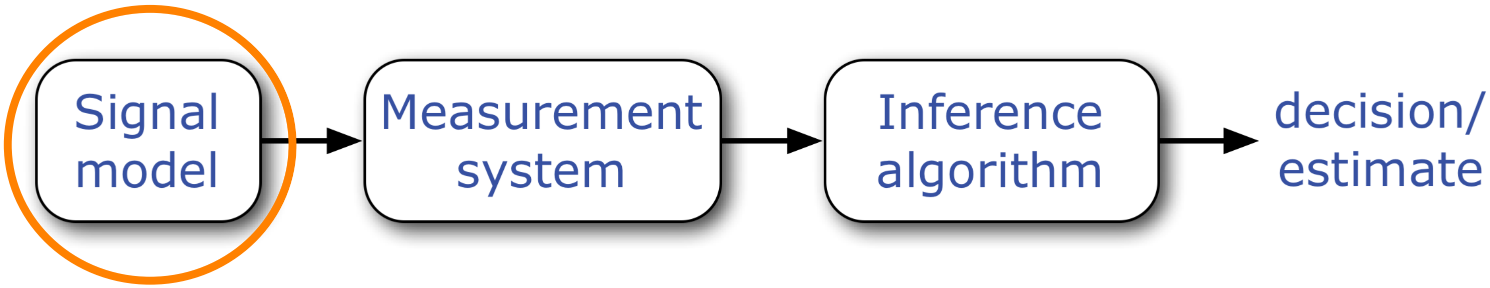


Today's plan: dynamics in the pipeline



- Dynamic sparse signal models
  - Stochastic filtering for sparse signals
- CS for structured systems
  - Short term memory in networks
- Inference using dynamical systems
  - Neuromorphic implementations
- Measurement of dynamical system attractors

Today's plan: dynamics in the pipeline



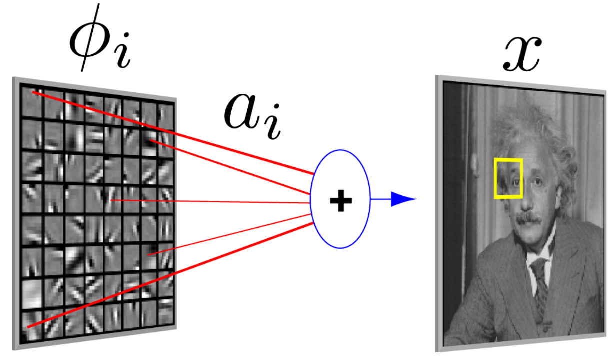
- Dynamic sparse signal models
  - Stochastic filtering for sparse signals
- CS for structured systems
  - Short term memory in networks
- Inference using dynamical systems
  - Neuromorphic implementations
- Measurement of dynamical system attractors

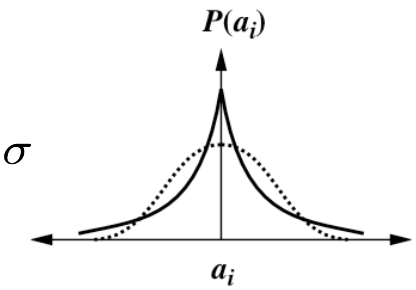
Static sparsity model

- Linear generative model:

Image Dictionary Coefficients

$$x = \sum_i \phi_i a_i + w$$





- Causes are iid and sparse: $p(a_i) \propto e^{-|a_i|\sqrt{2}/\sigma}$
- Noise is Gaussian: $p(x|a) \propto e^{-\|x-\Phi a\|_2^2/2\sigma_w^2}$
- Infer $\{a_i\}$ via MAP estimate called BPDN:

$$\widehat{a} = \arg\max_a p(a|x) = \arg\min_a \left[\frac{1}{2}\|x - \Phi a\|_2^2 + \lambda \sum_i |a_i|\right]$$

- Regularization parameter is inverse SNR: $\lambda \propto \frac{\sigma_w^2}{\sigma}$

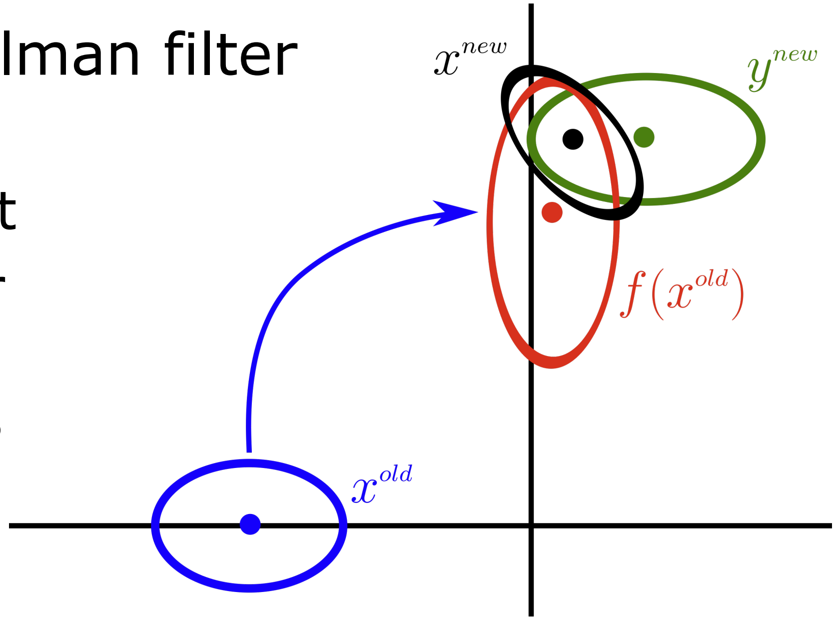
Dynamic signal estimation

State → $$x[n] = f(x[n-1]) + \nu[n]$$ ← Innovations

Observation → $$y[n] = g(x[n]) + \epsilon[n]$$ ← Measurement noise

- Linear/Gaussian model -> Kalman filter
  - Prior from prediction
  - Likelihood from measurement
  - x_{new} estimated from posterior

- Sparse states or innovations?



Challenges and inspiration

- Many approaches to sparse Kalman filter [Vaswani 2008,2010; Carmi et al. 2010; Zachariah et al. 2012; Ziniel et al. 2010; Charles, Asif, Romberg, & R. 2011; Sejdinovic et al. 2010]
- Existing methods lack combination of robustness, generality, and efficiency

- Idea from static model: re-weighted l1 (RWL1)
  - Gamma hyperprior on variances λ_i
  - EM algorithm yields iterative re-weighted l1

$$\lambda_i = \frac{\tau}{|x_i| + \eta}$$

$$\arg\min_x \|y - g(x)\|_2^2 + \sum_i \lambda_i |x_i|$$

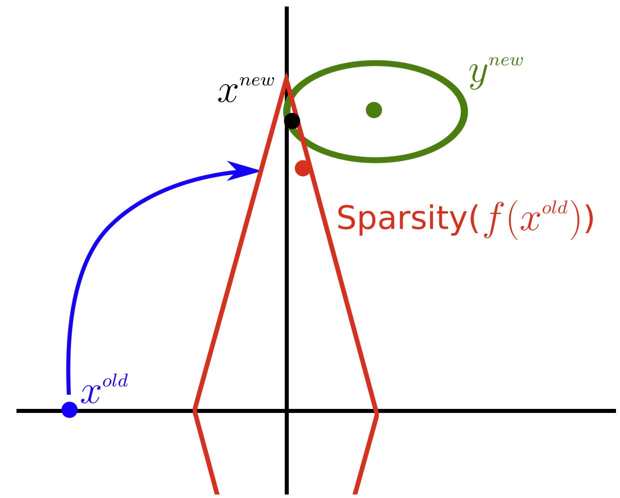
(Candès, et al. 2008; Garrigues & Olshausen 2010)

RWL1-DF algorithm

- Main idea: RWL1 with variances from model prediction

$$x_i^{new} | \lambda_i^{new} \sim \text{Laplacian}(0, \lambda_i^{new})$$

$$\lambda_i^{new} \sim \text{Gamma with } \mathbb{E}(\lambda_i^{new}) = \frac{1}{[f(x^{old})]_i}$$



- EM inference -> RWL1-DF:

$$\lambda_i^{new} = \frac{2\tau}{|x_i^{new}| + |[f(x^{old})]_i| + \eta}$$

$$x^{new} = \arg\min_x \|y^{new} - g(x)\|_2^2 + \lambda_0 \sum_i \lambda_i^{new} |x_i|$$

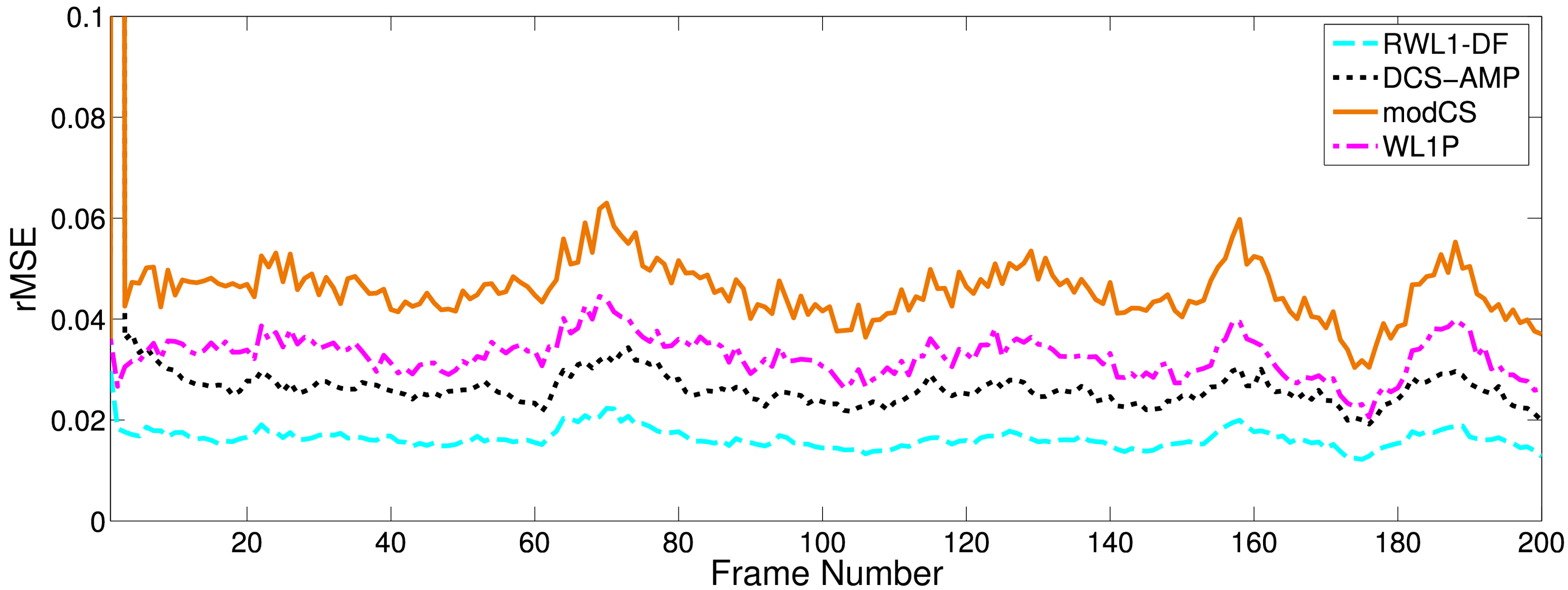
(Charles & R. 2013)

Video data

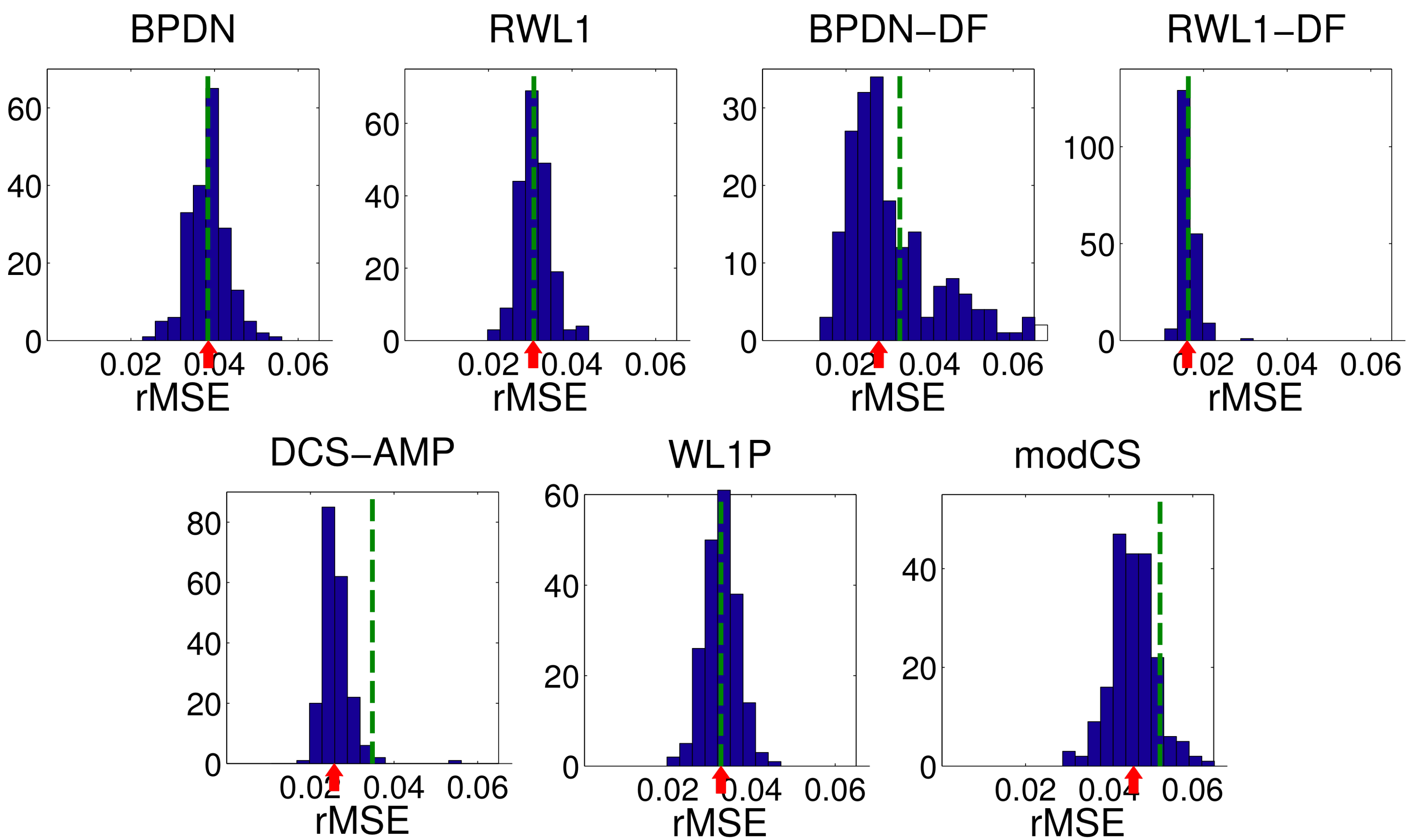
- Standard Foreman test video sequence (128x128)
- Measurements: compressed sensing setup with $M=0.25*128^2$ measurements
- Assume states are sparse wavelet (synthesis) coefficients with $f(x)=x$

- Methods based on standard KF not possible due to matrix inverses over large state space

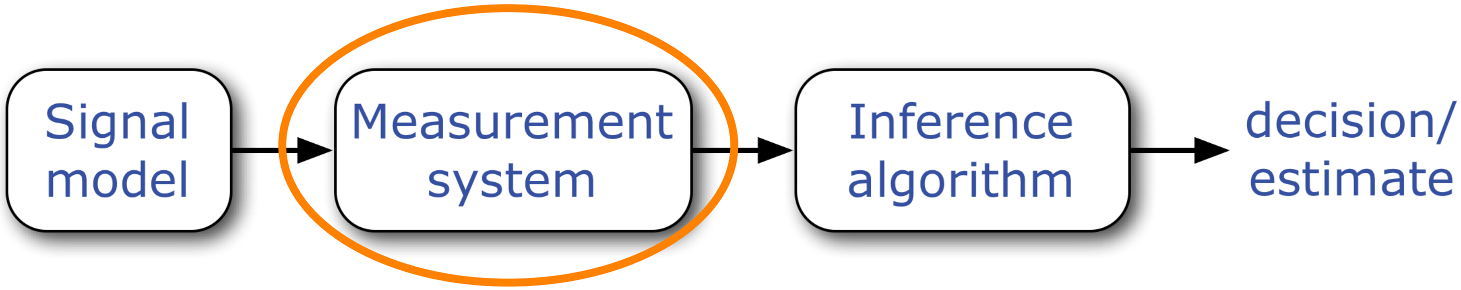
Lower steady-state recovery error



Lower steady-state recovery error



Today's plan: dynamics in the pipeline

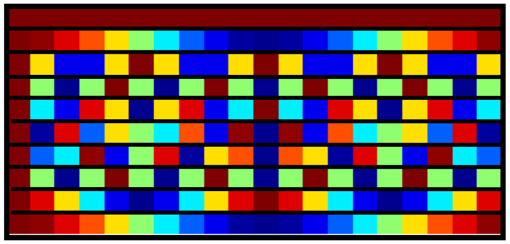


- Dynamic sparse signal models
  - Stochastic filtering for sparse signals
- CS for structured systems
  - Short term memory in networks
- Inference using dynamical systems
  - Neuromorphic implementations
- Measurement of dynamical system attractors

Structured Matrices in CS

- Subsampled Fourier matrices

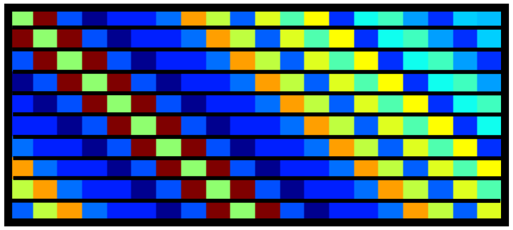
$$\text{RIP-}(S,\delta) \;\Leftrightarrow\; M \geq O\left(\frac{S}{\delta^2}\log^4(N)\right)$$



(Rudelson and Vershynin, 2008)

- Partial circulant matrices (with random probe)

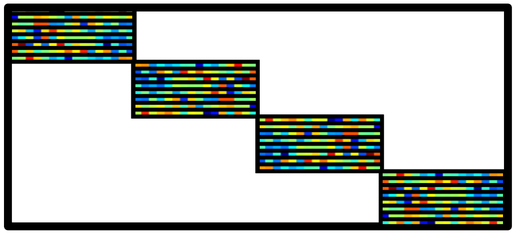
$$\text{RIP} - (S,\delta) \Leftrightarrow M \geq O\left(\frac{S}{\delta^2}\log^4(N)\right)$$



(Krahmer et al., 2012)

- Block diagonal matrices

$$\text{RIP} - (S,\delta) \Leftrightarrow M \geq O\left(\frac{S}{\delta^2}\mu^2\log^6(N)\right)$$



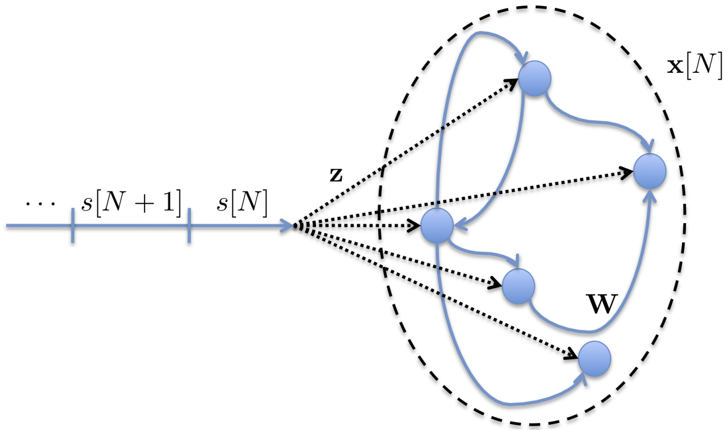
(Yap, Eftekhari, Wakin, & R., 2011)

- Any RIP matrix can produce stable embeddings of manifolds (Yap, Wakin, & R., 2013; Krahmer & Ward 2011; Baraniuk & Wakin 2009)

Sensing with a network

- Exogenous time series *s*[*n*] drives a network of *M* nodes

$$\boldsymbol{x}[n] = f\left(\boldsymbol{W}\boldsymbol{x}[n-1] + \boldsymbol{z}s[n] + \widetilde{\boldsymbol{\epsilon}}[n]\right)$$



(Maass, et al. 2002; Jaeger & Haas 2004; Jaeger 2001; White et al. 2004; Ganguli et al. 2008)

- Can *M* nodes recover a signal of length *N*>*M*?
- What if inputs *s*[*n*] are sparse in basis Ψ?

$$\boldsymbol{x}[N] = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{\epsilon} \qquad \boldsymbol{s} = [\boldsymbol{s}[N], \ldots, \boldsymbol{s}[1]]^T$$

$$\boldsymbol{A} = \left[\boldsymbol{z} \;\middle|\; \boldsymbol{W}\boldsymbol{z} \;\middle|\; \boldsymbol{W}^2\boldsymbol{z} \;\middle|\; \ldots \;\middle|\; \boldsymbol{W}^{N-1}\boldsymbol{z}\right]$$

(Ganguli et al. 2010)

Memory capacity of finite length inputs

- Choose a construction for the network
  - Random orthogonal connectivity matrix: $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{-1}$
    - Eigenvalues $\boldsymbol{d}_m = e^{jt_m}$ drawn iid from unit circle
    - Inputs weights: $\boldsymbol{z} = \frac{1}{\sqrt{M}}\boldsymbol{U}\boldsymbol{1}_M$
    - Decompose:

$$A \propto UF \quad \text{where} \quad F = [d^0|d|d^2|\ldots|d^{N-1}]$$

  - For S-sparse signal in basis Ψ, δ, and failure prob *η*, if:

$$M \geq C\frac{S}{\delta^2}\mu^2(\boldsymbol{\Psi})\log^4(N)\log\left(\eta^{-1}\right)$$

$$\mu(\boldsymbol{\Psi}) = \max_{n=1,\ldots,N}\sup_{t\in[0,2\pi]}\left|\sum_{m=0}^{N-1}\boldsymbol{\Psi}_{m,n}e^{-jtm}\right|$$

  - Then with probability exceeding (1-*η*), RIP:

$$(1-\delta) \leq \|As\|_2^2/\|s\|_2^2 \leq (1+\delta)$$

(Charles, Yap, & R., 2012)

Proof sketch

- Since *F* is Vandermonde, proof follows very closely from the proof of RIP for subsampled DTFT matrices

  (Rauhut, 2010)

- Proof sketch:
  - Express RIP conditioning as a random variable

$$\delta_S = \sup_{s:\|s\|_0=S} \left| \frac{\|As\|_2^2}{\|s\|_2^2} - 1 \right|$$

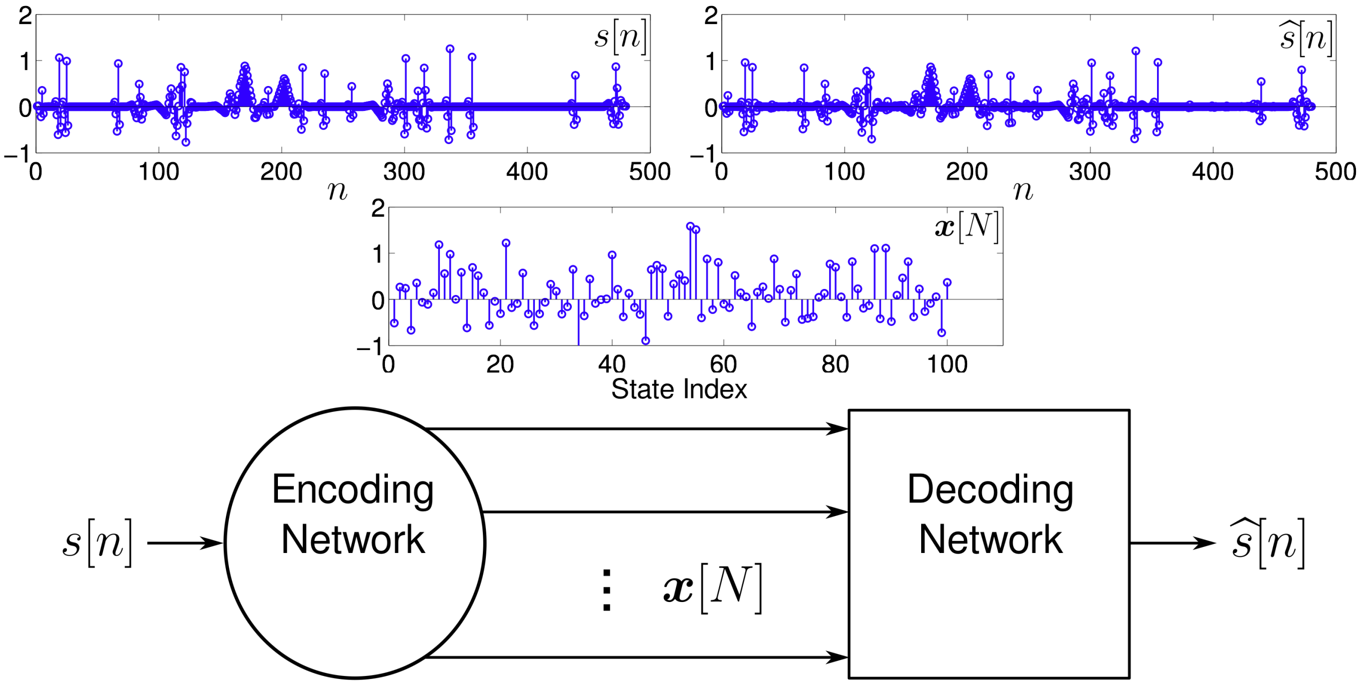
  - Bound moments $E((\delta_S)^p)$ using recent results for bounding expected supremum of random processes

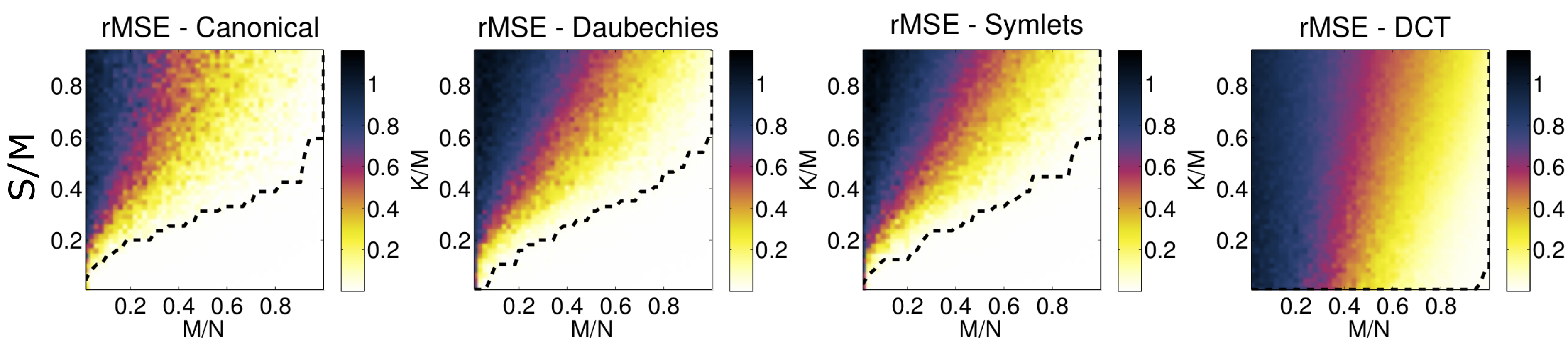
    (Rudelson and Vershynin, 2008)

  - Use moment bounds to get tail bounds characterizing the RIP failure probability

- Extensions to multiple inputs (sparse or low rank)

Empirical recovery

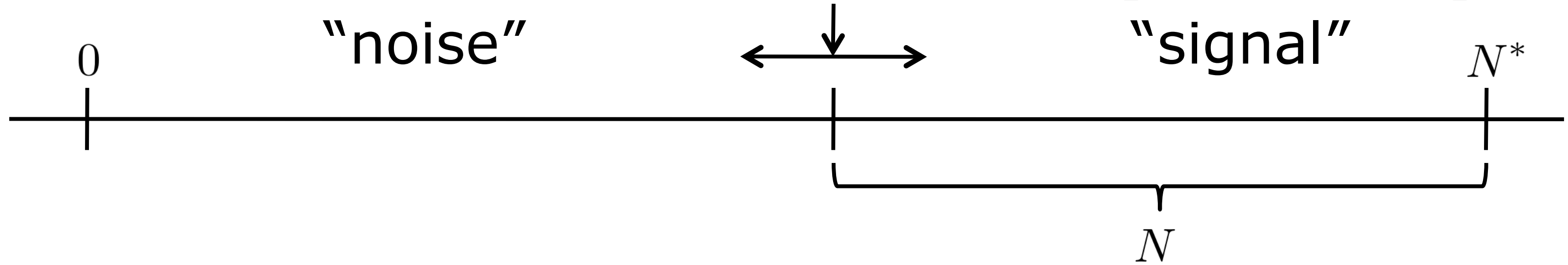




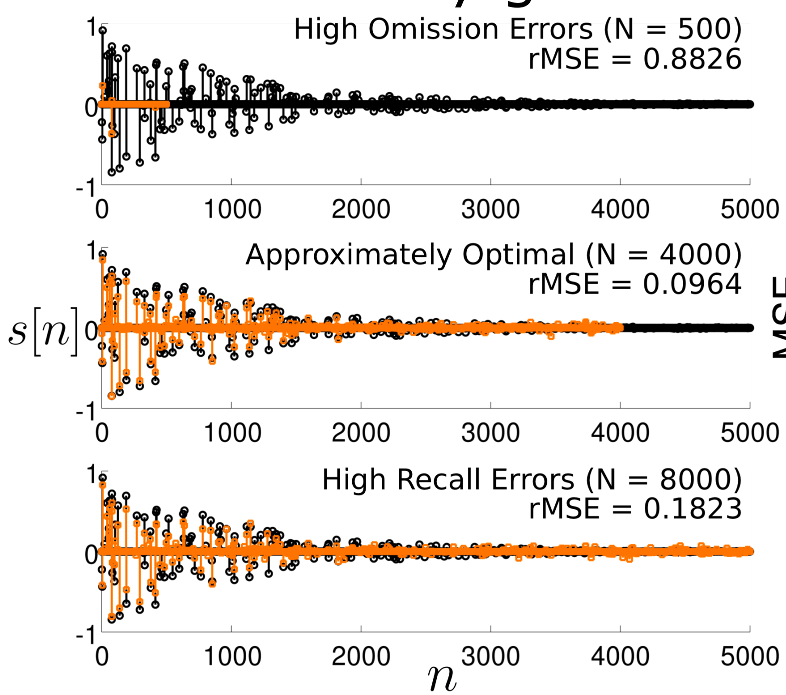
One man's signal is another man's noise

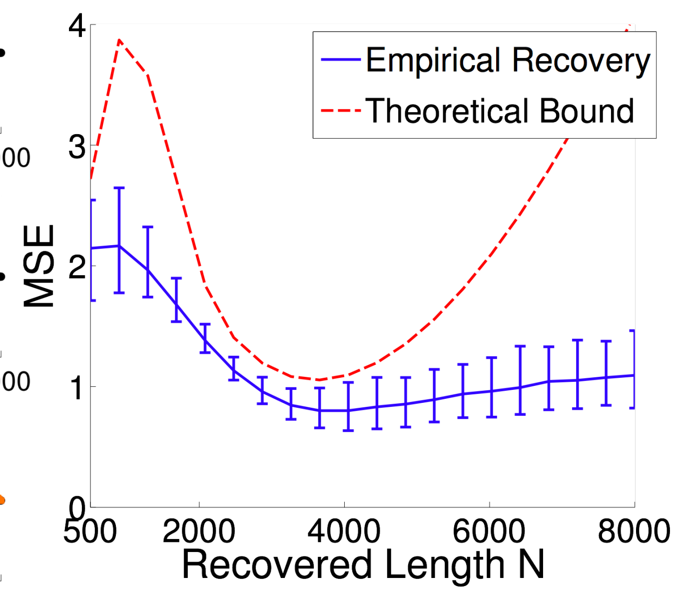
$$\epsilon = \left\| \sum_{n=0}^{N^*-N} \boldsymbol{W}^n \boldsymbol{z} s(n) \right\|_2 \qquad \boldsymbol{s} = \begin{bmatrix} s(N^*) \\ \vdots \\ s(N^* - N + 1) \end{bmatrix}$$

Choose this

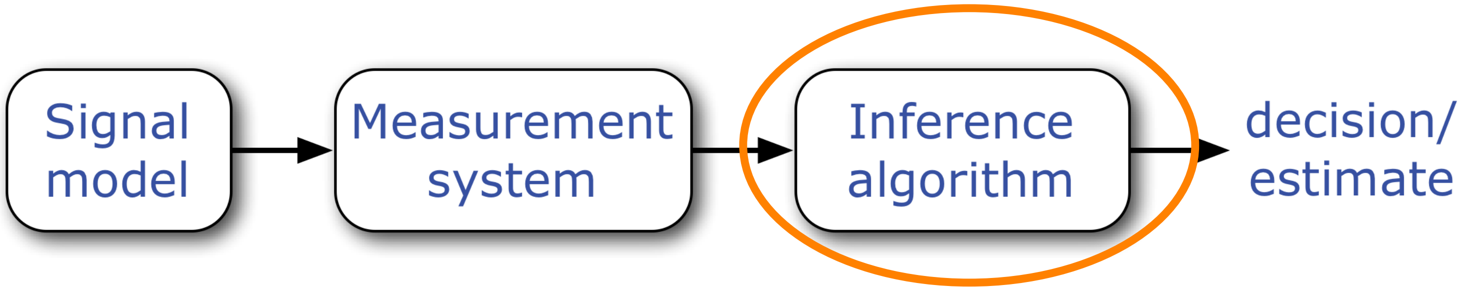


- Use RIP recovery guarantees to bound recovery error





Today's plan: dynamics in the pipeline



- Dynamic sparse signal models
  - Stochastic filtering for sparse signals
- CS for structured systems
  - Short term memory in networks
- Inference using dynamical systems
  - Neuromorphic implementations
- Measurement of dynamical system attractors

Inference in dynamical systems

- Many algorithms for using computers to solve

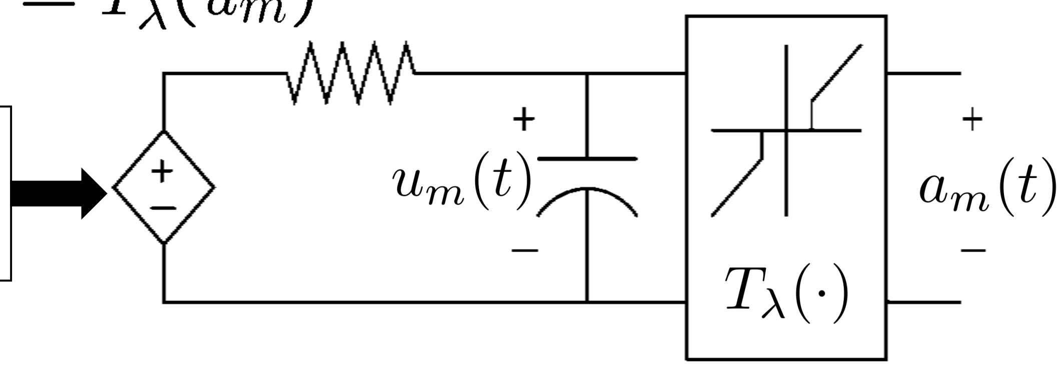
$$\{a_m\} = \arg\min_a \frac{1}{2}\left\| x - \sum_m \phi_m a_m \right\|_2^2 + \lambda \sum_m C(a_m)$$

- Can a dynamical system compute sparse codes?
- Locally competitive algorithms (LCA)

$$\dot{u}_m(t) = \frac{1}{\tau}\left[\langle \boldsymbol{\phi}_m, \boldsymbol{x}(t)\rangle - u_m(t) - \sum_{n \neq m} \langle \boldsymbol{\phi}_m, \boldsymbol{\phi}_n\rangle a_n(t)\right]$$

$$a_m = T_\lambda(u_m)$$

$$\boxed{\langle \boldsymbol{\phi}_m, \boldsymbol{x}(t)\rangle - \sum_{n \neq m} \langle \boldsymbol{\phi}_m, \boldsymbol{\phi}_n\rangle a_n(t)}$$



(R., Johnson, Baraniuk & Olshausen, 2008)

Sparse approximation with LCAs

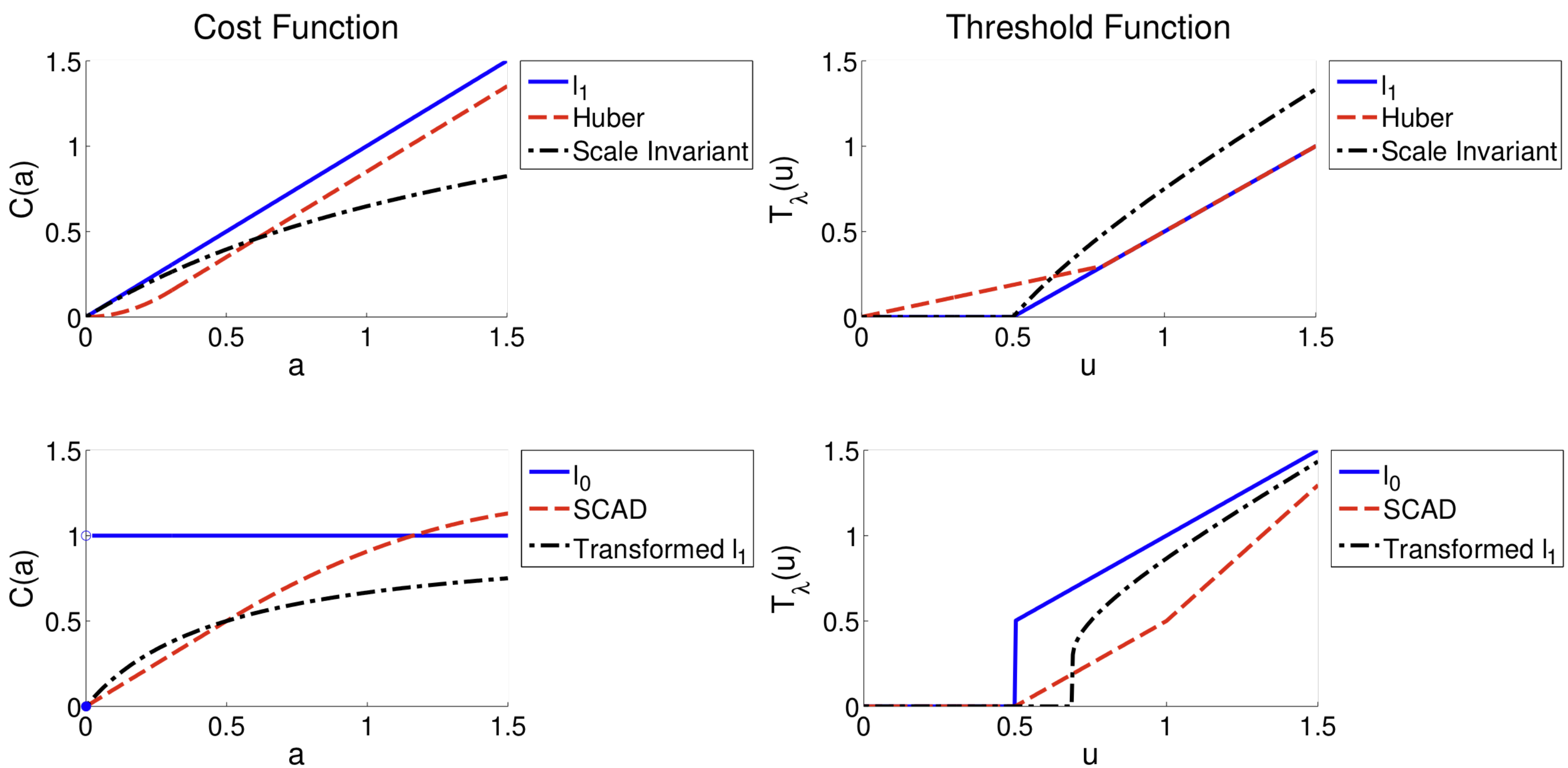
- System descends via warped gradient descent:

$$\dot{u}_m \propto -\frac{\partial E}{\partial a_m} \quad \text{with} \quad u_m = T_\lambda^{-1}(a_m) = a_m + \lambda \frac{\partial C(a_m)}{\partial a_m}$$

- With some assumptions on the non-linear function:
  - Is globally asymptotically stable if E is strictly convex
  - Converges to fixed point even with connected solutions
  - Converges exponentially fast: MSE $\leq ke^{-ct}$
- In CS recovery, can establish stronger bounds
  - No extraneous coefficients in support if $M=O(K^2 \log N)$
  - Strong bounds on convergence rate if $M=O(K \log N)$

(Balavoine, Romberg & R., 2012;
Balavoine, R. & Romberg, 2013a,b)

Many sparse cost functions

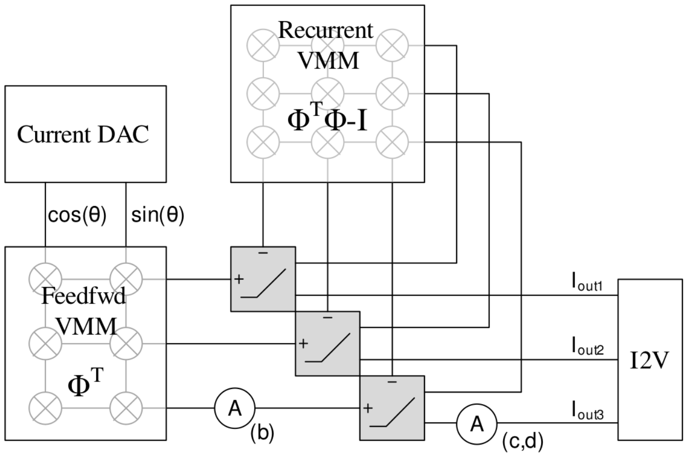


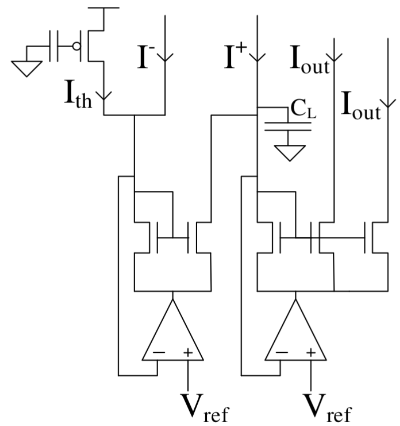
- Also RWL1 and block L1 with non-overlapping blocks

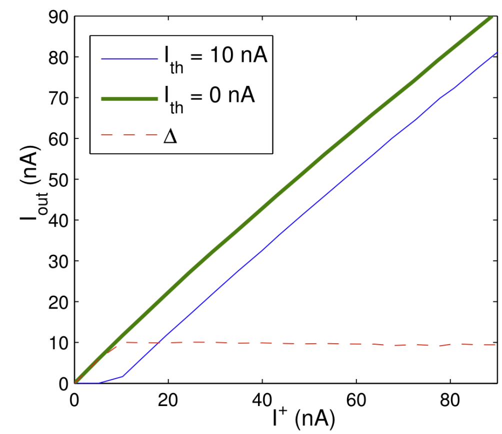
(Charles, Garrigues & R., 2012)

Implementation in analog VLSI

- Implementation on reconfigurable analog hardware





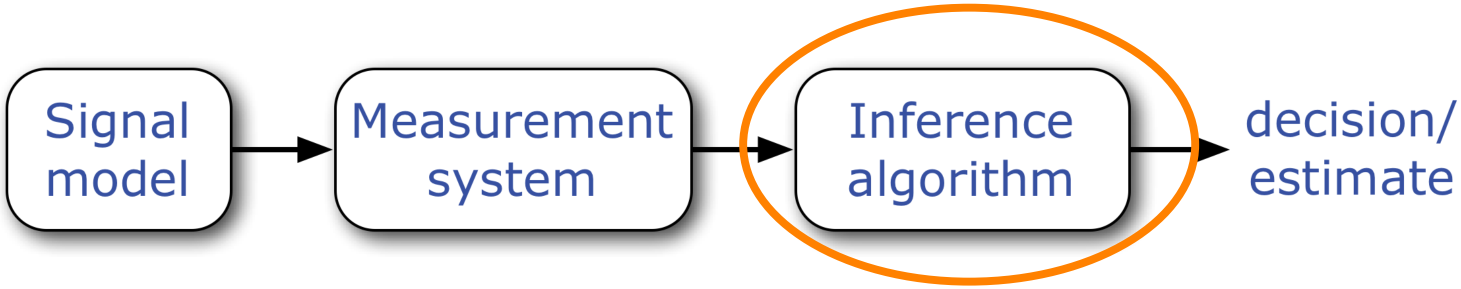


- Sublinear scaling of convergence time with *N*

System	12×18 Spiking LCA	666×1k Spiking LCA (Hypothetical)	666×1k Analog LCA (Hypoth.)[4]	1k CPU [3]
Power (Active)	1.34mW	7.68mW	149mW	≈3.8W
(Total)	3.02mW	9.79mW	151mW	≈100W
Time (Converge)	25μs	≈25μs	≈240μs	46ms
Time (Total)	1.03ms	1.03ms	4.62ms	46ms
Error (RMS)	4.8% (@ K=3)	≈ 4.8%	≈ 5%	-
Extra Cost (Avg)	1.7% (@ K=3)	≈ 1.7%	≈ 1%	-

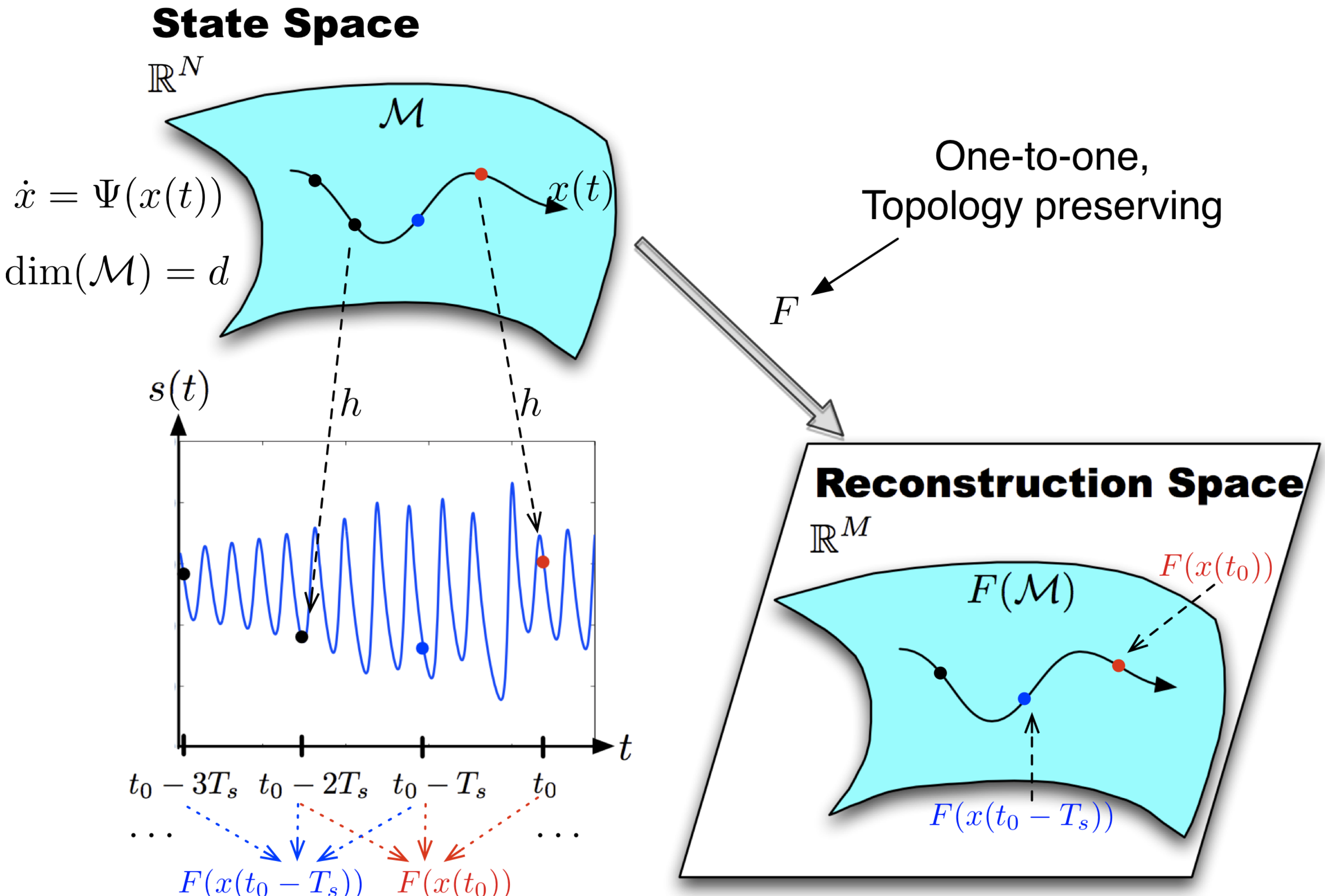
(Shapero,Charles, R. &Hasler 2012; Shapero, R. &Hasler 2012,2013; Shapero,Zhu, Hasler &R. 2013)

Today's plan: dynamics in the pipeline



- Dynamic sparse signal models
  - Stochastic filtering for sparse signals
- CS for structured systems
  - Short term memory in networks
- Inference using dynamical systems
  - Neuromorphic implementations
- Measurement of dynamical system attractors

Takens' Embedding Theorem

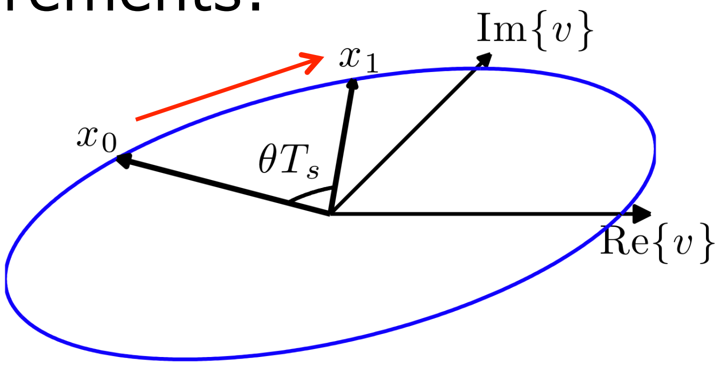


(Takens 1981; Sauer et al. 1991)

Stable Takens' Embeddings?

- RIP works because pairwise distances are stable
- Stable embedding extended to manifolds
  - Unlike typical CS, get one measurement M times
- Linear system and linear measurements:

$$\dot{x} = \Psi x$$



Dimension: *d* Speed: *v*

- If $M > 2(2d-1)v\varepsilon^{-1}$, then Takens' embedding is stable with conditioning $\delta_0+\varepsilon$

$$(1-(\delta_0+\epsilon)) \leq \frac{\|F(x)-F(y)\|_2^2}{\|x-y\|_2^2} \leq (1+(\delta_0+\epsilon)) \qquad x,y \in \mathcal{M}$$

(Yap & R., 2011)

Observations

- M doesn't depend directly on N
- Possible that $M>N$
  - Would be crazy in standard CS, but reasonable here
- Plateau in conditioning: limit to improvement with M
  - Real effect and not a proof artifact
  - δ_0 depends on system and interaction with measurements
  - Only eliminated for systems that fill state space and measurements that observe them evenly

- Extension can be derived for nonlinear systems

(Yap, Eftekhari, Wakin & R., in preparation)

More information

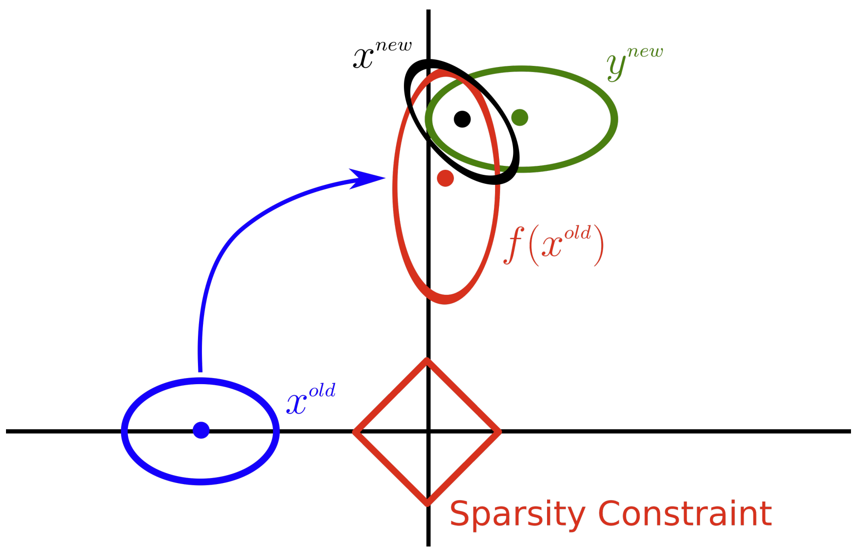
http://users.ece.gatech.edu/~crozell

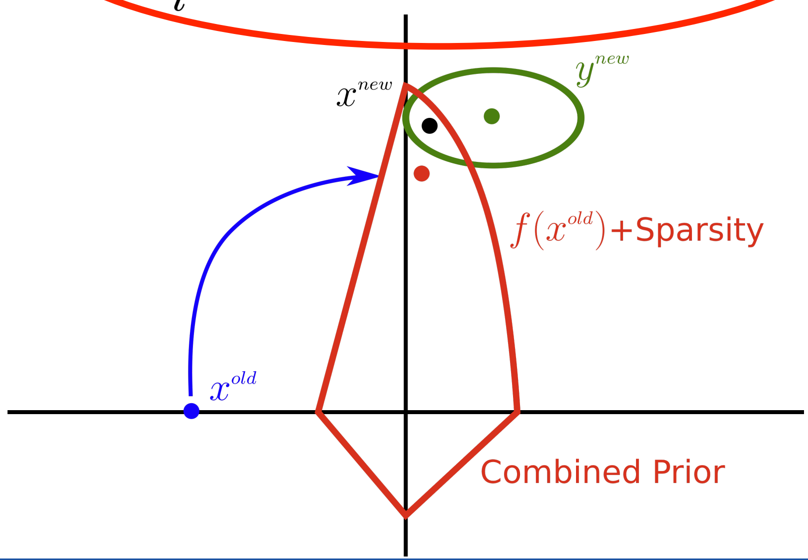
crozell@gatech.edu

Current approaches

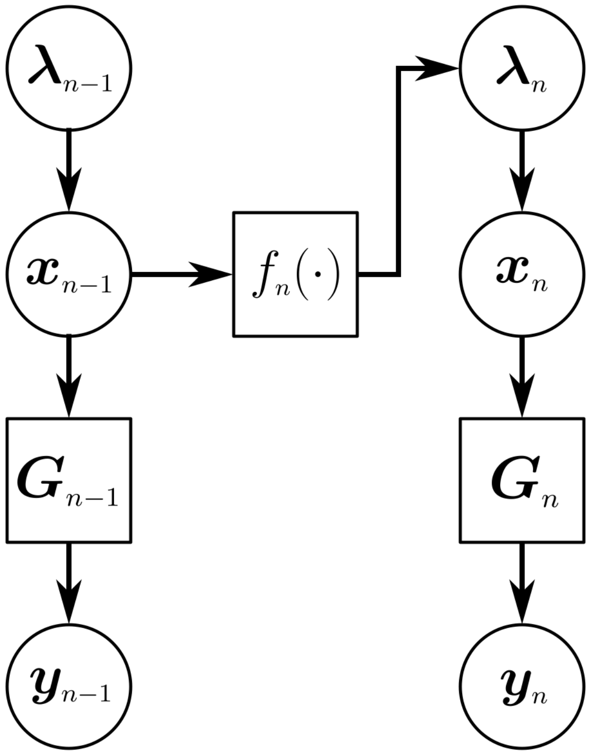
- Modify KF (e.g., restricted support, sparsify solution, propagate covariance) [Vaswani 2008; Carmi, et al. 2010]
- Direct coefficient transition modeling (e.g., MP or modified OMP) [Zachariah et al. 2012; Ziniel, et al. 2010]
- L1 penalty in optimization (BPDN dynamic filtering) [Charles, Asif, Romberg, & R. 2011; Vaswani 2010; Sejdinovic et al. 2010]

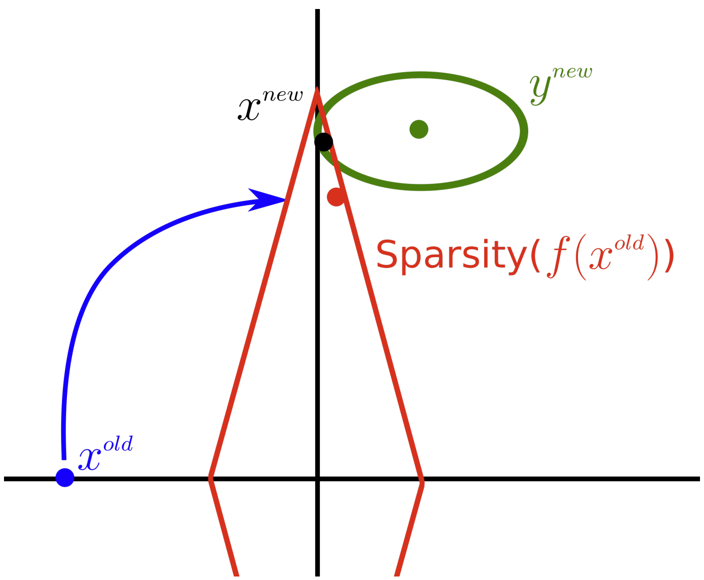
$$x^{new} = \arg\min_x \|y^{new} - g(x)\|_{2,R}^2 + \lambda \sum_i |x_i| + \gamma \|x - f(x^{old})\|_2^2$$





Propagating second order statistics





Compressed Sensing (CS)

- Signal acquisition framework: $y = \Psi x + w$
  - Undersampled: *Ψ* is *M*x*N* with *M*<<*N*

- Sufficient condition: Restricted Isometry Property
  - For all 2*S*-sparse signals *x*, we have RIP(2*S*,δ) if:

$$C(1-\delta) \leq \|\Psi x\|_2^2 / \|x\|_2^2 \leq C(1+\delta)$$

  - Recovery via BPDN:

$$\widehat{a} = \arg\min_a \|a\|_1 \quad \text{such that} \quad \|y - \Psi\Phi a\|_2 \leq \|w\|_2$$

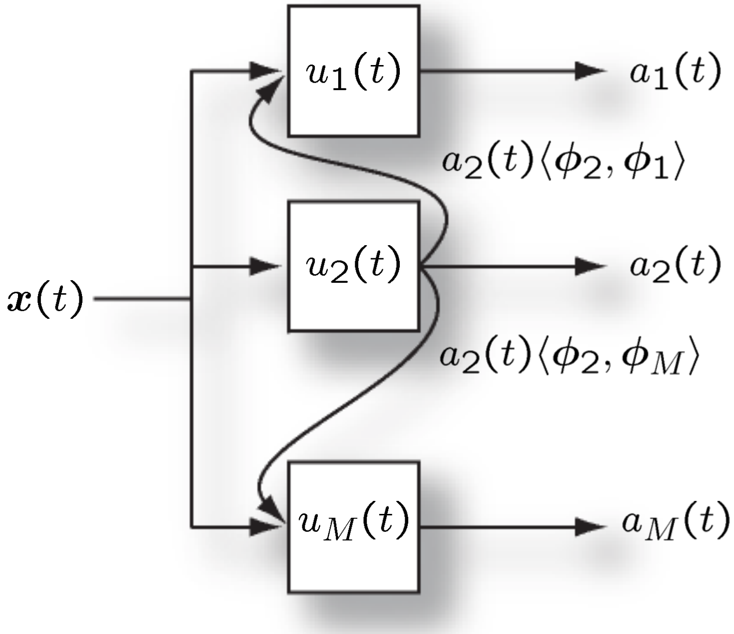
$$\widehat{x} = \Phi\widehat{a}$$

  - Recovery guarantee:

$$\|x - \widehat{x}\|_2 \leq \alpha\|w\|_2 + \beta\frac{\|\Phi^T(x - x_S)\|_1}{\sqrt{S}}$$

[Candès 2006]

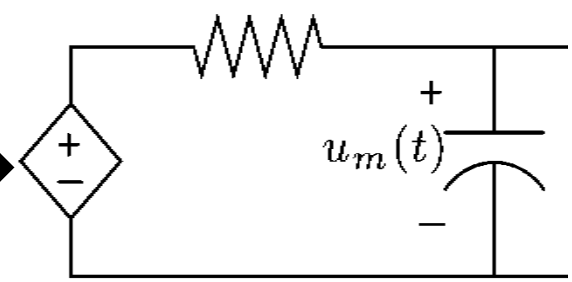
LCA dynamical system architecture



Computational primitives

- Leaky integration
- Nonlinear activation
  $$a_m = T_\lambda(u_m)$$
- Inhibition/excitation

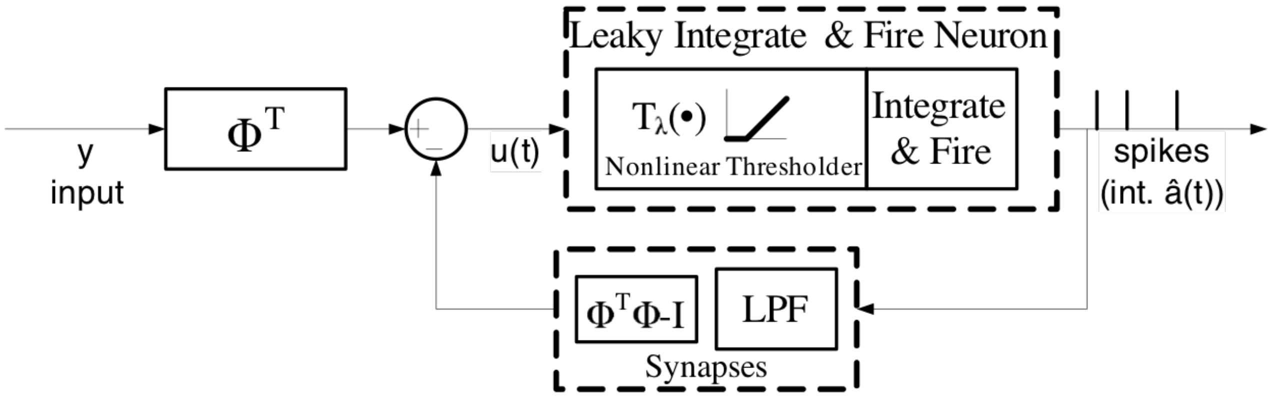
$$\boxed{\langle\phi_m,\boldsymbol{x}(t)\rangle - \sum_{n\neq m}\langle\phi_m,\phi_n\rangle\, a_n(t)}$$



$$\dot{u}_m(t) = \frac{1}{\tau}\Big[\qquad\qquad - u_m(t) \qquad\qquad \Big]$$

Implementation in spiking VLSI

- Implementation in rate coded integrate and fire neurons
  - Same dynamics as analog LCA on expectation of spike
  - Efficiency gains in spike representation due to sparsity



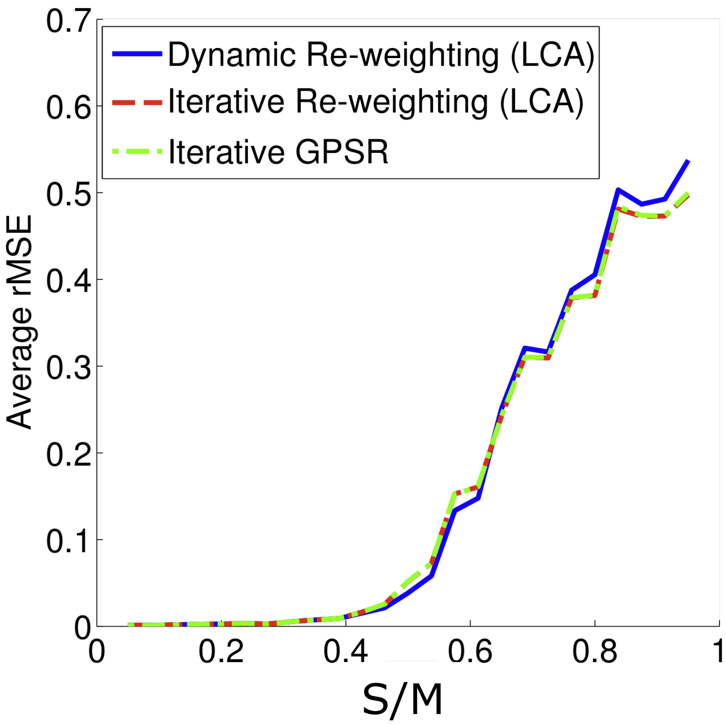
System	12×18 Spiking LCA	666×1k Spik. LCA (Hypoth.)	666×1k Anlg LCA (Hyp.)[18]	1k CPU[29]
Power (Active)	1.34mW	7.68mW	149mW	≈3.8W
(Total)	3.02mW	9.79mW	151mW	≈100W
Time (Converge)	25µs	≈25µs	≈240µs	46ms
Time (Total)	1.03ms	1.03ms	4.62ms	46ms
Error (RMS)	4.8% (@K=3)	≈ 4.8%	≈ 5%	-
Extra Cost (Avg)	1.7% (@K=3)	≈ 1.7%	≈ 1%	-

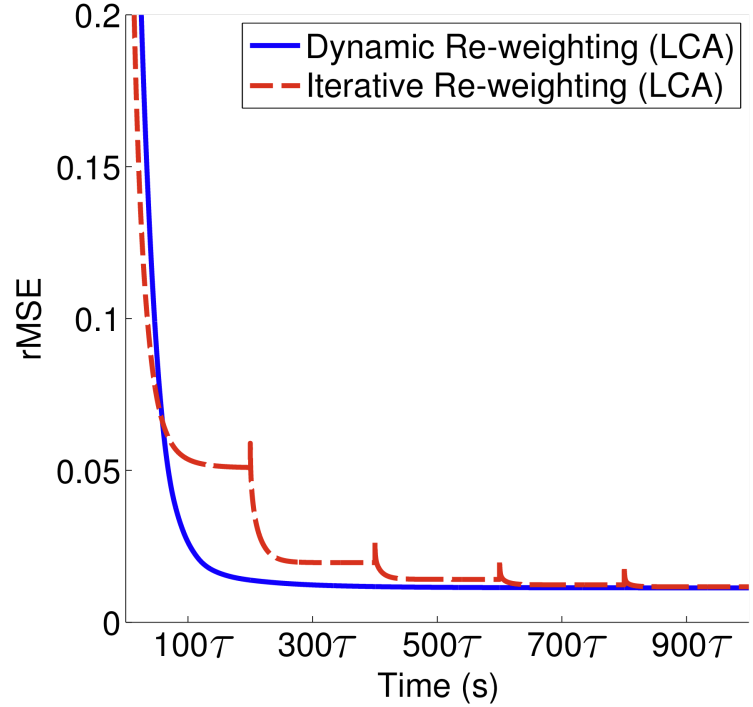
x500

x1800

(Shapero, R. and Hasler 2013; Shapero, Zhu, Hasler and R. 2013b)

RWL1 via dynamic thresholds





Stable Takens' Embedding (Linear Systems)

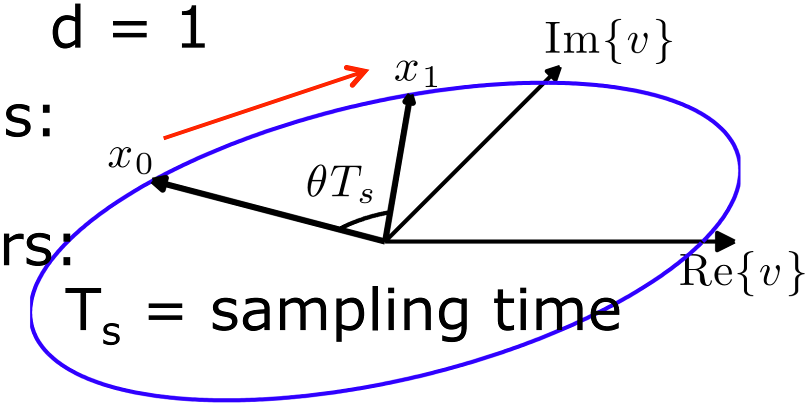
$\boxed{\dot{x} = \Psi x}$

imaginary eigenvalues: $\{\pm j\theta_p\}_{p=1}^{d}$

eigenvectors: $\{v_p, v_p^*\}_{p=1}^{d}$

<u>Example:</u>

d = 1



T_s = sampling time

observation function: $h \in \mathbb{R}^N$

<u>System parameters</u>

A_1, A_2 - attractor shape

κ_1, κ_2 - attractor "observability"

ν " speed of flow

Theorem: (Y. & Rozell, 2011)

Suppose (i) $v_p^H h \neq 0$, (ii) some constraints on T_s. Then, F is a stable embedding of $\mathcal{M}$ with conditioning $\delta(M) := \delta_0 + \delta_1(M)$ where

$$\delta_0 = \frac{A_2\kappa_2^2 - A_1\kappa_1^2}{A_2\kappa_2^2 + A_1\kappa_1^2}, \quad \delta_1(M) = \frac{(2d-1)\nu}{M} \frac{2A_2\kappa_2^2}{A_2\kappa_2^2 + A_1\kappa_1^2}.$$

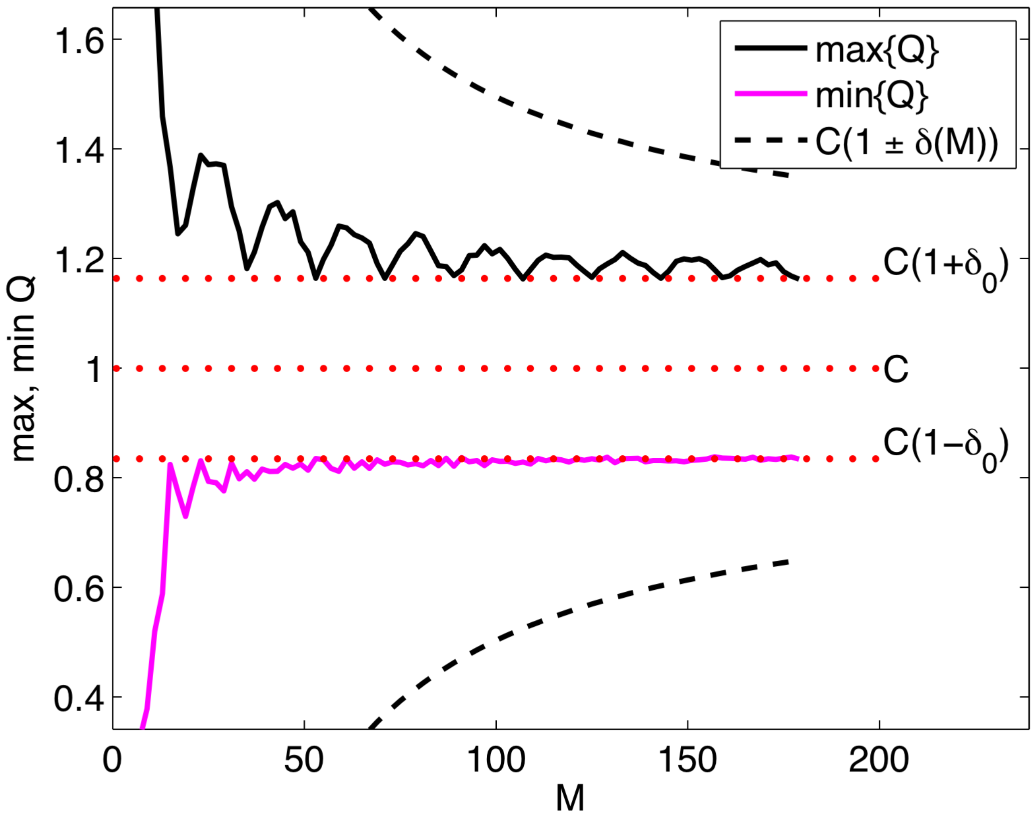
In other words (and after manipulation), if we pick $M \geq \frac{2(2d-1)\nu}{\epsilon}$, then F is a stable embedding with conditioning $\delta_0 + \epsilon$.

1) plateau in conditioning

2) possible that M > N

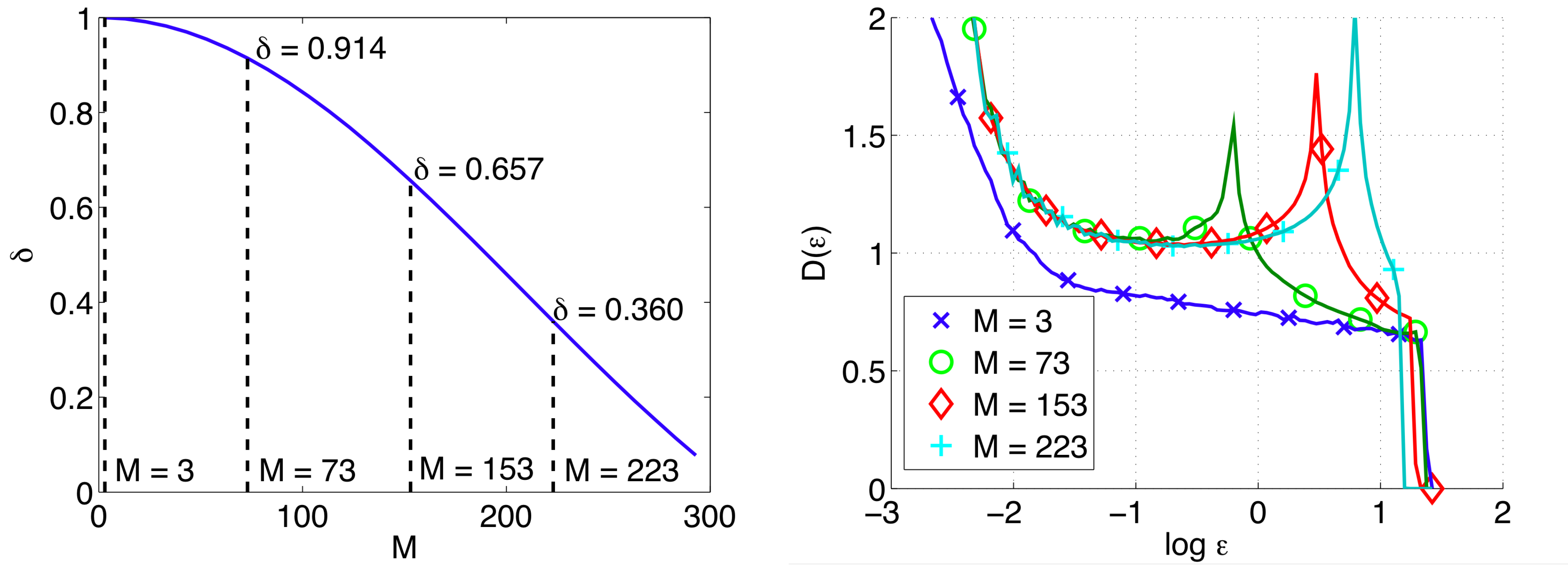
3) Deterministic result

Simulations



- d = 3, $A_1 = A_2$, $\kappa_1 \neq \kappa_2$
- $Q(x,y) := \dfrac{\|F(x) - F(y)\|_2^2}{\|x - y\|_2^2}$

Dimensionality Estimation (GP Algo.)



- d = 1
- Better conditioning implies:
  - improved dimension estimation
  - increased width of plateau