ARL Visit July 8, 2015

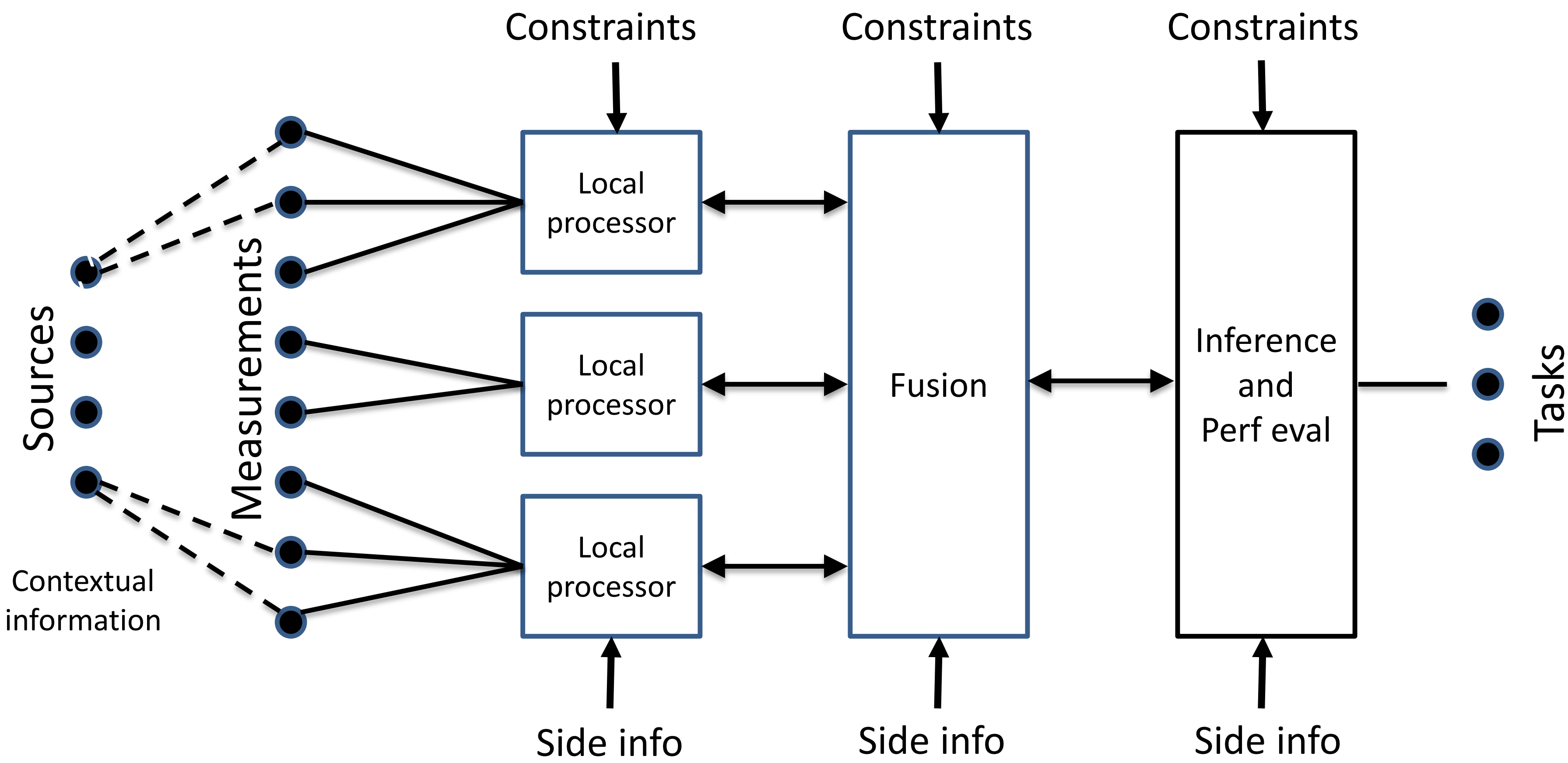
Value-centered Information Theory for Adaptive Learning, Inference, Tracking, and Exploitation

[http://wiki.eecs.umich.edu/voimuri]

ARO W911NF-11-1-0391
Program manager: Liyi Dai

Investigators: Al Hero (PI), Raj Nadakuditi, John Fisher, Jon How, Alan Willsky, Randy Moses, Emre Ertin, Angela Yu, Michael Jordan, Stefano Soatto, Doug Cochran

Learning, inference, tracking, and exploitation system

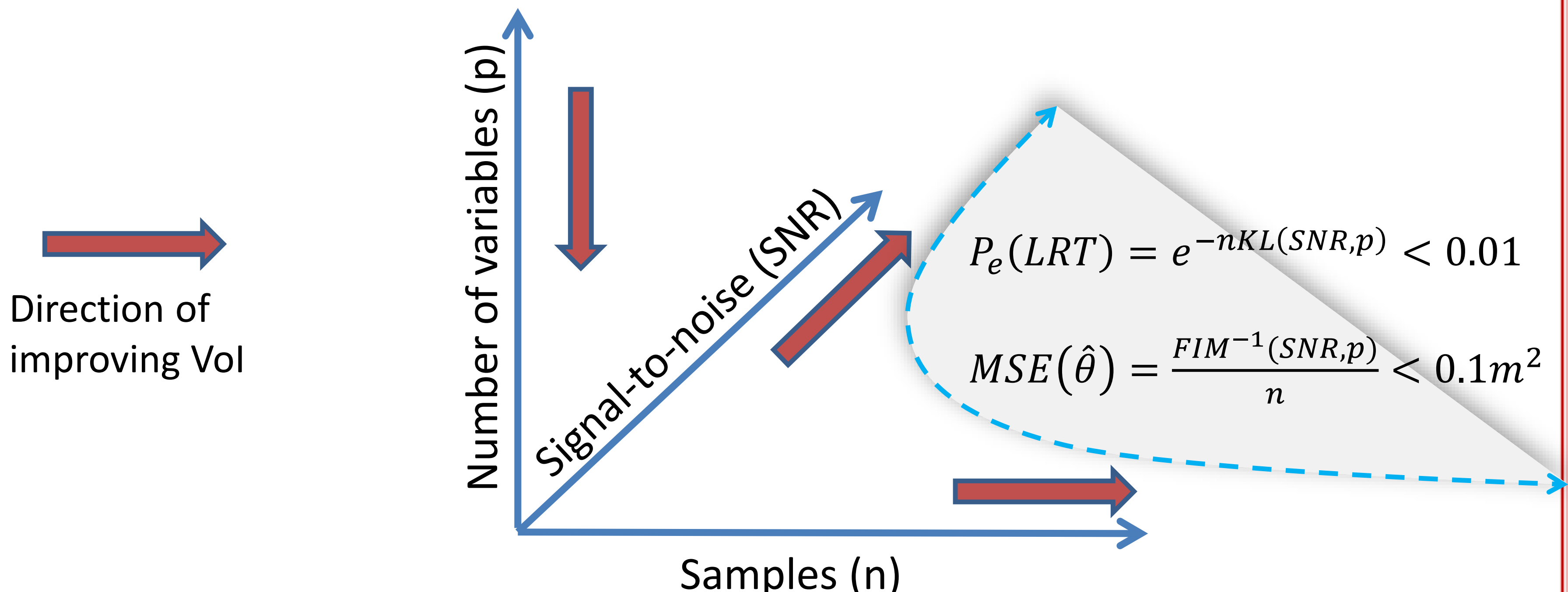


Learning, inference, tracking, and exploitation system

- Different types of data collection actions
  - Take a measurement from selected sensor S
  - Share information between two local processors
  - Acquire contextual knowledge and side info
  - Incorporate human experts in the loop
- Different types of constraints
  - Communications, computation, delay, privacy
- Central questions
  - What data collection actions maximize VoI?
  - How is VoI affected by constraints?
  - How can VoI be translated into design principles?

Value-centric information theory for learning, inference, tracking, and exploitation

- Develop theory of VoI in terms of fundamental limits depending on the task, the selected performance criterion, and the resource constraints.
- Apply VoI theory to develop better feature learning algorithms, information fusion methods, and sensor planning strategies.



Why is this problem difficult?

- There is lots of relevant theory…
  1. **Communication theory**: Shannon theory, rate exponents, entropy inequalities
  2. **Signal representation theory**: statistical modeling/representation, detection/estimation, convex relaxation
  3. **Control theory**: Markov Decision Processes (MDPs), value-function bounds, bandit approximations

- …and these are foundational building blocks that we are using…
- …but, there are gaps that have to be filled
  – Existing theories are inadequate when there are algorithmic complexity constraints
  – Existing theories are inadequate when there are human factors
  – Shannon theory was developed for communications and almost all propositions hold only for infinite block length (not real-time)
  – MDPs do not scale well to practical problem sizes

- We have made progress on filling these gaps

VoI in communications

- Value of information in communications theory
  - Primary task: reliable communication over noisy channel
  - Cost function: bit error rate (ber), information rate (bps)
  - Information measures: Shannon entropy, Shannon mutual information
  - Foundational principle: Shannon's coding and data processing theorems, Nyquist thm
  - Asymptotic regimes: infinite block length, high SNR
  - Deficiencies: asymptotic, no feedback, no real time or timeliness constraints

34 *The Mathematical Theory of Communication*

CE Shannon, BSTJ 1948

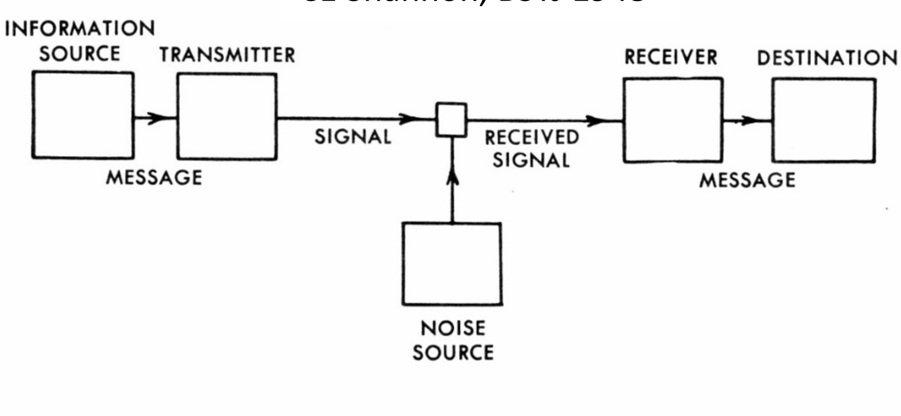
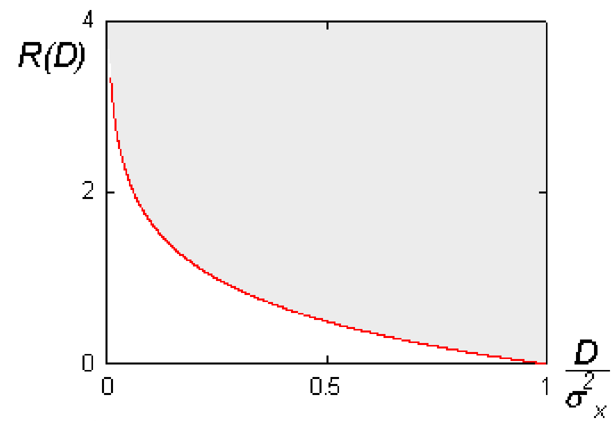
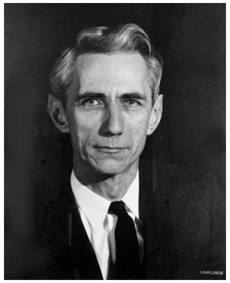
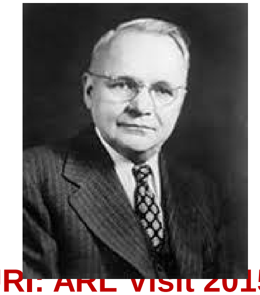


Fig. 1. — Schematic diagram of a general communication system.



http://en.wikipedia.org/wiki/Ratedistortion_theory

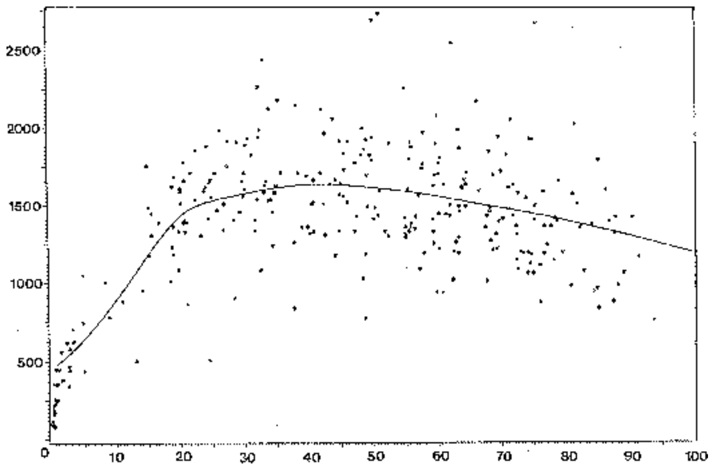


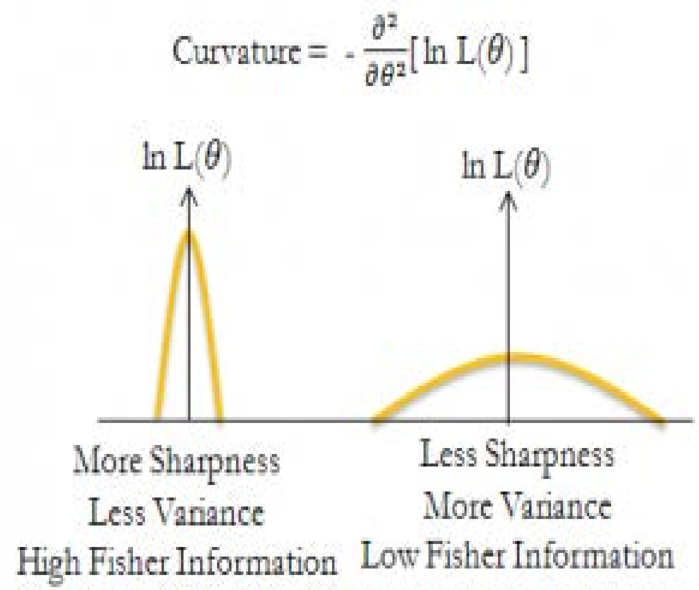


VoI in mathematical statistics

- Value of information in mathematical statistics
  - Primary task: fitting data to a model, parameter estimation or prediction
  - Cost function: risk, empirical risk, bias, variance, misclassification error
  - Information measures: Fisher information, Hellinger affinity
  - Foundational principle: Fisher sufficiency, asymptotic efficiency, Wiener filtering
  - Asymptotic regimes: stationarity, CLT, LLNs, concentration inequalities
  - Deficiencies: no real time constraint, no timeliness or human-in-the-loop

Hardle and Marron, 1985

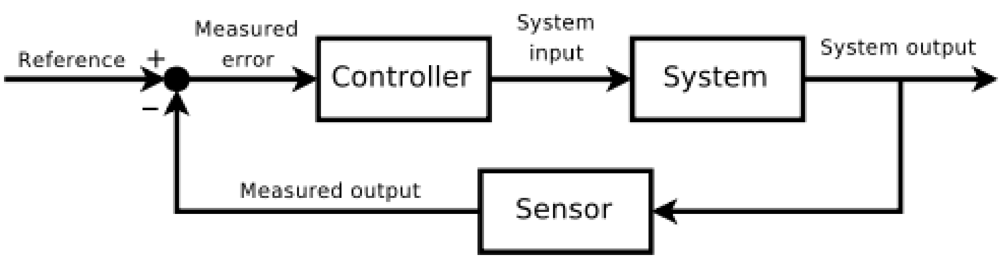




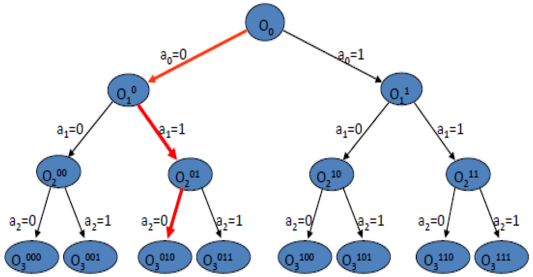
http://www.gaussianwaves.com/tag/fisher-information/

VoI in stochastic control theory

- Value of information in stochastic control theory
  - Primary task: maximize a reward by controlling actions on a measurement space
  - Cost function: general risk function, cost-to-go, future payoff, value functions
  - Information measures: information invariants on belief states (KL divergence)
  - Foundational principle: value-optimal policies via Bellman’s DP, Gittins index thm
  - Asymptotic regimes: myopic limit, limit of infinite horizon
  - Deficiencies: high computational complexity, approximation complexity, no HMI



http://en.wikipedia.org/wiki/Control_theory



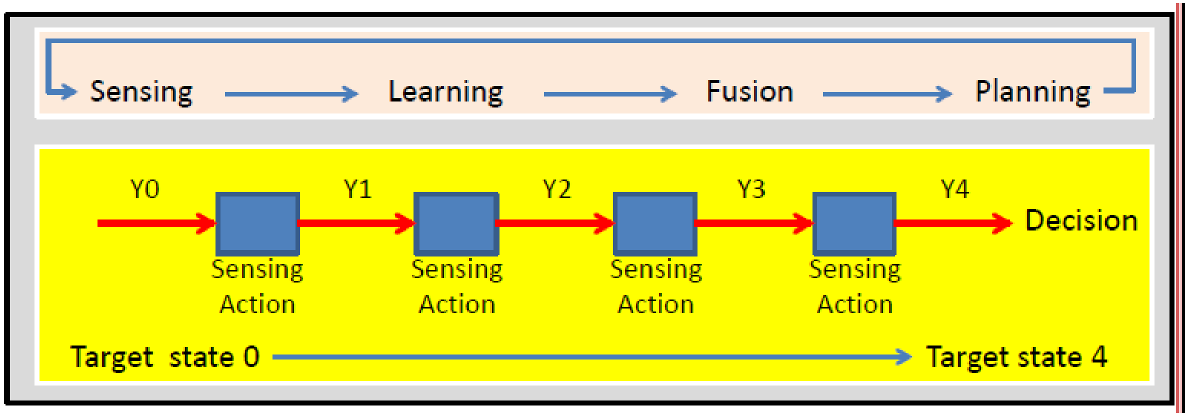
Source: Blatt H 2006

贝尔曼, R.

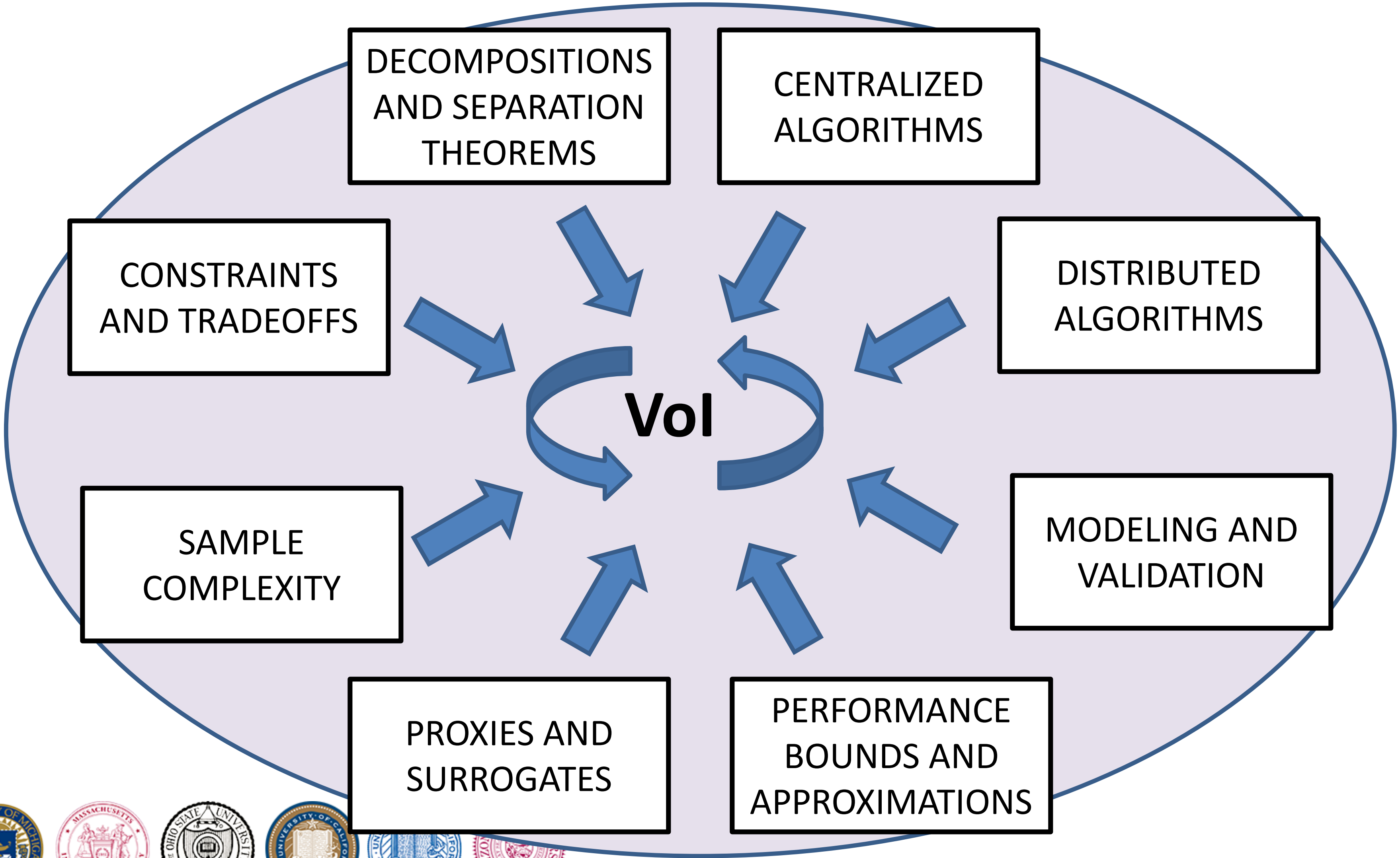
$$V_t(\pi_t) = \max_{a_t} \{R(a_t, \pi_t) + E[V_{t+1}(\pi_{t+1})|a_t, \pi_t]\}$$

VoI theory for sensing and data collection

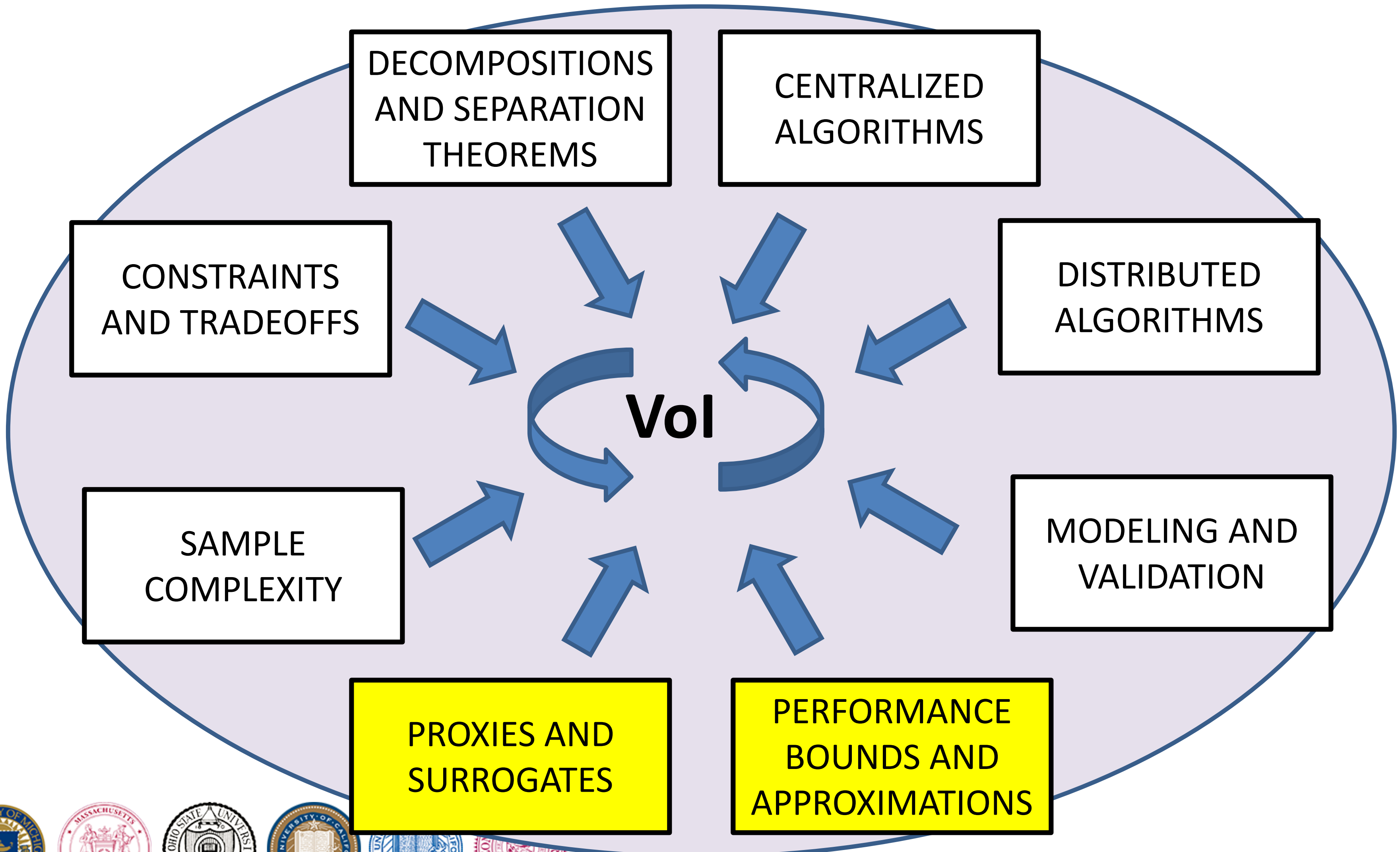
- Value of information in sensing and data collection
  - Primary task: data collection, distributed fusion, and decision-making
  - Cost function: decision-error with constraints on comms, computation, energy, etc
  - Information measures: invariant surrogates and proxies for performance
  - Foundational principle: a data processing theorem for multi-modal sensing
  - Asymptotic regimes: must apply to small sample sizes, distributed collection
  - Opportunities: real-time, time sensitive, mission sensitive, human-in-the-loop



Components of VoI theory



Components of VoI theory



Performance and proxies: some refs

- Non-parametric estimation of information directly from data
  - V. Berisha and A. O. Hero, "Empirical non-parametric estimation of the Fisher Information," IEEE Signal Processing Letters, vol. 22, no. 7, pp 988-992, July 2015.
- Information theoretic bounds on distributed estimation
  - J. C. Duchi, M. I. Jordan, M. J. Wainwright, and Y. Zhang, "Information-theoretic lower bounds for distributed statistical estimation with communication constraints,“ Proceedings of NIPS, 2014.
- Information geometric proxies for planning
  - S. D. Howard, W. Moran, and D. Cochran, "Intrinsic Fisher information on manifolds," *Proceedings of the Asilomar Conference on Signals, Systems, and Computers,* November 2014
- Convex proxies for resource allocation and mission planning
  - D. Wei and A.O. Hero, " Performance Guarantees for Adaptive Estimation of Sparse Signals," IEEE Transactions on Information Theory, to appear 2015.
  - G. Newstadt, B. Mu, D. Wei, J. P. How and A.O Hero, ""Importance-weighted adaptive search for multi-class targets," IEEE Trans on Signal Processing, to appear 2015
- Minimal sufficient and invariant proxies for deep learning networks
  - S. Soatto and A. Chiuso, "Visual Scene Representations: sufficiency, minimality, invariance and deep approximation", *International Conference on Learning Representations*, 2015.

What constitutes a good proxy for VoI?

- P0: Computability. Computational complexity of optimizing over policy space A
  - Example: greedy myopic approximation to multi-stage planning (Fisher, Hero, How, Cochran)
- P1: Intrinsic monotonicity: If proxy increases then expected reward should too
  - Example: correct classification probability asymptotically non-decreasing in KL divergence (Hero, Soatto)
- P2: Order preservation. Highest ranked actions same/similar to reward-intrinsic ranking.
  - Example: rank-preserving index proxies in multi-armed bandit approximations (Yu)
- P3: Intrinsic-to-task. Proxy is a fundamental limit or bound
  - Example: Fisher information, Chernoff information (Ertin, Soatto, Hero)
- P4: Side-information. Proxy approximation adapts to availability of new side information.
  - Example: Bayes update to fuse side-information into belief function (Hero)
- P5: Sample monotonicity of information: taking additional measurements can do no harm
  - Example: All proxies without measurement penalty terms or stopping rules
- P6: Action monotonicity: expanding available actions cannot decrease reward
  - Example: Any proxy that is maximized over nested action spaces
- P7: Data processing theorem: proxy is non-decreasing over processing resources
  - Example: Mutual information (Fisher)
- P8: Information invariance: information preserving transformations preserve proxy
  - Example: Mutual information (Fisher)
- P9: Proxy captures exploitation vs exploration tradeoff
  - Example: knowledge-gradient, single stage proxies with intervisibility penalties (Yu, Hero)

Proxies and surrogates: Multistage proxies

Proxy design strategies for multistage planning

- Multistage planning **index proxies** to reduce search complexity
  - Assign score functions to actions and choose action w/ max score at each stage
  - Reduces search complexity from $O(|A|^T)$ to $O(|A|T)$
  - Optimal when action space is finite and feasible performance space is polymatroid
  - Example: multi-armed bandit approximation (Yu)

- Multistage planning **myopic proxies** to reduce search complexity
  - Implement a sequence of T one-step lookahead policies
  - Reduction of search complexity from $O(|A|^T)$ to $O(|A|T)$
  - Optimizing φ_a over myopic policies comes within a constant factor of gold-standard VoI when avg reward R is a sub-modular (Krause and Guestrin 2007).
  - Examples: E[R]=mutual information (Fisher), E[R]=community detection (Hero)

- Multistage planning **myopic+ proxies**
  - Add randomization to myopic policy to promote exploration of action space
  - Example: Mission adaptive search policies [Hero and How], Knowledge gradient [Yu]

Proxies and surrogates: single-stage proxies

Proxy design strategies for single stage planning

- **Asymptotic proxies** for VoI
  - Use limiting form VoI measures: large sample size, large action space, high SNR,…
  - Asymptotic expressions for performance are often tractable and lead to insight
  - Asymptotic optimizers of asymptotic proxies often close to finite sample optimal
  - Example: CLT, LLNS, high dimensional limiting forms [Nadakuditi], [Hero].

- **Information theoretic proxies** for VoI
  - Optimize an information measure instead of the average reward
  - Information theoretic planning asymptotically concentrates posterior distribution
    - Data-processing theorem, Cramer-Rao theorem, Hoeffding-Chernoff theorem, Duncan's theorem, Shannon's rate distortion theorem.
  - Using an information proxy for planning is a "hedge" against overfitting to task
  - Example: information geometric planning [Cochran], Fisher info [Ertin and Moses], divergence estimation [Hero], mobile sensing and navigation [Soatto]

Proxies and surrogates: single-stage proxies

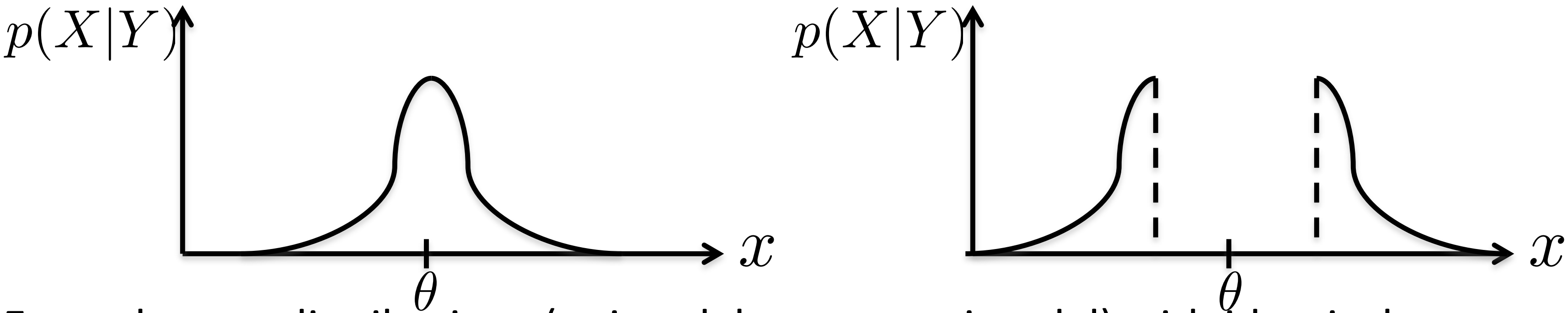
Proxy design strategies for single stage planning (ctd)

- **Convex proxies** for VoI
  - Drive the choice of proxy by computational considerations: convex optimization
  - Especially useful when multiple actions are to be selected in combination
  - Reduce combinatorial problem to a continuous optimization, e.g., via L1 relaxation
  - Example: L1/L0/trace norm feature selection [Hero], [Nadakuditi], [Jordan]

- **Bound proxies** for VoI
  - Maximize a lower bound on average reward over action space
  - Performance upper bounded by that of an oracle with access to hidden info
  - Sometimes bound proxies are also asymptotic proxies **and** information proxies
    - Cramer-Rao lower bound on MSE, Chernoff lower bound on Pe, Fano lower bound on Pe
  - Example: heterogenous mission planning [Hero and How], adversarial planning [Ertin], sensor networks [Moses]

A theorem on entropy proxies

Q: When does an entropy proxy progressively concentrate posterior?



Example: two distributions (unimodal vs. non-unimodal) with identical entropy

Theorem: unimodal distribution $p(X|Y)$, exponentially decreasing tails,

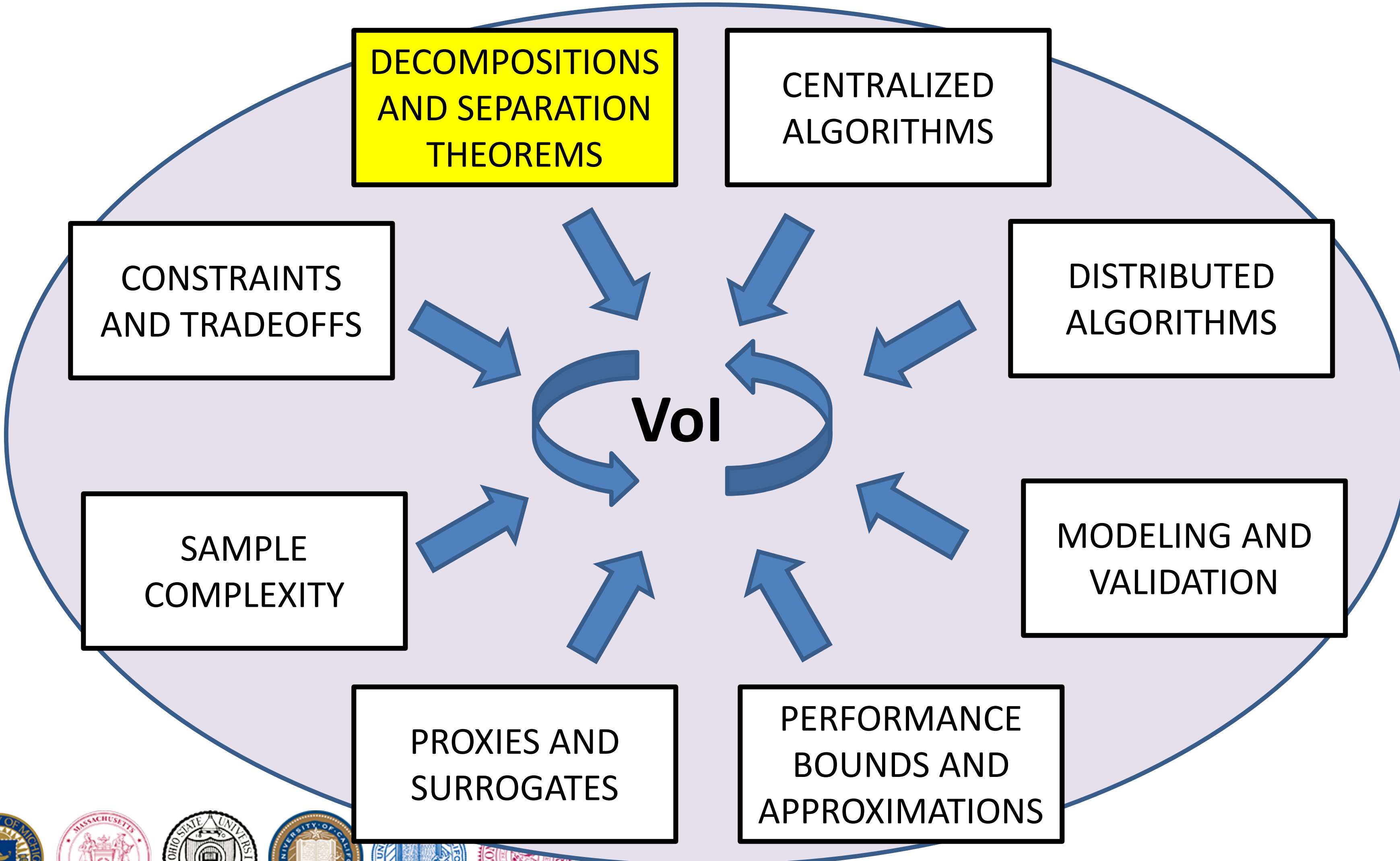
$$\frac{e^{2\cdot h(p(X|Y))}}{2\pi e} \overset{(a)}{\leq} \mathrm{var}(X|Y) \overset{(b)}{\leq} \frac{\alpha}{(\mathcal{I}(p(X|Y)))^{\beta}} \overset{(c)}{\leq} \frac{\alpha \cdot e^{2\beta\cdot h(p(X|Y))}}{2\pi e}$$

Differential entropy $h(p(X|Y))$ and Fisher information $\mathcal{I}(p(X|Y))$

Proof: (a): Estimation counterpart to Fano's inequality, (b):[Chung et al 2015], (c): Stam's inequality

Ref: Chung, Sadler, Hero. submitted 2015.

Components of VoI theory



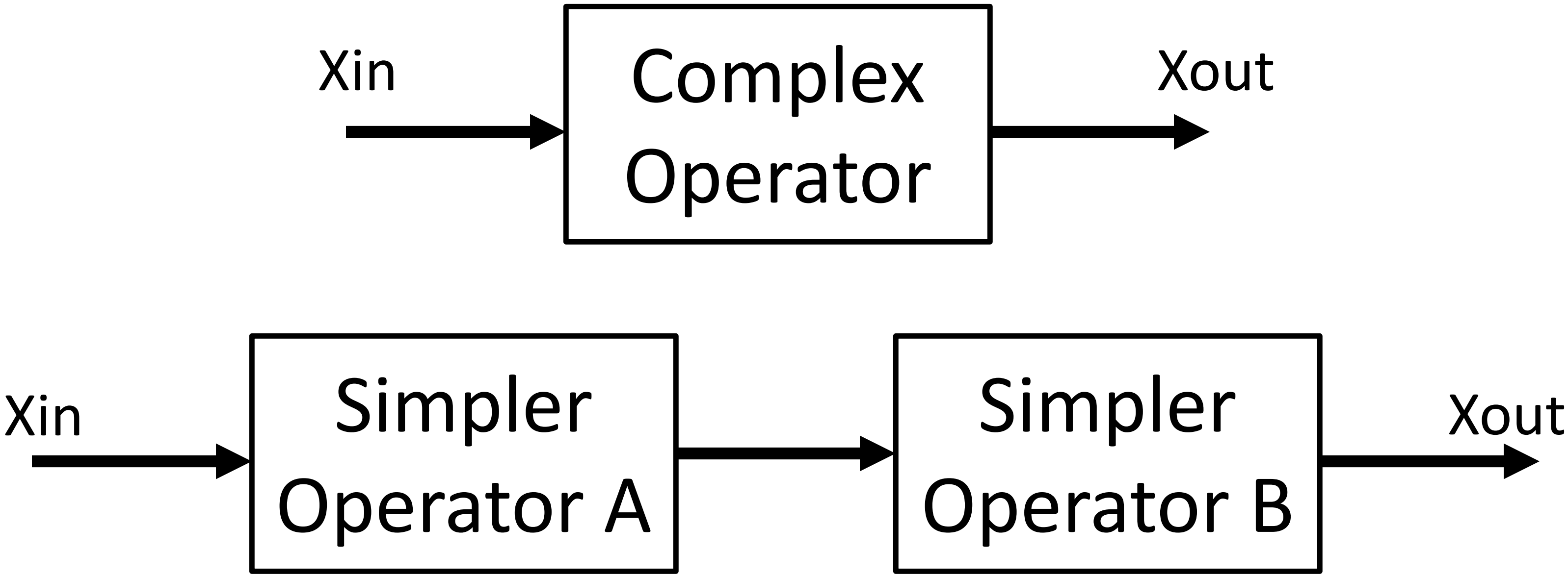
VOI

Decompositions and separations: some refs

- Myopic multistage-planning decompositions
  - G. Papachristoudis and J. W. Fisher III, "Efficient information planning in Markov chains" (in review 2015).
  - P-Y Chen and A.O. Hero, "Phase transitions in spectral community detection of large noisy networks," IEEE Intl Conf on Acoustics, Speech, and Signal Processing (ICASSP), Brisbane, April 2015.
- Separation theorem for collaborative 20 questions
  - T. Tsiligkaridis, B. M. Sadler, and A. O. Hero, "Collaborative 20 questions for localization," *IEEE Transactions on Information Theory.* vol. 60, no. 4, pp 2233-2252, April 2014.
- Separation of fusion and denoising
  - R. R. Nadakuditi, "OptShrink: An Algorithm for Improved Low-Rank Signal Matrix Denoising by Optimal, Data-Driven Singular Value Shrinkage," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 3002 - 3018, May 2014
- Kronecker decomposition for spatio-temporal analysis
  - T. Tsiligkaridis and A. O. Hero, "Covariance estimation in high dimensions via Kronecker product expansions," *IEEE Transactions on Signal Processing,* vol. 61, no. 21, pp. 5347 - 5360, Nov 2013.
  - K. Greenewald, E. Zelnio, A.O. Hero, "Kronecker PCA Based Robust SAR STAP,“ (in review 2015).

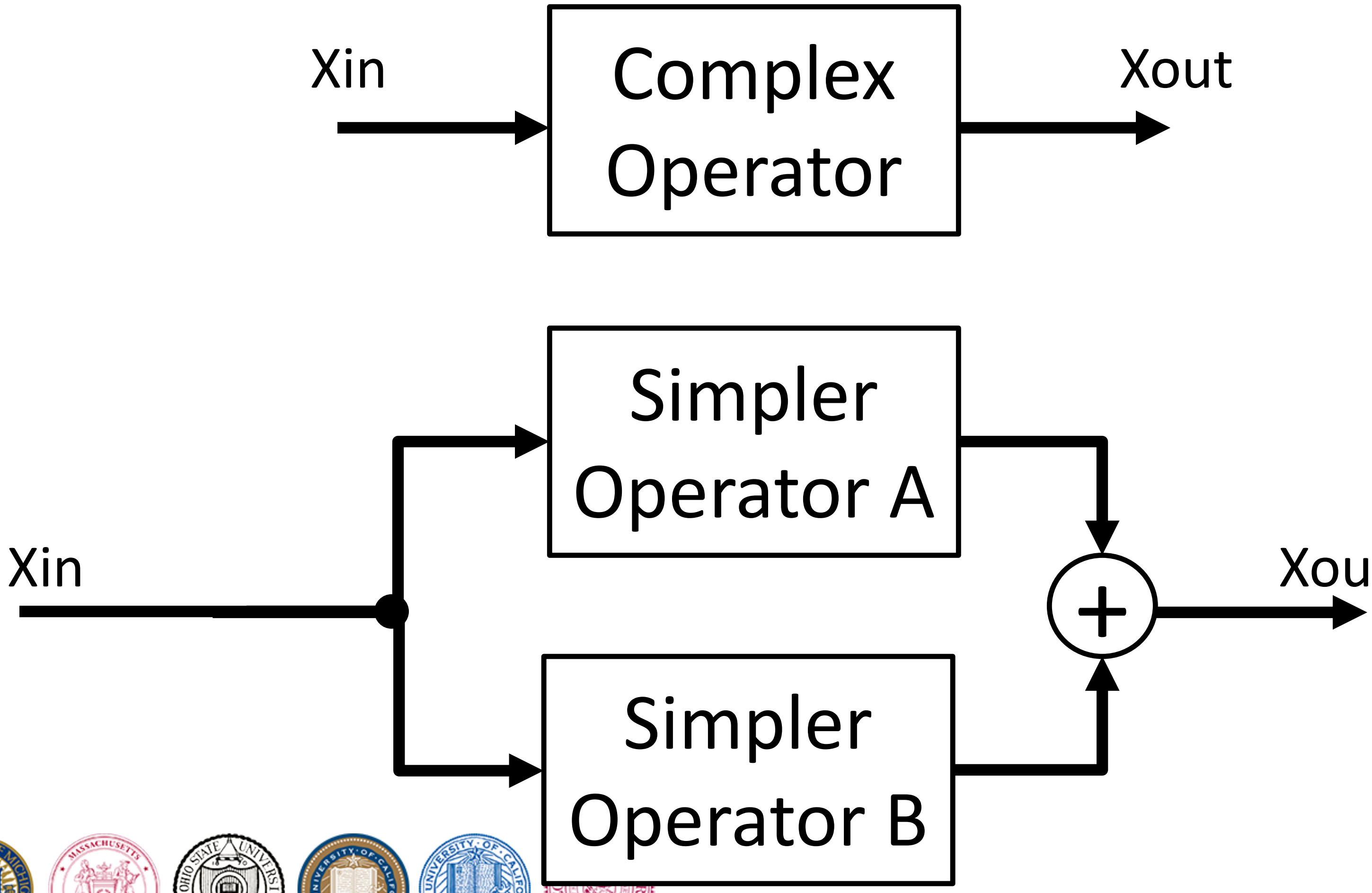
Decompositions and separation theorems

A serial **separation principle** is a decomposition of an operator into a cascade of simpler operators A and B



Decompositions and separation theorems

A parallel **separation principle** is a decomposition of an operator into a sum of simpler operators



Examples of serial separation principles

- Communication
  – Source/channel coding separation theorem for discrete memoryless channels (DMC) (Shannon 1948)
- Stochastic control
  – Separation of estimation and control in LQG (Wonham 1968)
- Mathematical statistics
  – Separation property for estimator of a function of a parameter – transformation invariance property of MLE.
- Convex optimization
  – Variables splitting via augmented Lagrangian - alternating direction method of multipliers (ADMM)

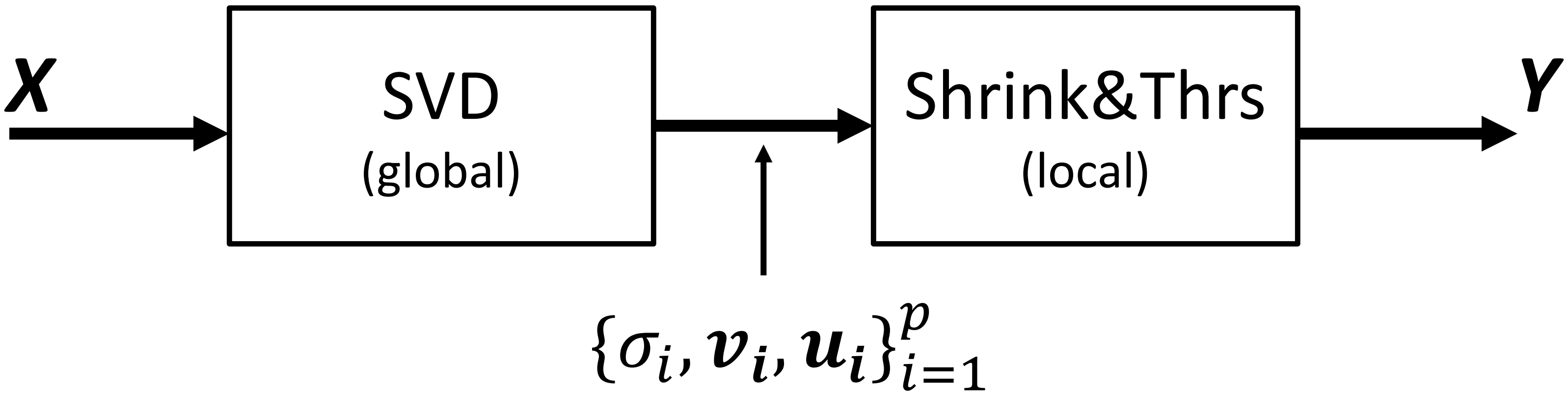
Serial separation principles for data collection

VOI

Application to subspace processing and information fusion

- Low rank plus noise model: $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{W}$. $\boldsymbol{X} = \boldsymbol{A}\boldsymbol{\Lambda}\boldsymbol{B}$.
- Estimation proxy: $||\boldsymbol{X} - \boldsymbol{Y}||^2 + \beta||\boldsymbol{Y}||_*$. Frobenius + nuclear.
- Solution ***Y*** that minimizes proxy is (Lounici 2013, Tsiligkaridis 2014)

$$\boldsymbol{Y} = \sum_{i=1}^{p} \left[\sigma_i - \frac{\beta}{2}\right]_+ \boldsymbol{v}_i \boldsymbol{u}_i^T$$

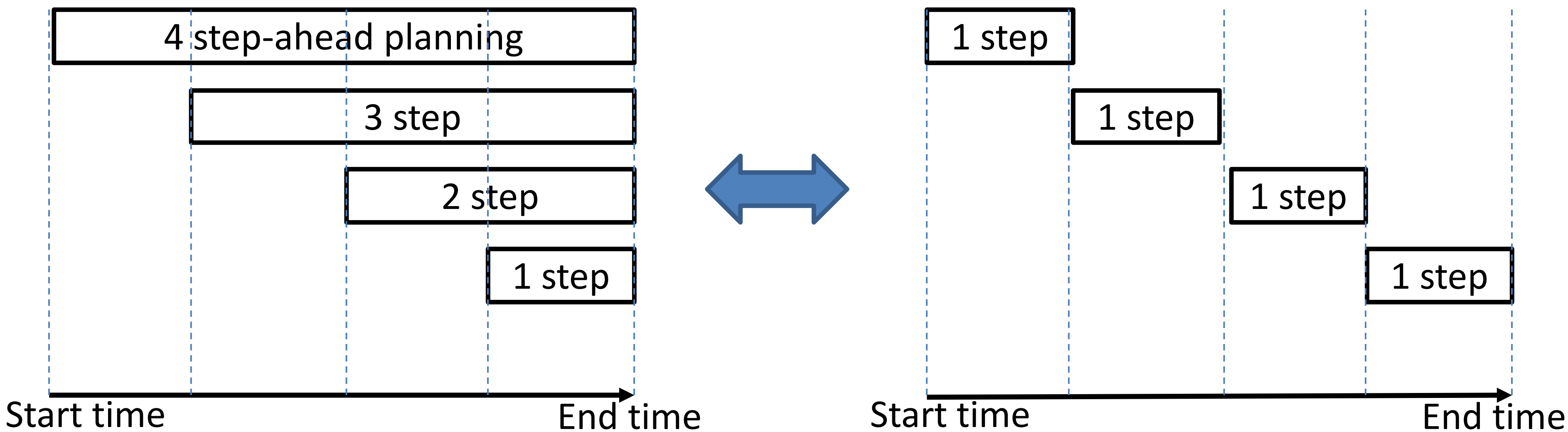


- Denoising proxy (Nadakuditi 2014) gives better local Shrink&Thrs

Serial separation principles for data collection

- Multistage planning for state tracking (Fisher 2013) and for network community detection (Chen, Hero 2014).
- Weighted MI proxy is sub-modular (Fisher 2013)
- Fiedler spectral centrality proxy is sub-modular (Chen, Hero 2014)
- Myopic proxy approximates optimal Value up to constant factor

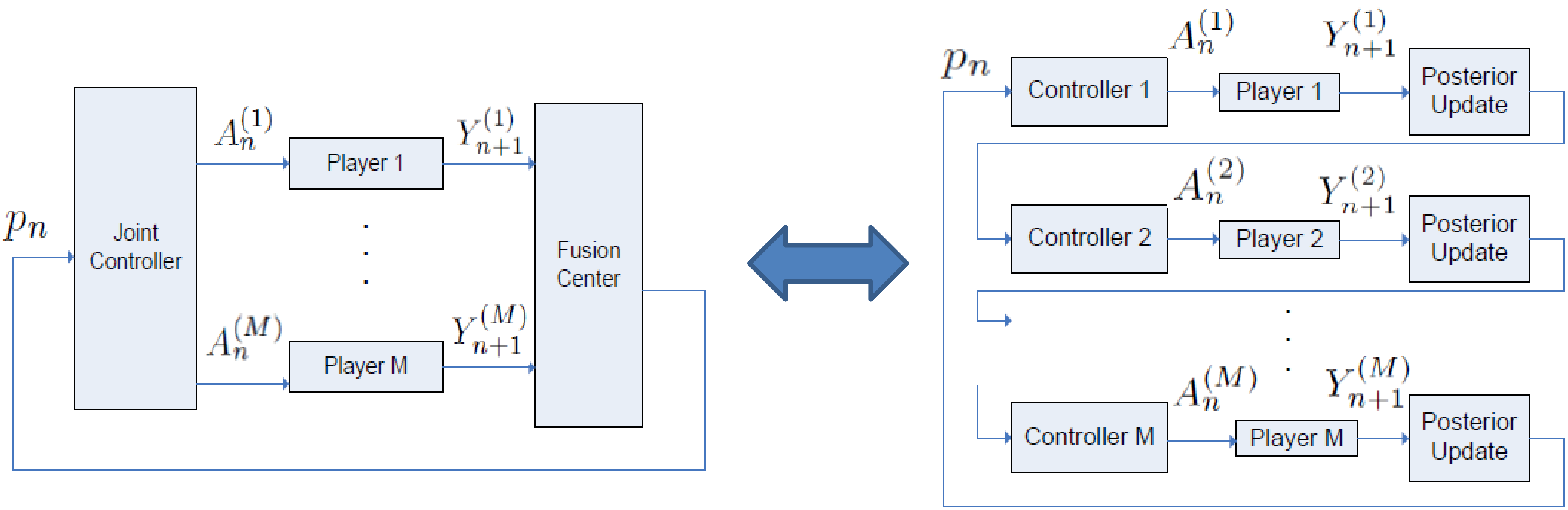
Optimal multistage planning **Myopic multistage planning**



$$E[\varphi_\pi] \geq (1 - \mathrm{e}^{-1})\, V_\pi$$

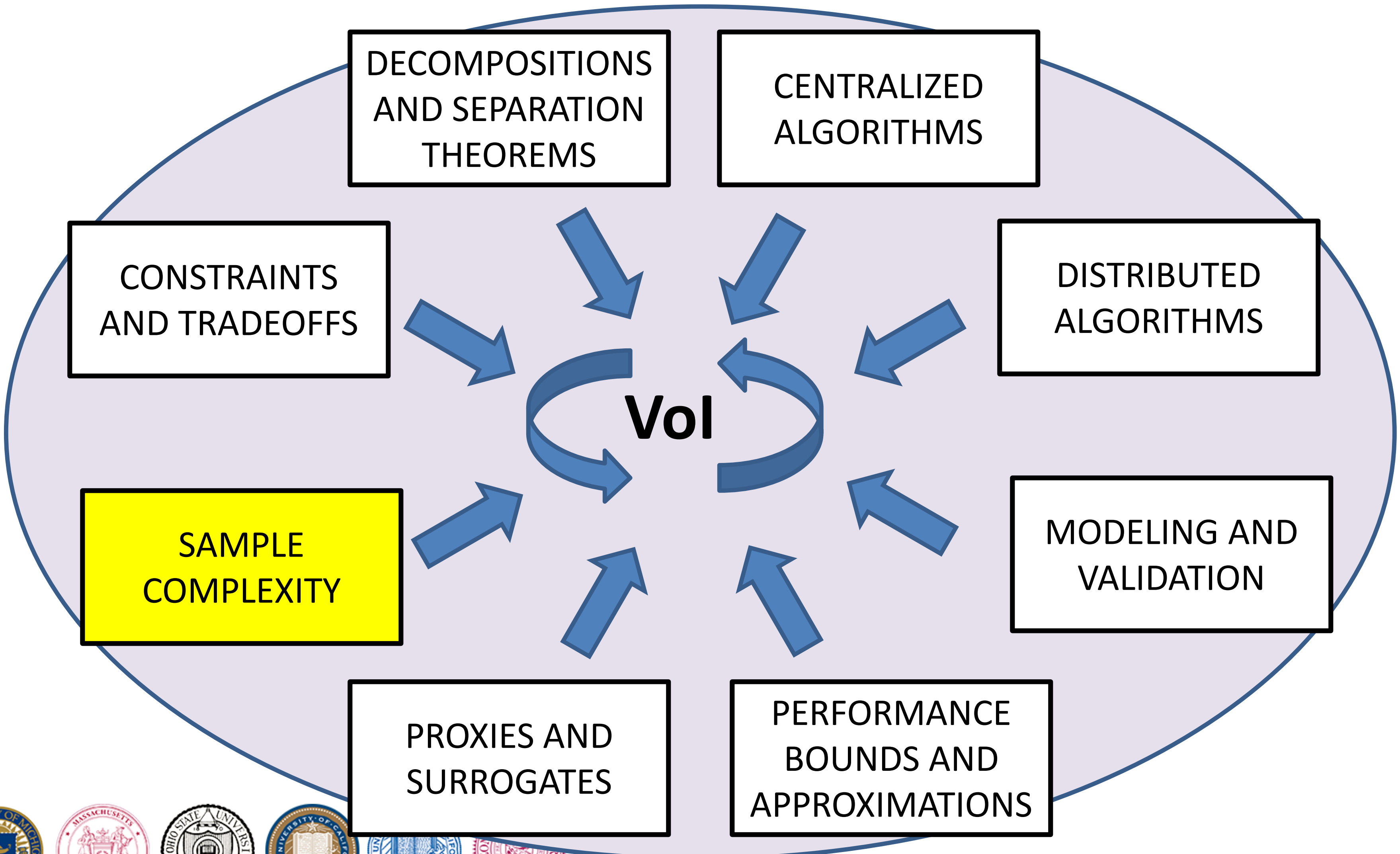
Example: parallel vs serial querying

- Collaborative 20 questions target search with multiple agents
- Entropy-optimal joint query control policy is same as entropy-optimal serial query control policy (Tsiligkaridis, Sadler, H 2013)
- This is a serial commutative separation principle
- May not hold for non-entropic proxies



Ref: Tsiligkaridis,Sadler, H IEEE TSP 2013

Components of VoI theory



Sample complexity: some refs

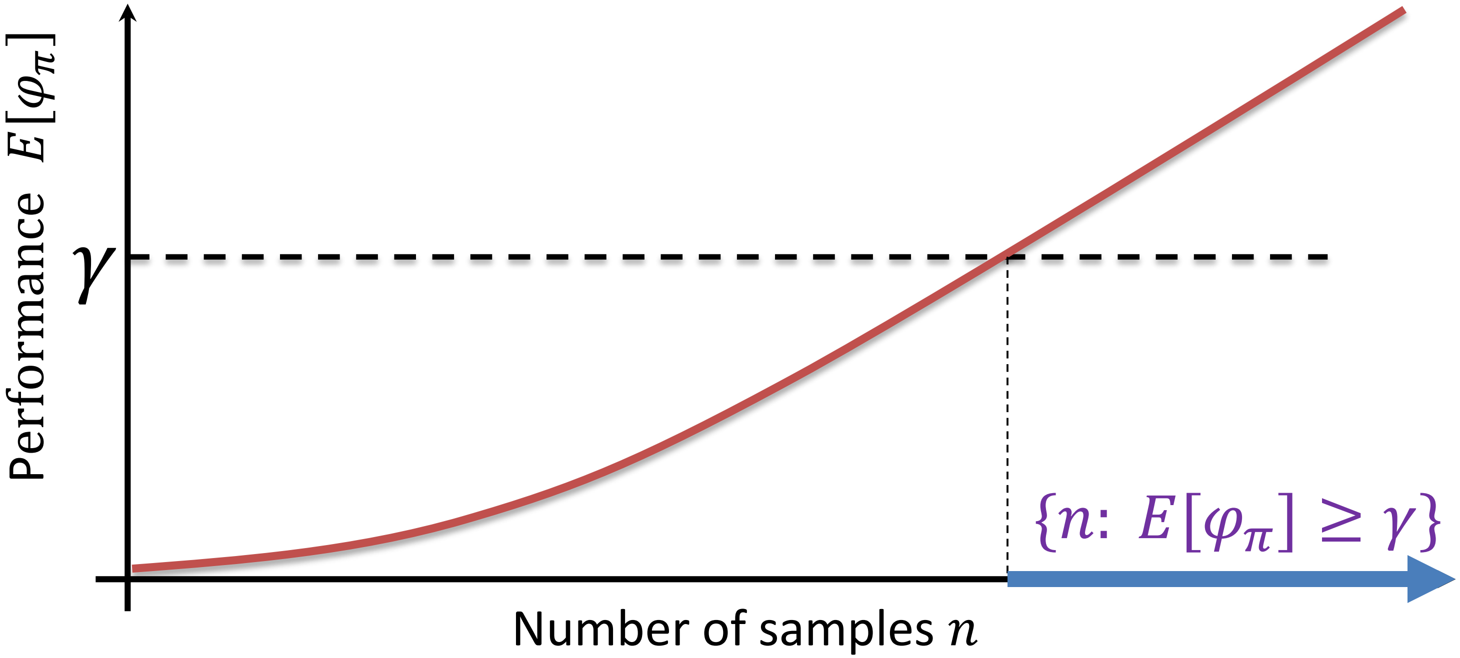
- High dimensional learning rates for low rank + sparse GGM
  - Z. Meng, B. Erikson, A.O. Hero, "Learning Latent Variable Gaussian Graphical Models," *Proceedings of the International Conference on Machine Learning (ICML),* Beijing, July 2014.
- High dimensional learning rates for spatio-temporal covariance estimation
  - T. Tsiligkaridis and A. O. Hero, "Covariance estimation in high dimensions via Kronecker product expansions," IEEE Transactions on Signal Processing, vol. 61, no. 21, pp. 5347 - 5360, Nov 2013.
  - T. Tsiligkaridis and A.0. Hero, and S. Zhou, "On convergence of Kronecker graphical lasso algorithms," IEEE Transactions on Signal Processing, vol. 61, no. 9, pp. 1743--1755, 2013
  - A. O. Hero and B. Rajaratnam, "Foundational principles for large scale inference: Illustrations through correlation mining," IEEE Proceedings. In press 2015.
- Tradeoffs between performance and computation as function of samples
  - M. I. Jordan, "On statistics, computation, and scalability," Bernoulli, 19, 1378-1390, (2013).
- High dimensional learning rates for uncertainty (divergence) estimation
  - K. Moon and A.O. Hero, "Ensemble estimation of multivariate f-divergence," Proc. of IEEE Intl. Symposium on Information Theory (ISIT), Hawaii, June 2014.

Sample sufficiency

- Let γ specify a desired performance benchmark for a task

$$E[\varphi_\pi] \geq \gamma$$

  - Sample sufficiency: $\min\{\text{samples n} \mid \varphi_n \geq \gamma\}$
  - Computation sufficiency: $\min\{\text{flops f} \mid \varphi_f \geq \gamma\}$
  - Communications sufficiency: $\min\{\text{transmissions t} \mid \varphi_t \geq \gamma\}$

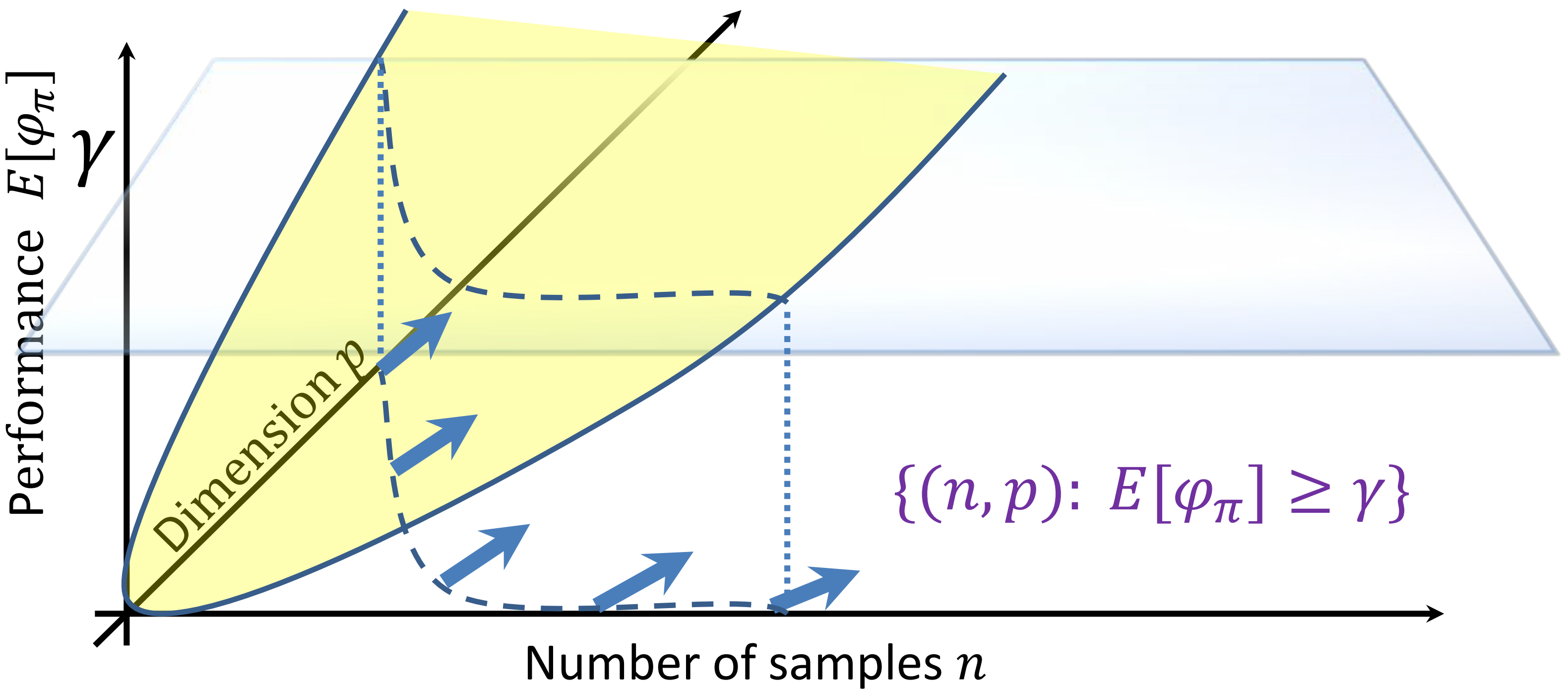


Sample complexity

- Let γ specify a desired performance benchmark for a task

$$E[\varphi_\pi] \geq \gamma$$

- Assume n samples and p unknown variables (dimension)
- Sample complexity is: minvalue of $n = n(p)$ ensuring benchmark



Sample complexity

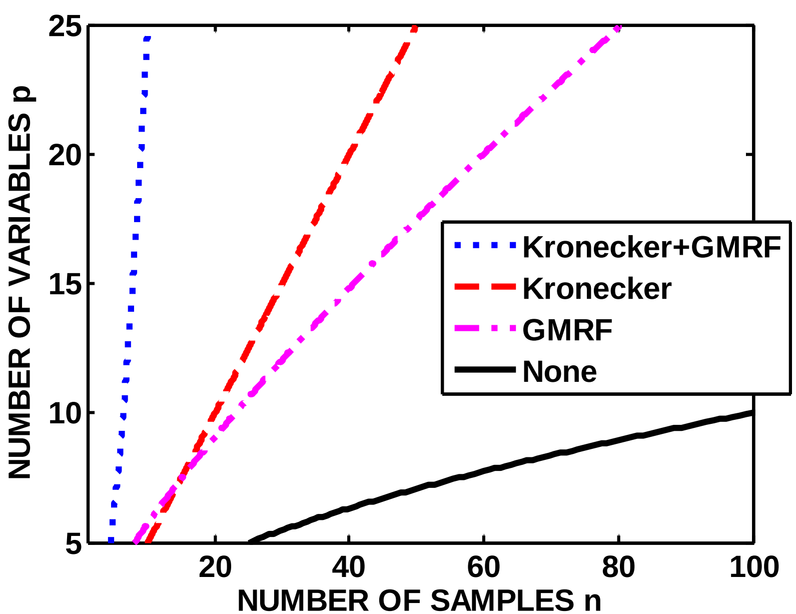
- Questions of interest
  - Question 1: what is intrinsic value of contextual info for a task?
    - How would environmental domain knowledge improve performance?
  - Question 2: how does the VoI intrinsically depend on the task?
    - How many samples required to attain performance benchmarks?
  - Question 3: what is the effect of data collection or processing constraints on attainable VoI?
    - What additional resources are required to achieve benchmark?

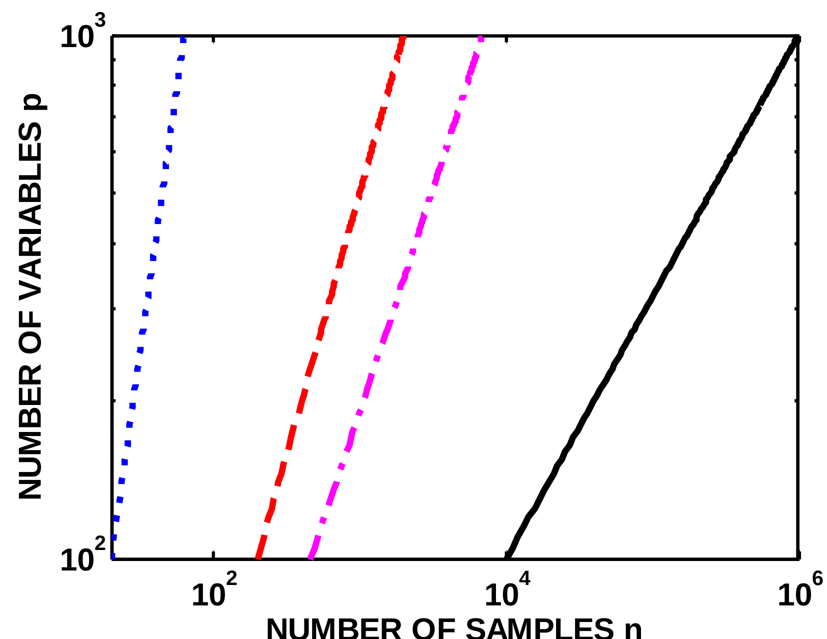
Sample complexity: inference on covariance

Effect of **contextual information** on VoI and sample complexity

- **Regimes:** The sample complexity decreases as more info is available

Information	None	GMRF	Kronecker	Kronecker+GMRF
Model	saturated $\mathbf{\Omega}$	sparse $\mathbf{\Omega}$	$\mathbf{\Omega} = \mathbf{A} \otimes \mathbf{B}$	sparse $\mathbf{\Omega} = \mathbf{A} \otimes \mathbf{B}$
$\log f(\mathbf{\Omega})$	constant	$\lambda \Vert\mathbf{\Omega}\Vert_1$	$\delta\left(\text{rank}\mathcal{R}(\mathbf{\Omega}) - 1\right)$	$\delta\left(\text{rank}\mathcal{R}(\mathbf{\Omega}) - 1\right) + \lambda_1 \Vert\mathbf{A}\Vert_1 + \lambda_2 \Vert\mathbf{B}\Vert_2$
Proxy	$\frac{1}{2}\log\left(\frac{q^2r^2}{n}\right)$	$\frac{1}{2}\log\left(\frac{qr\log qr}{n}\right)$	$\frac{1}{2}\log\left(\frac{(q^2+r^2)\log M}{n}\right)$	$\frac{1}{2}\log\left(\frac{(q+r)\log M}{n}\right)$
Regime	$\frac{q^2r^2}{n} \to \alpha$	$\frac{qr\log qr}{n} \to \alpha$	$\frac{(q^2+r^2)\log M}{n} \to \alpha$	$\frac{(q+r)\log M}{n} \to \alpha$





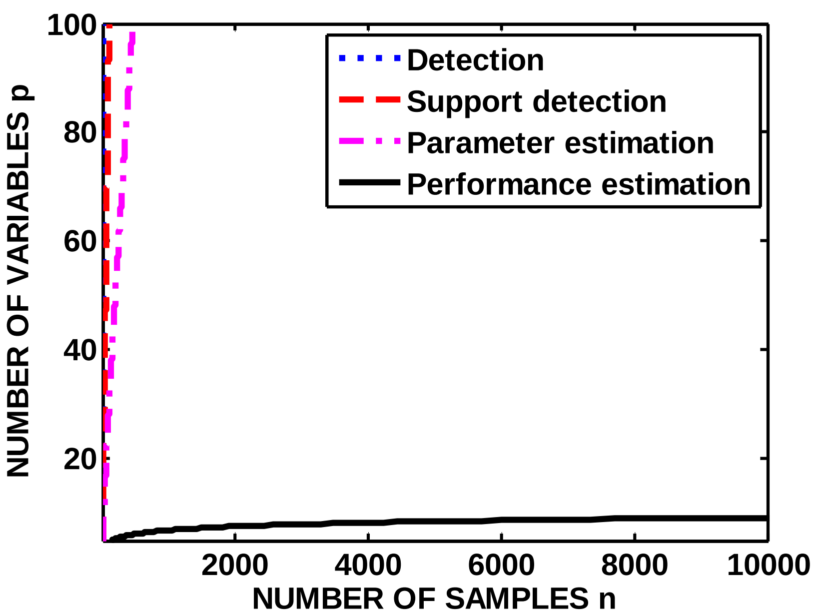
Hero and Rajaratnam, IEEE Proc. (2015)

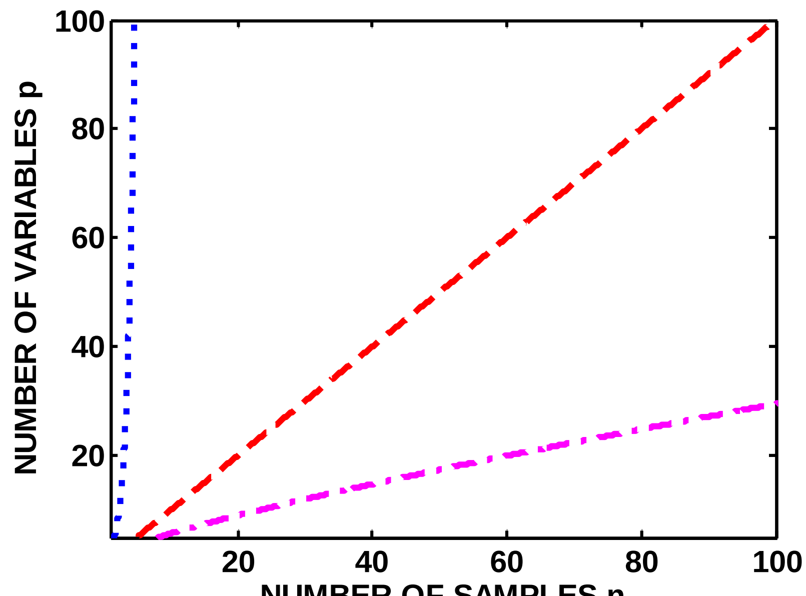
Sample complexity: inference on covariance

Effect of **task** on VoI and sample complexity

- **Regimes:** The sample complexity decreases as task becomes easier

Task	Screening	Detection	Support detection	Param. estimation	Perform. estimation
Risk	$P(N_e > 0)$	$P(N_e > 0)$	$P(\mathrm{card}\{\mathcal{S}\Delta\hat{\mathcal{S}}\} = \phi)$	$E[\lVert\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}\rVert_F^2]$	$\int E[(f_{\Sigma}(\mathbf{x}) - \hat{f}(\mathbf{x}))^2]d\mathbf{x}$
Proxy	$1 - e^{-\kappa_n}$	$pe^{-n\alpha}$	$2^p e^{-n\alpha}$	$\frac{p}{n}\alpha$	$n^{-2/(1+p)}\alpha$
Regimes	$\frac{\log p}{n} \to \infty$	$\frac{\log p}{n} \to \alpha$	$\frac{p}{n} \to \alpha$	$\frac{p \log p}{n} \to \alpha$	$\frac{p}{\log n} \to \alpha$
Threshold	$\rho_c \to 1$	$\rho_c \to \rho^*$	$\rho_c \to 0$	$\rho_c \to 0$	$\rho_c \to 0$

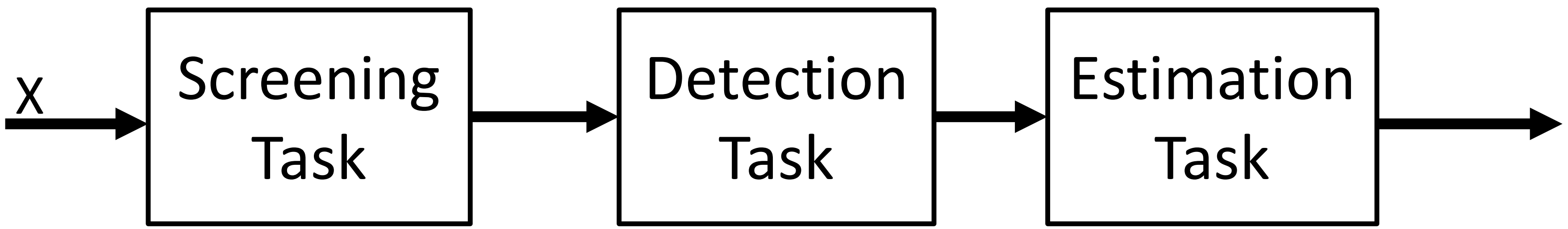


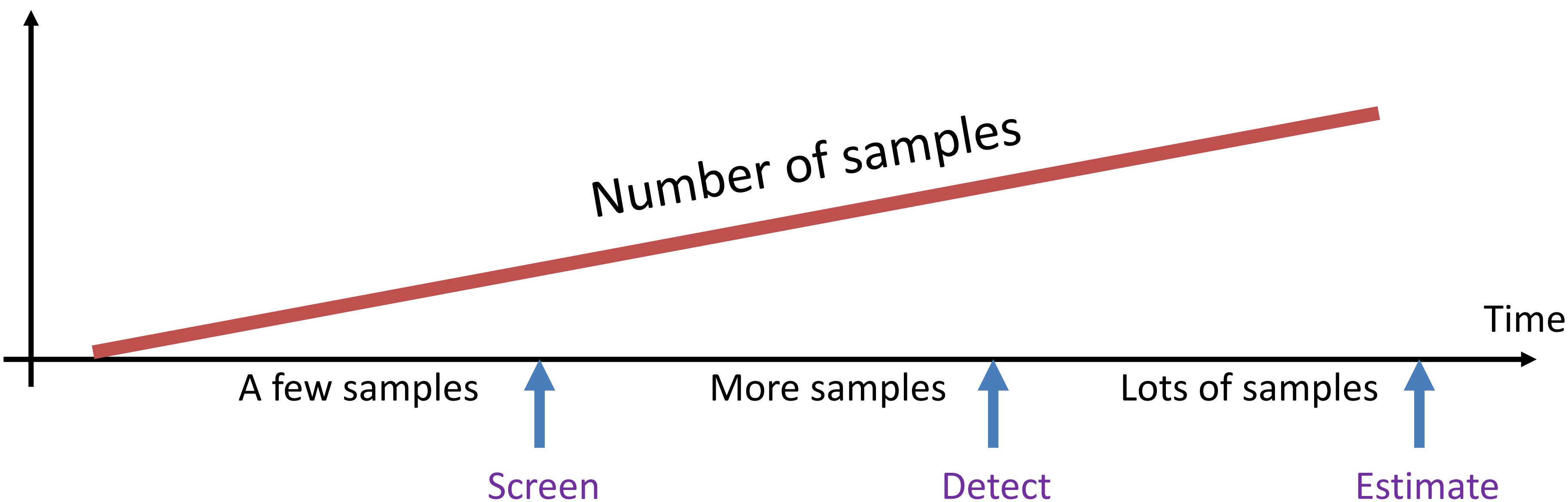


Hero and Rajaratnam, IEEE Proc. (2015)

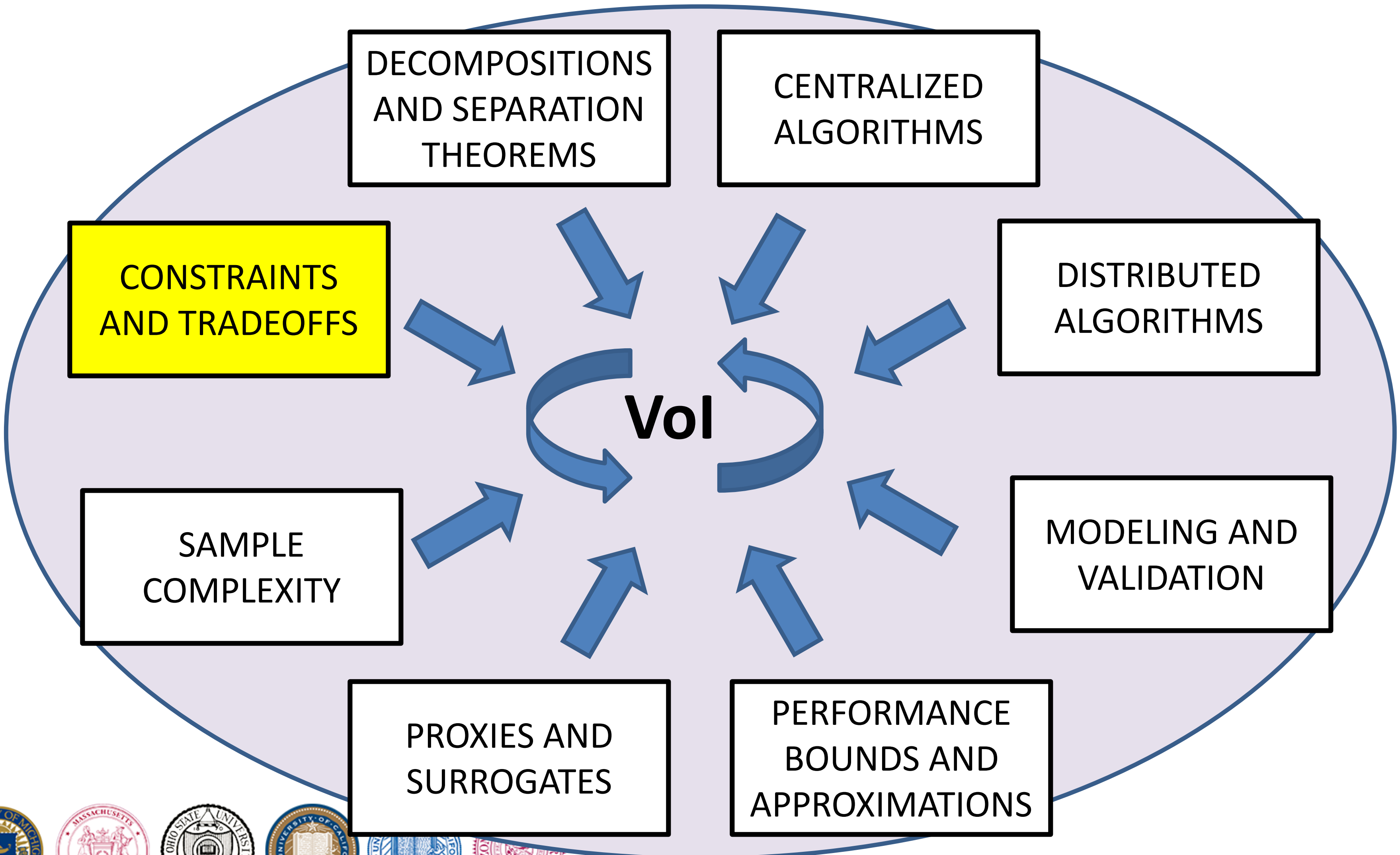
Implication: matching tasks to sample size

Sequence tasks in order of increasing sample complexity requirements (Firouzi, Hero, Rajaratnam, 2014)





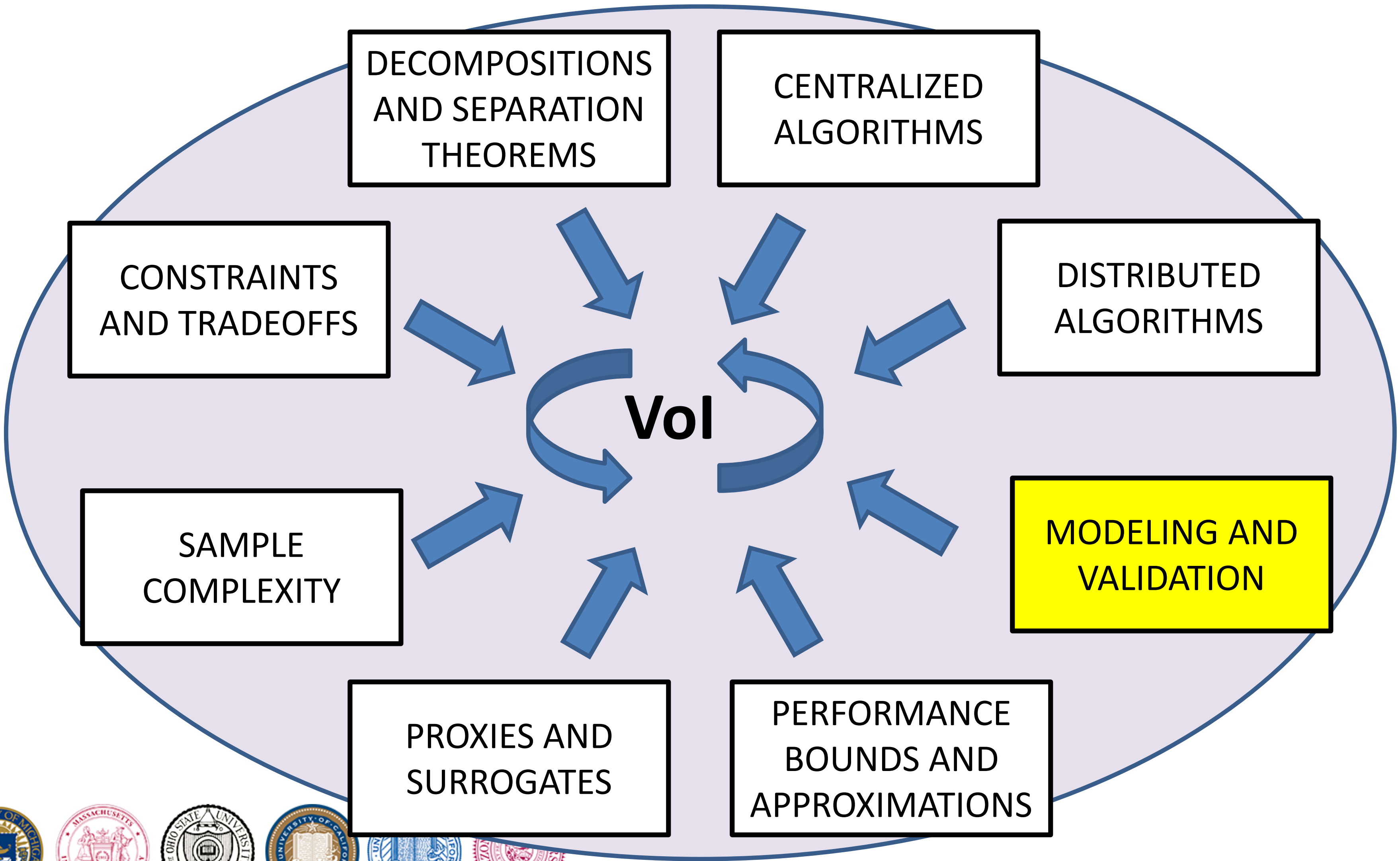
Components of VoI theory



Constraints and tradeoffs: some refs

- Fusion with missing or corrupted data
  - T. Xie, N. Nasrabadi and A.O. Hero, "Learning to classify with possible sensor failures," Proc. of IEEE Conf. on Acoustics, Speech and Signal Processing (ICASSP), Florence, May 2014.
- Fundamental limits on estimation with privacy constraints
  - J. Duchi, M. I. Jordan, and M. Wainwright, "Local privacy and minimax bounds: Sharp rates for probability estimation," (2014). Advances in Neural Information Processing (NIPS) 26, Red Hook, NY: Curran Associates.
- Fundamental limits on fusion with communications constraints
  - J. C. Duchi, M. I. Jordan, M. J. Wainwright, and Y. Zhang, "Information-theoretic lower bounds for distributed statistical estimation with communication constraints," Proceedings of NIPS, 2014.

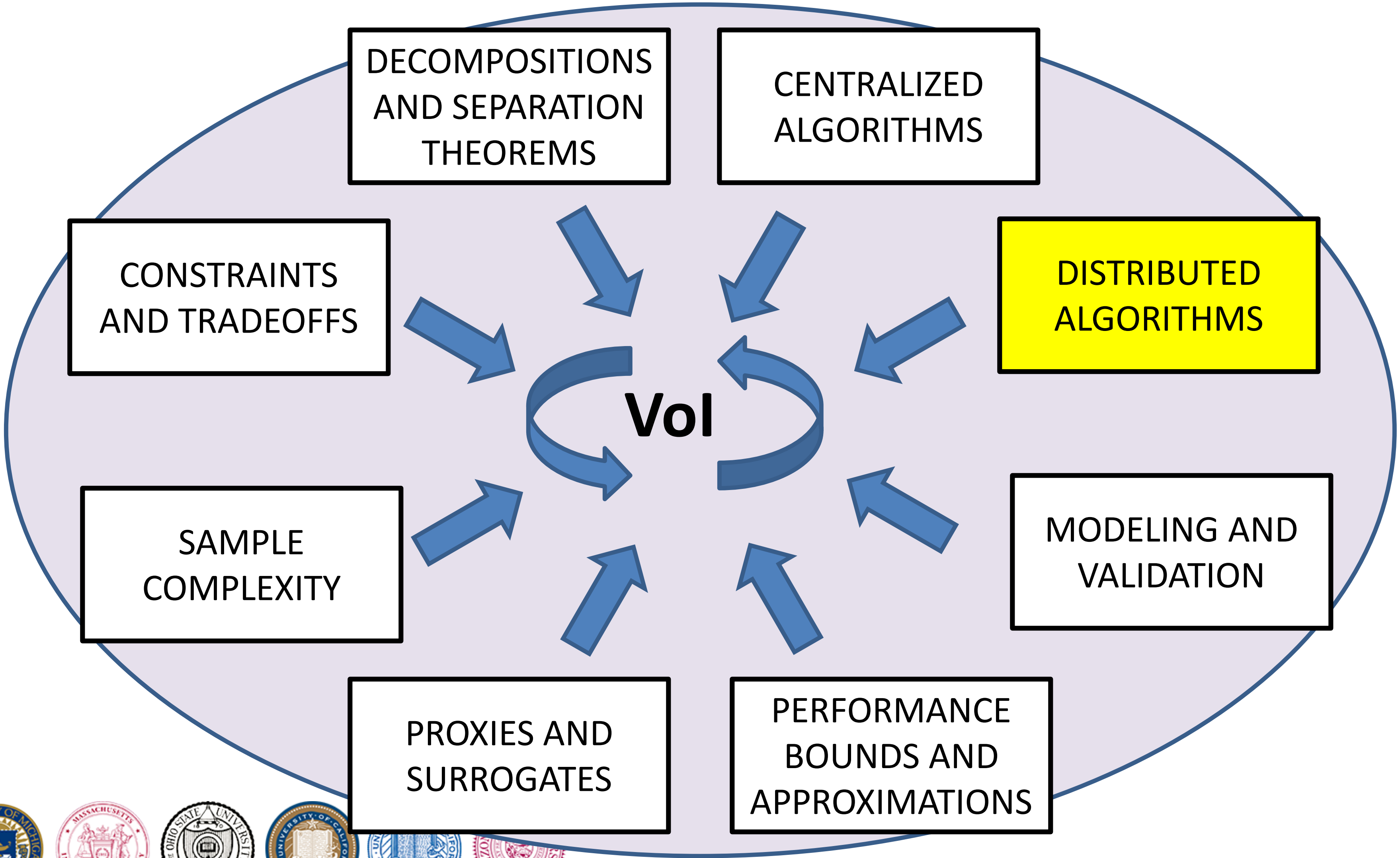
Components of VoI theory



Modeling and validation: some refs

- Validation of radar models for SAR imaging
  - J. Ash, E. Ertin, L. C. Potter, and E. Zelnio, "Wide-Angle Synthetic Aperture Radar Imaging: Models and algorithms for anisotropic scattering" IEEE Signal Processing Magazine, vol. 31, no. 4, pp 16-26, 2014
- Human response models for collaborative 20 questions problems
  - T. Tsiligkaridis, B. M. Sadler, and A. O. Hero, "Collaborative 20 questions for localization," *IEEE Transactions on Information Theory.* vol. 60, no. 4, pp 2233-2252, April 2014
- Validation of human action and interaction models
  - S. Zhang and A. J. Yu. "Forgetful Bayes and myopic planning: Human learning and decision-making in a bandit setting," *Advances in Neural Information Processing Systems*, 26. MIT Press, Cambridge, MA, 2013.
  - S. Ahmad, and A. J. Yu, "A socially aware Bayesian model for competitive foraging", Proceedings of the Cognitive Science Society Conference, 2014.
- Models in "learning to rank" problems
  - J. Duchi, L. Mackey, and M. I. Jordan, "The asymptotics of ranking algorithms," *Annals of Statistics* 41(5):2292-2323, 2013.
- Models and validation for crowdsourcing
  - F. L. Wauthier and M. I. Jordan, "Bayesian bias mitigation for crowdsourcing," In P. Bartlett, F. Pereira, J. Shawe-Taylor and R. Zemel (Eds.), Advances in Neural Information Processing Systems (NIPS) 24. MIT Press, Cambridge, MA, 2012.
  - Y. Zhang, X. Chen, D. Zhou, and M. I. Jordan, "Spectral methods meet EM: A provably optimal algorithm for crowdsourcing," *Advances in Neural Information Processing Systems*, to appear 2015

Components of VoI theory



Distributed algorithms: some refs

- Distributed network fusion without message passing
  - Z. Meng, D. Wei, A. Wiesel, and A.O. Hero, "Distributed Learning of Gaussian Graphical Models via Marginal Likelihoods," *IEEE Transactions on Signal Processing* vol 62, no. 20, pp. 5425-5438. Nov. 2014
- Target localization via collaborative 20 questions over a network
  - T. Tsiligkaridis, B. M. Sadler and A. O. Hero III, "On decentralized estimation with active queries," *IEEE Trans on Signal Processing*. In press 2015.
- Distributed subspace estimation
  - Y. Zhang, M. J. Wainwright, and M. I. Jordan, "Distributed estimation of generalized matrix rank: Efficient algorithms and lower bounds," *International Conference on Machine Learning*, 2015.
- Pooling for deep architectures in computer vision
  - J. Dong and S. Soatto, "Domain-size pooling in local descriptors: DSP-SIFT", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- Multi-agent networks for competitive foraging
  - S. Ahmad, and A. J. Yu, "A socially aware Bayesian model for competitive foraging", *Proceedings of the Cognitive Science Society Conference,* 2014.
- Information sharing with adaptive censoring
  - B. Mu, G. Chowdhary, and J. P. How, "Efficient distributed sensing using adaptive censoring-based inference," *Automatica,* vol. 50, no. 6, pp. 1590 - 1602, 2014.

Conclusions

- Team is making fundamental advances in building a VoI framework in sensing and data collection
- Elements of framework
  – Performance bounds and approximations
  – Proxies and surrogates for performance
  – Sample complexity (sampling requirements)
  – Constraints and tradeoffs (computation, communication, privacy)
  – Decompositions and separation theorems
- Fresh VoI perspective to design of (de)centralized algorithms
- Importance of experimental validation
  – Experiments on real data from domains including: radar, vision, acoustic, knowledge bases, social media feeds, web crowds, IRB approved human experiments (Hero, Ertin, Soatto, Fisher, Nadakuditi, Jordan, Yu)
  – Software radar testbed is being deployed (Ertin, Cochran, Fisher, How)