Coupling Multiple Hypothesis Testing with Proportion Estimation in Heterogeneous Categorical Sensor Networks

DTRA SBIR Phase II Contract: W911SR-10-C-0038

Chris Calderon, Ph.D.

Austin Jones

Scott Lundberg

Randy Paffenroth, Ph.D.

Presentation Outline

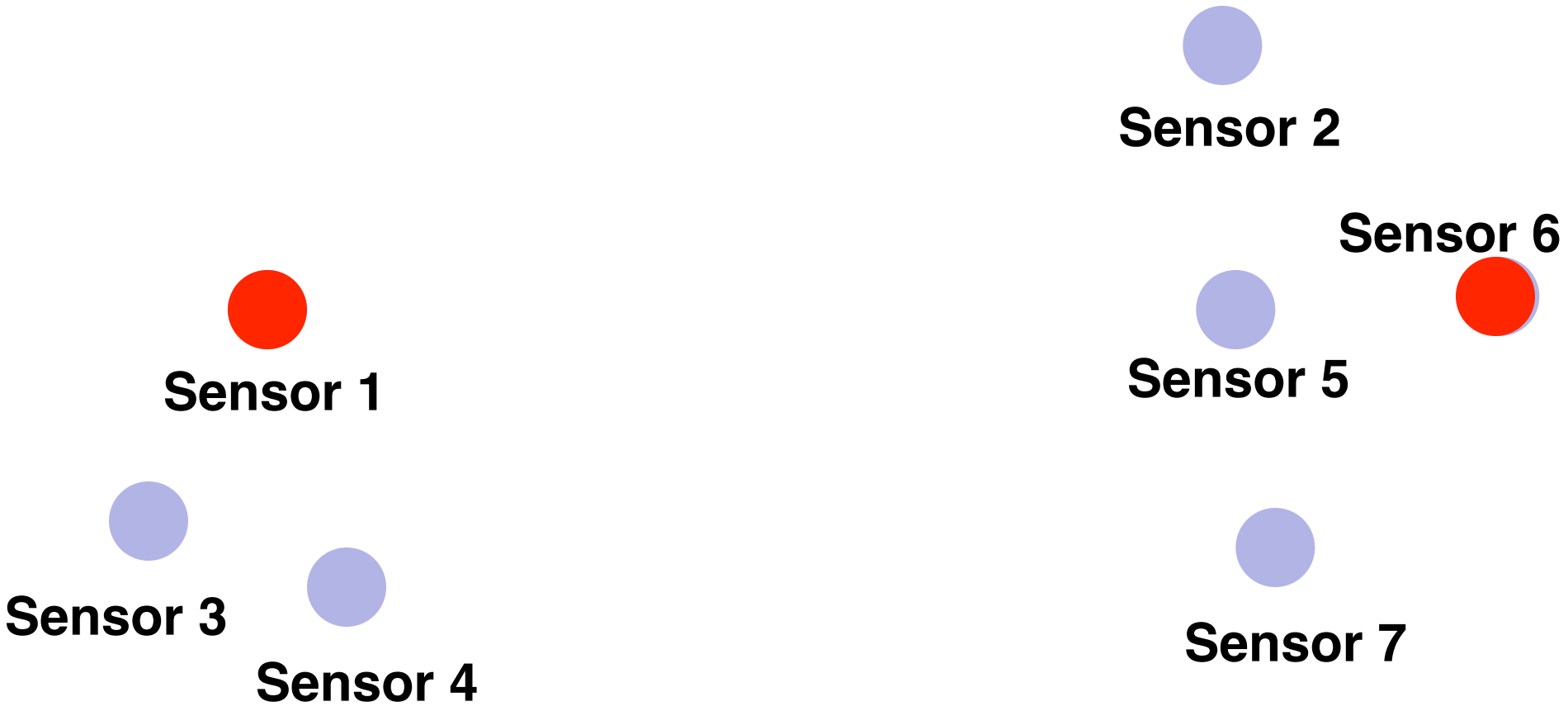
- Illustrate various flavors of “False Alarm” problem facing Department of Threat Reduction Agency (DTRA) chemical detection applications
- Coupling estimation with testing. Quickly highlight proportion estimation.
- Discuss our thoughts on future directions
- Goal: Estimate environment specific effects and “fuse” information (through multiple testing) while maintaining high power and an “Operationally Acceptable” type I error (more “liberal” multiple testing required)

Some Cartoon Sketches of the Problem(s)

NOTE:

Throughout Every “Sensor Alarm” Registered Does NOT Correspond to a Harmful Chemical Threat / Attack. *It is a False Alarm.*

Case I (Identical Point Sensors "Behaving")



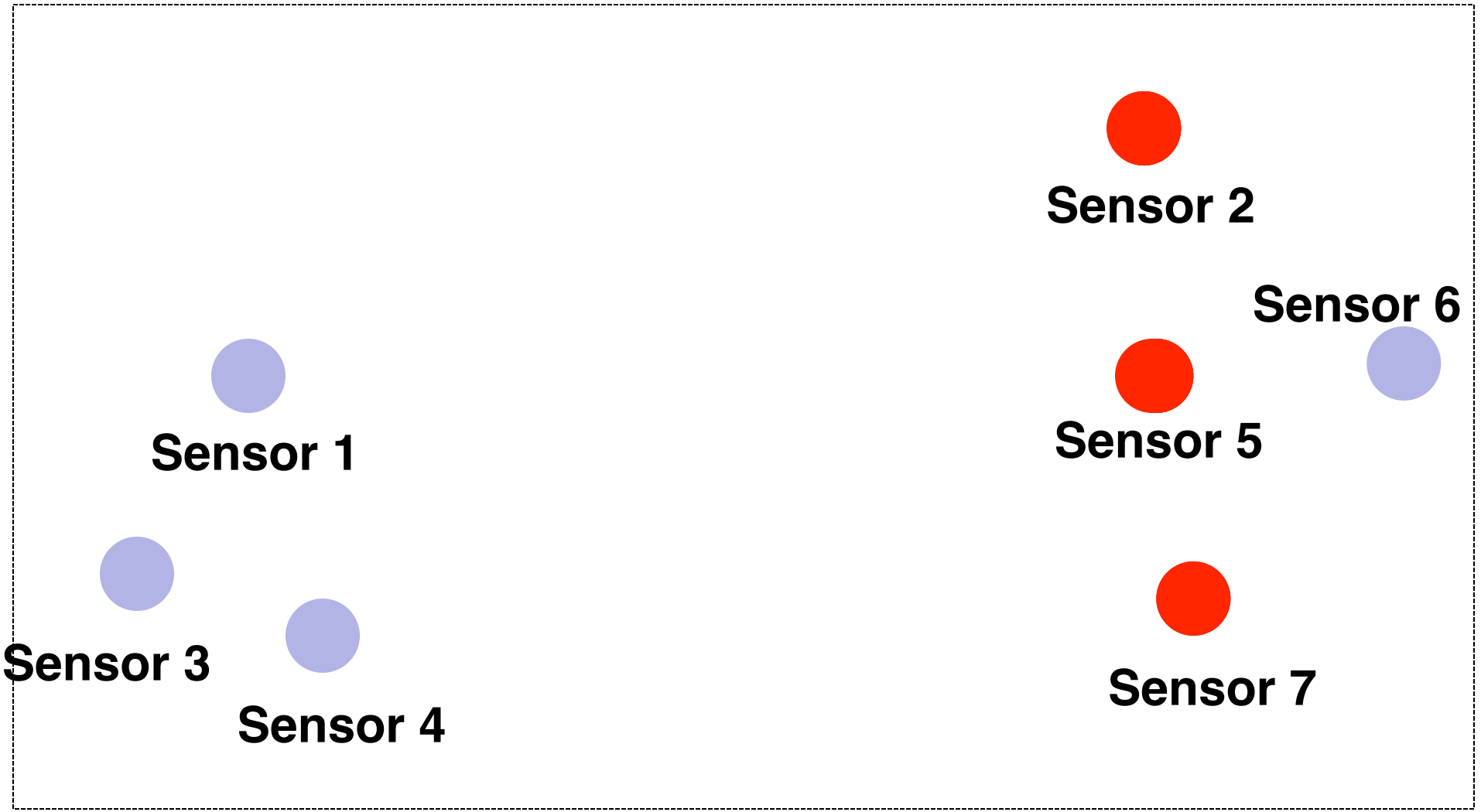
Case I (Identical Point Sensors "Behaving")

Sensors Independent and Identically Distributed (i.i.d.)

Recall the "Y" and "N" Bars Denote the Probability of an Alarm

Extension to Multiple Categorical Threats Possible

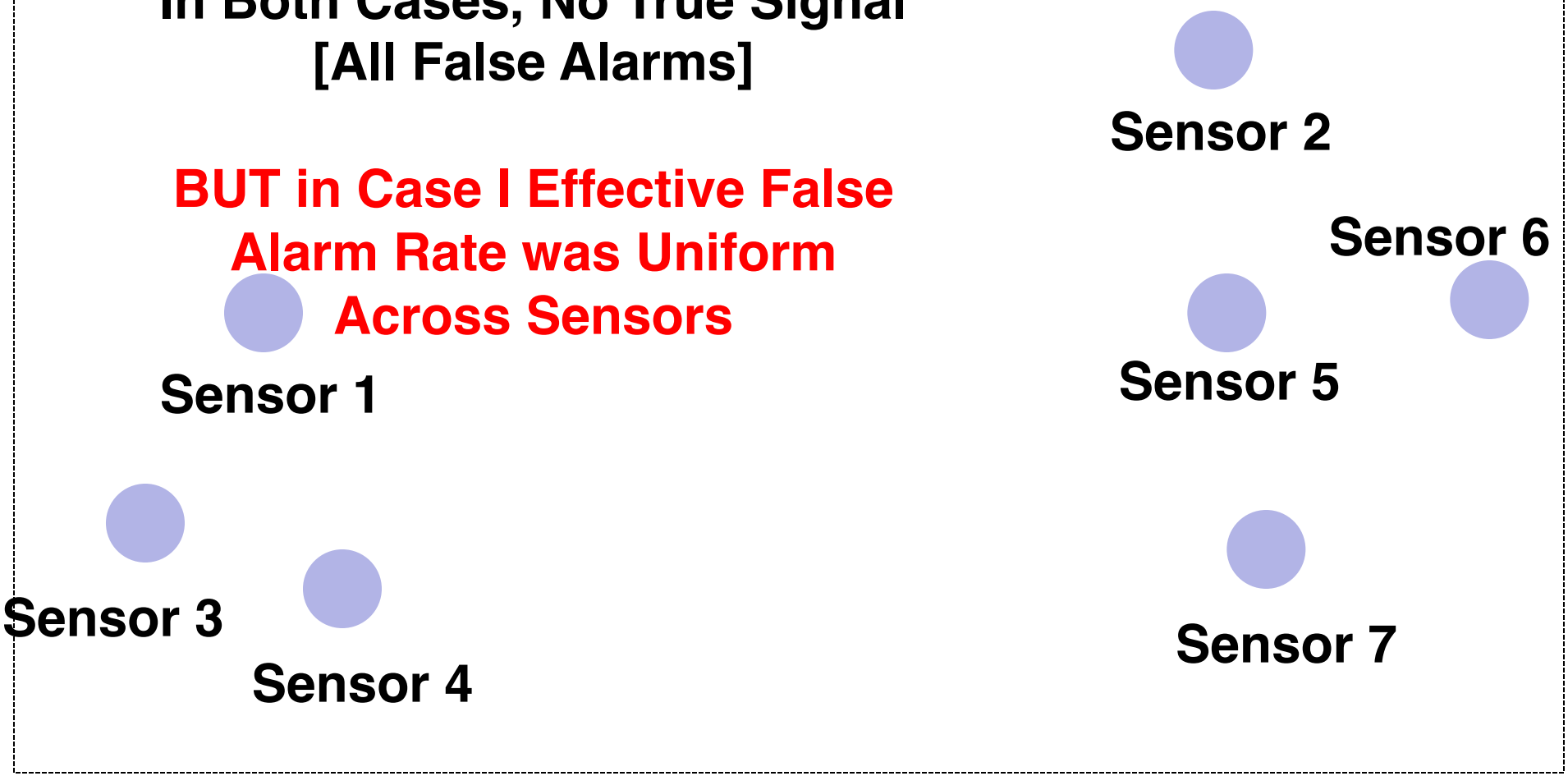
Case II (Identical Point Sensors BUT….)



Case II vs Case I

In Both Cases, No True Signal [All False Alarms]

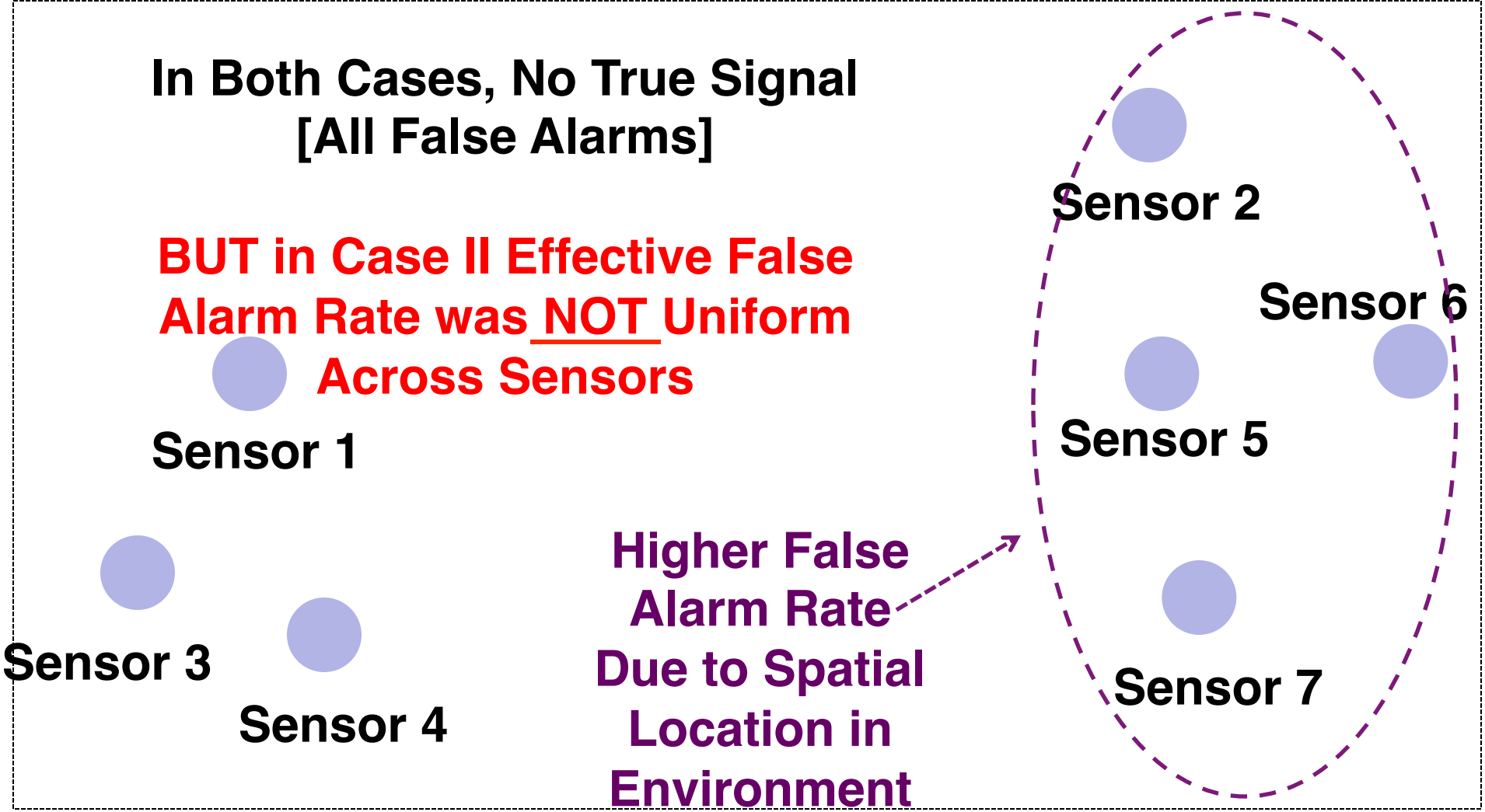
BUT in Case I Effective False Alarm Rate was Uniform Across Sensors



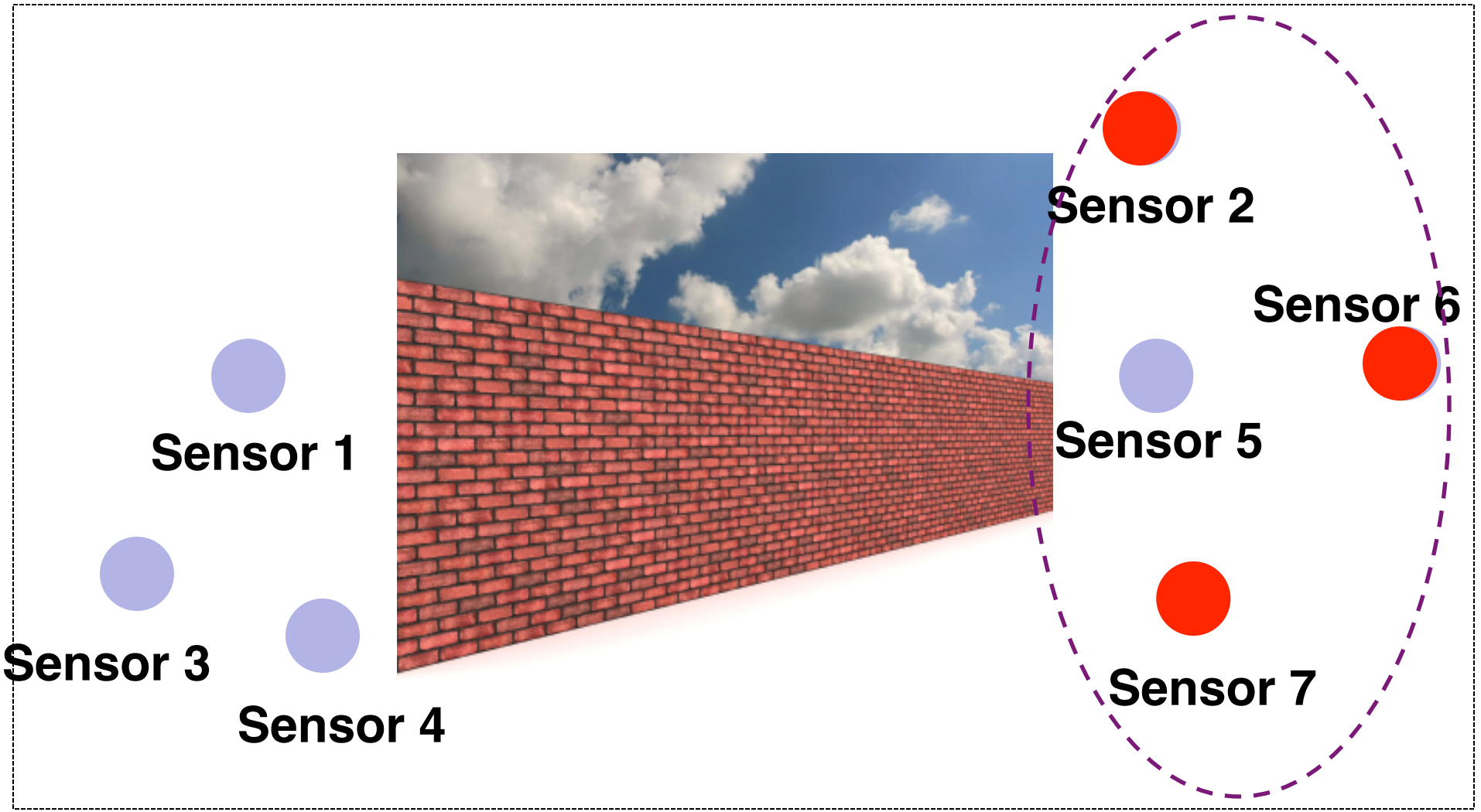
Case II vs Case I

In Both Cases, No True Signal [All False Alarms]

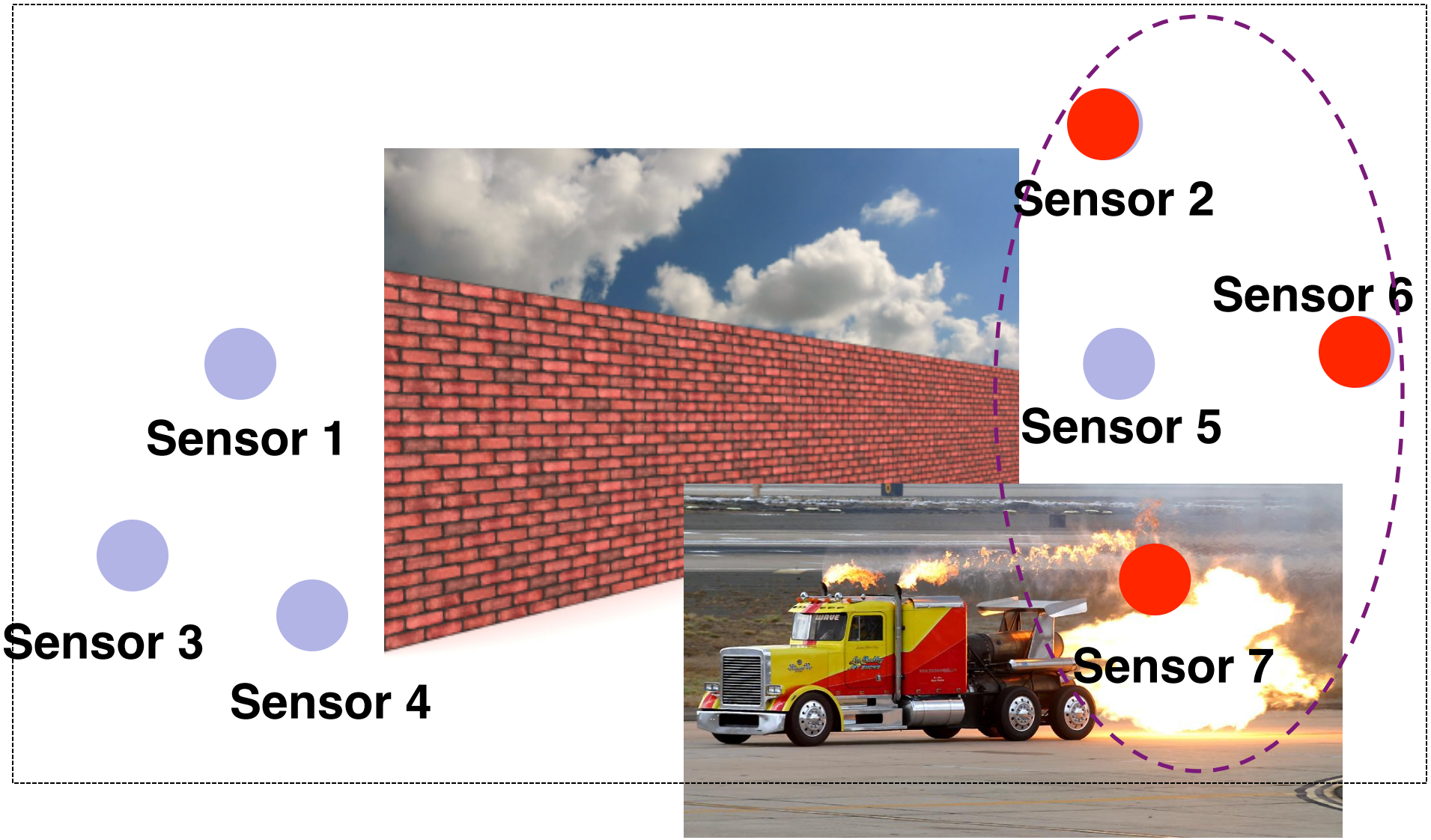
BUT in Case II Effective False Alarm Rate was NOT Uniform Across Sensors



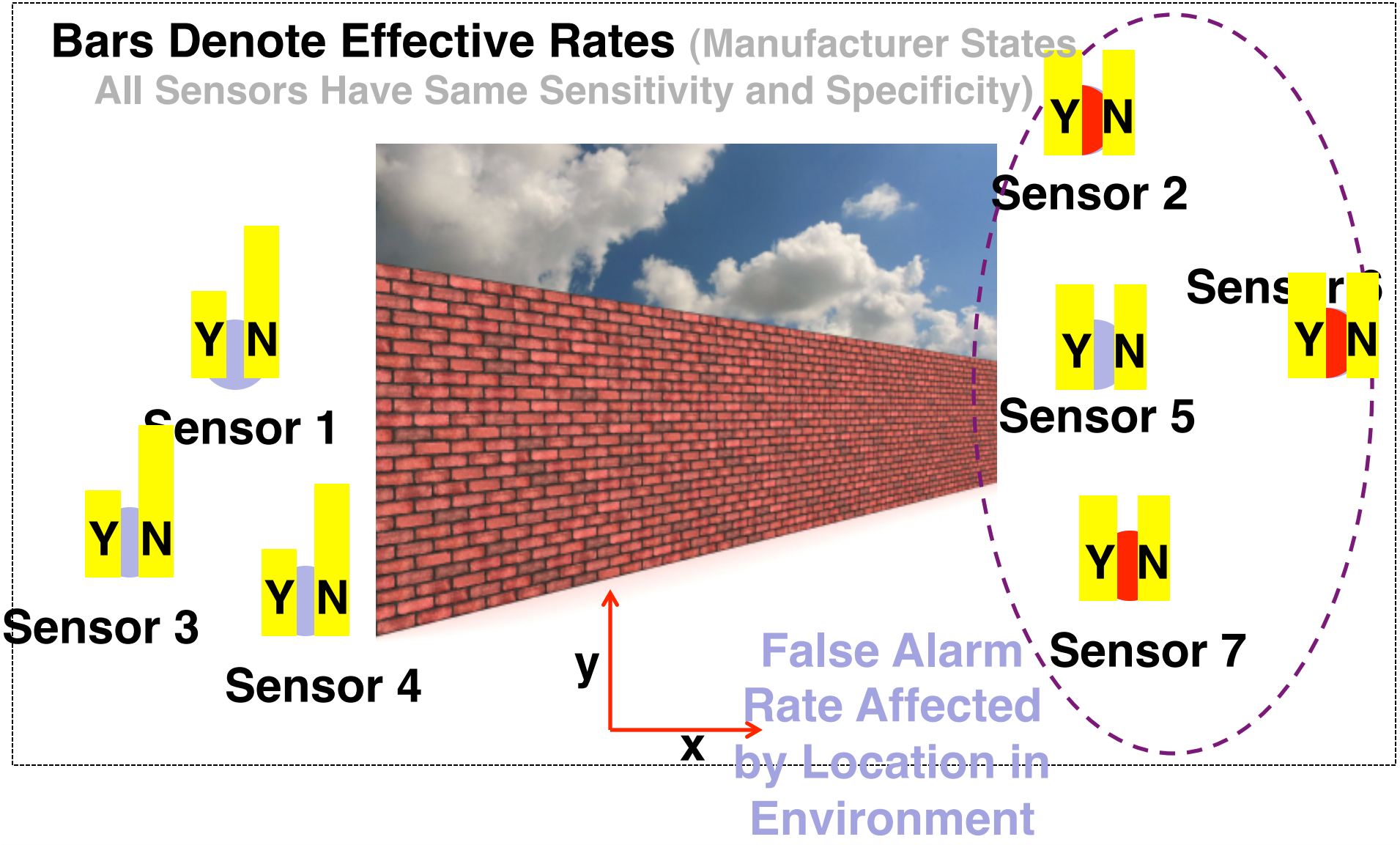
Case II (Identical Point Sensors BUT…. False Alarm Rate Affected by Environment)



Case II (Identical Point Sensors BUT…. False Alarm Rate Affected by Environment)



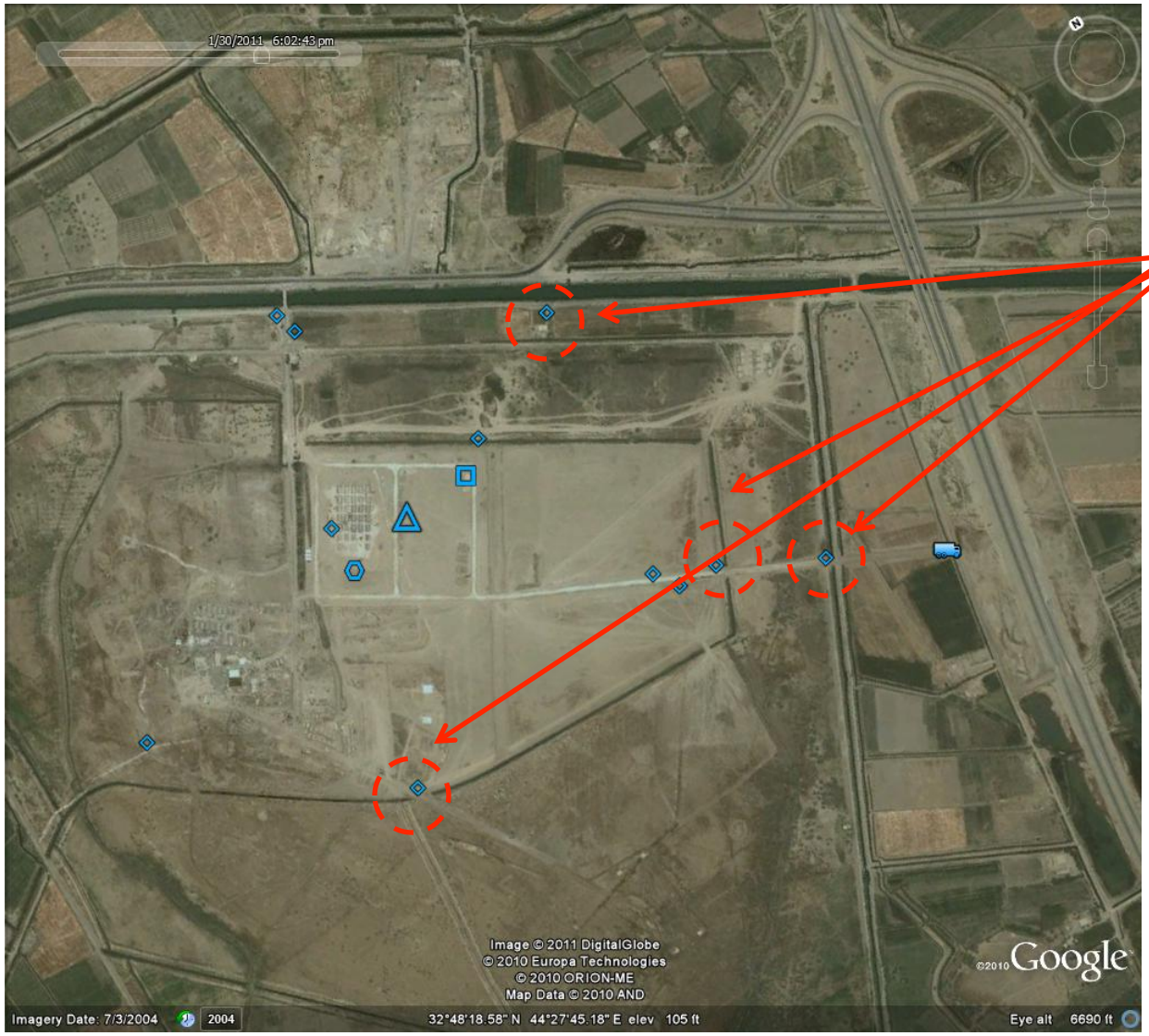
Case II (Identical Point Sensors BUT….)



Make Better Use of Existing Resources

1. Frequent Positives (False Alarms) May Causes Operators to Ignore Sensor Warnings.
2. The Algorithms We Propose Aim at Respecting the Features Specific to a Sensor Configuration in a Real Operating Environment (Face Complications Outside of a Controlled Lab Testing Situation).
3. We Stress that the Algorithms Utilize Field Data and Attempt to Continually Calibrate and Monitor Sensors at Individual and Network Levels (Both Accuracy and Uncertainty)
4. Focus on Collection of Different Sensors Under Various Conditions

A More Realistic Sensor Layout



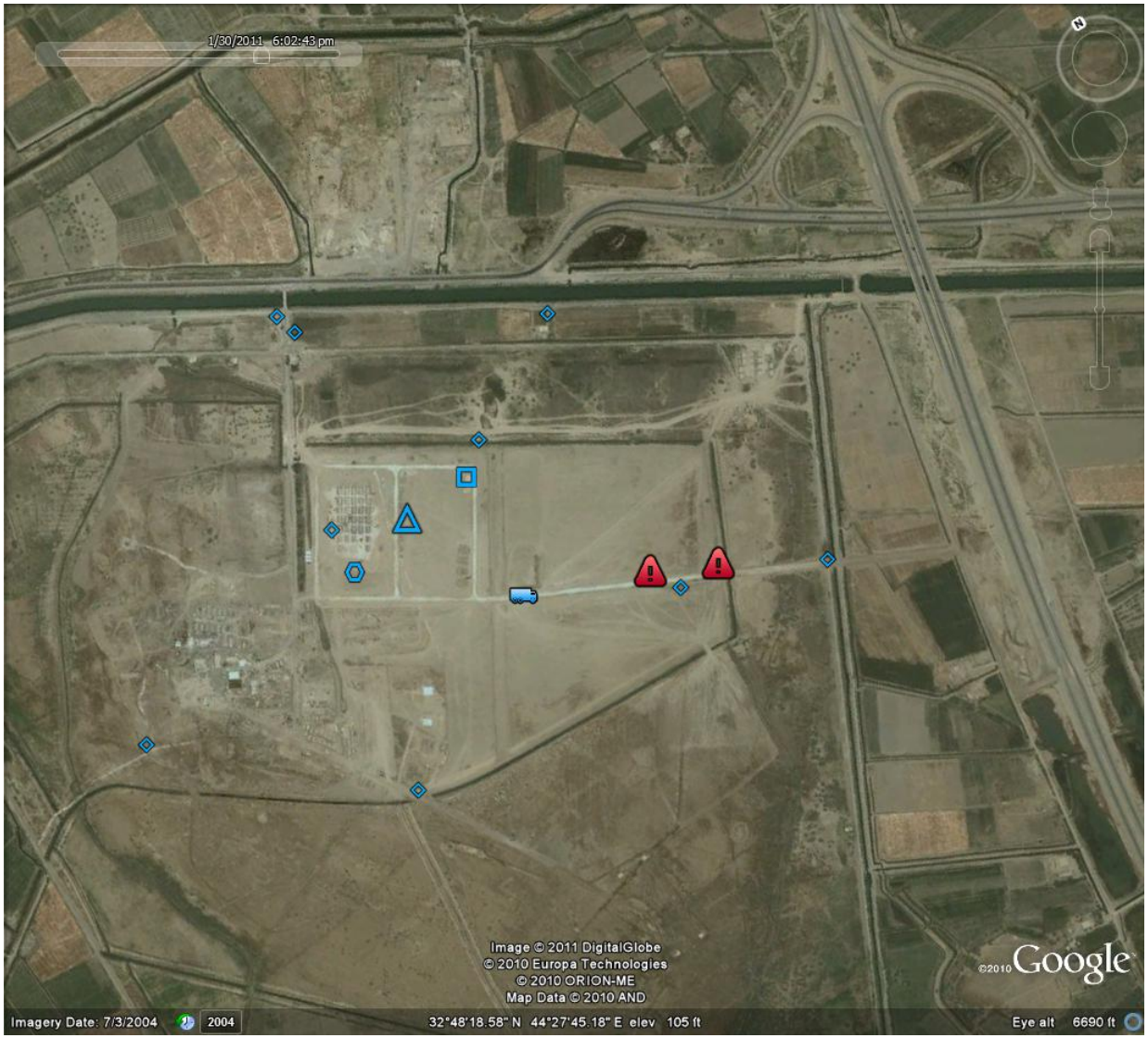
Sensors

A More Realistic Sensor Layout

A More Realistic Sensor Layout

A More Realistic Sensor Layout

A More Realistic Sensor Layout



A More Realistic Sensor Layout

Complications We Face And Should Address to Mitigate/Reduce False Alarms

1. Nominal Rates of Manufacturer Likely "Imperfect"
   Proportion Estimation

2. A Variety of Sensor Types Deployed (Blessing and a Curse)
   Heterogeneous Sensor Network

3. Fusion of Multiple Sensor Readings
   Multiple Hypothesis Testing

4. Sensor Network Posses Complex Spatial and Temporal Dependencies (Application Specific)

5. Environment Conditions Change with Time

“On the Fly” Calibration of Sensors

Nominal vs. Actual Sensor Performance in Field

Setup Notation and *Quickly* Go Through Estimation Problem in Binary Signal Context
(Many of Our Sensors Are Categorical But Problem Easiest to Describe in Binary Setting).

Output of Calibration Fed To Multiple Hypothesis Testing or “Fusion” Algorithms

Measurement Error Notation

Truth (X) \ Sensor Measurement (W)	1	0
1	**Sensitivity** $\mathbb{P}(W = 1 \mid X = 1)$ **What You See is What You Get**	**Type II Error / False Negative** $\mathbb{P}(W = 0 \mid X = 1)$ **The "Worst Sin"**
0	**Type I Error / False Positive** $\mathbb{P}(W = 1 \mid X = 0)$ **A Burden We Carry if We Don't "Sin" Often**	**Specificity** $\mathbb{P}(W = 0 \mid X = 0)$ **What You See is What You Get**

Binary Sensor Signals

Truth (X)	Sensor Measurement (W) 1	Sensor Measurement (W) 0
1	**Sensitivity** $\mathbb{P}(W = 1 \mid X = 1) \equiv \theta_{1\mid 1}$	**Type II Error / False Negative** $\beta = 1 - \theta_{1\mid 1}$
0	**Type I Error / False Positive** $\alpha = 1 - \theta_{0\mid 0}$	**Specificity** $\mathbb{P}(W = 0 \mid X = 0) \equiv \theta_{0\mid 0}$

Binary Sensor Signals

Truth (X) \ Sensor Measurement (W)	1	0
1	**Sensitivity** $\theta_{1\|1}$	**Type II Error / False Negative** $\beta = 1 - \theta_{1\|1}$
0	**Type I Error / False Positive** $\alpha = 1 - \theta_{0\|0}$	**Specificity** $\theta_{0\|0}$

It is a Difficult Task to Design A Sensor with Accurate Specificity. Our Aim: "Tune" Using Network Information

Binary Sensors Signals

Truth (X) \ Sensor Measurement (W)	1	0
1	**Sensitivity** $\theta_{1\|1}$	**False Negative** $\beta = 1 - \theta_{1\|1}$
0	**False Positive** $\hat{\alpha} = 1 - \hat{\theta}_{0\|0}$	**Specificity** $\hat{\theta}_{0\|0}$

Our Focus: Estimate (and Assess Performance/Uncertainty) Using Both "Bad" and Good Sensor Readings

Misclassification / Errors in Variables

Complications Arise When Sensitivity and Specificity are Not Precisely Known. We Will Utilize Specificity Estimates: $\hat{\theta}_{0|0}$

Specificity for "Point" (e.g., Ion Mobility Sensor) and "Stand-Off" Sensors (Video Imaging) Vary and Depend on Environment

We Usually Assume Manufacturers Estimate of Sensitivity, $\hat{\theta}_{1|1}$, is "Within Spec" but Still Utilize Uncertainty (if provided)

Misclassification / Errors in Variables

1) Collect Streams of 1's and 0's for Length "N" Coming From Usual Sensor (Select "N" via Edgeworth Expansions)

Brown, L., Cai, T., and Dasgupta, A. *Annals of Statistics* **30**, 160–201 (2002).

2) Still Practice Standard Safety Procedure for Alarm (i.e. Verify No Threat Condition; i.e., Check "Control Case").

3) Estimate $\hat{\theta}_{0|0}$ Using "Internal" or "External Data"

4) Utilize Training Data to Generate Collection of Prevalence Estimates $\hat{\pi}_X$

5) Form Various Test Statistics (Using Different Proxies of SE, Specificity, and Sensitivity) of the Form $\longrightarrow \dfrac{\hat{\pi}_X}{SE[\hat{\pi}_X]}$

Misclassification / Errors in Variables

5) Form Various Test Statistics (Using Different Proxies of SE, Specificity, and Sensitivity) of the Form $\frac{\hat{\pi}_X}{SE[\hat{\pi}_X]}$

Analytic Results/Approximations in Measurement Error Methods Useful.

Buonaccorsi, J. *Measurement Error: Models, Methods, and Applications.* Chapman and Hall/CRC, Boca Raton, FL, (2010).

Test Stat Provides "Metric" Which Can Be Used to Quantitatively Compare Different Sensors Under "Null" (Facilitates Fusion of Different Sensor Types)

Misclassification / Errors in Variables

Naïve Estimates Known to Be Biased in Independent and Identically Distributed (i.i.d.) Setting, e.g.

$$\pi_W := \mathbb{E}\left[p_W := \frac{1}{N}\sum_{i=1}^{N} W_i\right] \neq \pi_X := \mathbb{E}\left[p_X := \frac{1}{N}\sum_{i=1}^{N} X_i\right]$$

$$\pi_W - \pi_X = \pi_X(\theta_{1|1} + \theta_{0|0} - 2) + (1 - \theta_{0|0})$$

If Sensitivity and Specificity Known Precisely, Unbiased Estimators and Hypothesis Tests Can Be Constructed.

$$\hat{\pi}_X = \frac{p_W - (1 - \theta_{0|0})}{\theta_{0|0} + \theta_{1|1} - 1}$$

Point Estimate

$$SE(\hat{\pi}_X) = \left(\frac{p_W(1 - p_W)}{N(\theta_{0|0} + \theta_{1|1} - 1)^2}\right)^{1/2}$$

Large Sample "Uncertainty" of Estimate

Misclassification / Errors in Variables

For Usual Data (Non-Threat) We Know X=0, and for Population of Binary Responses $\pi_X \approx 0$

Utilize Usual Data to Construct Point Estimates and "Diagnostic" Hypothesis Tests

$$\hat{\pi}_X = \frac{p_W - (1 - \hat{\theta}_{0|0})}{\hat{\theta}_{0|0} + \hat{\theta}_{1|1} - 1}$$

Ratio of Two Inherently Noisy Quantities.

Unambiguous "Metrics" Need to Account for this Uncertainty

Buonaccorsi, J. *Measurement Error: Models, Methods, and Applications.* Chapman and Hall/CRC, Boca Raton, FL, (2010).

Different Unbiased Proxies of SE

"Wald-Delta"

$$\hat{V}_1 = \frac{p_W(1-p_W)}{N} + \frac{\hat{\theta}_{0|0}(1-\hat{\theta}_{0|0})}{n_0}$$

$$\hat{V}_2 = \frac{\hat{\theta}_{1|1}(1-\hat{\theta}_{1|1})}{n_1} + \frac{\hat{\theta}_{0|0}(1-\hat{\theta}_{0|0})}{n_0}$$

$$\hat{V}_{12} = \frac{\hat{\theta}_{0|0}(1-\hat{\theta}_{0|0})}{n_0}$$

$$\hat{V} = \frac{(\hat{V}_1 - 2\hat{\pi}_X \hat{V}_{12} + \hat{\pi}_X{}^2 \hat{V}_2)}{(\hat{\theta}_{0|0} + \hat{\theta}_{1|1} - 1)^2}$$

$$SE^{C2} = \sqrt{V}$$

Fieller CI

$$A = p_W - (1-\hat{\theta}_{0|0}); B = \hat{\theta}_{0|0} + \hat{\theta}_{1|1} - 1$$

$$\hat{\pi}_X = \frac{A}{B}$$

$$f_1 = A^2 - Z^2_{\alpha/2}\hat{V}_1$$

$$f_2 = B^2 - Z^2_{\alpha/2}\hat{V}_2$$

$$f_{12} = AB - Z^2_{\alpha/2}\hat{V}_{12}$$

$$f = f_{12}^2 - f_1 f_2$$

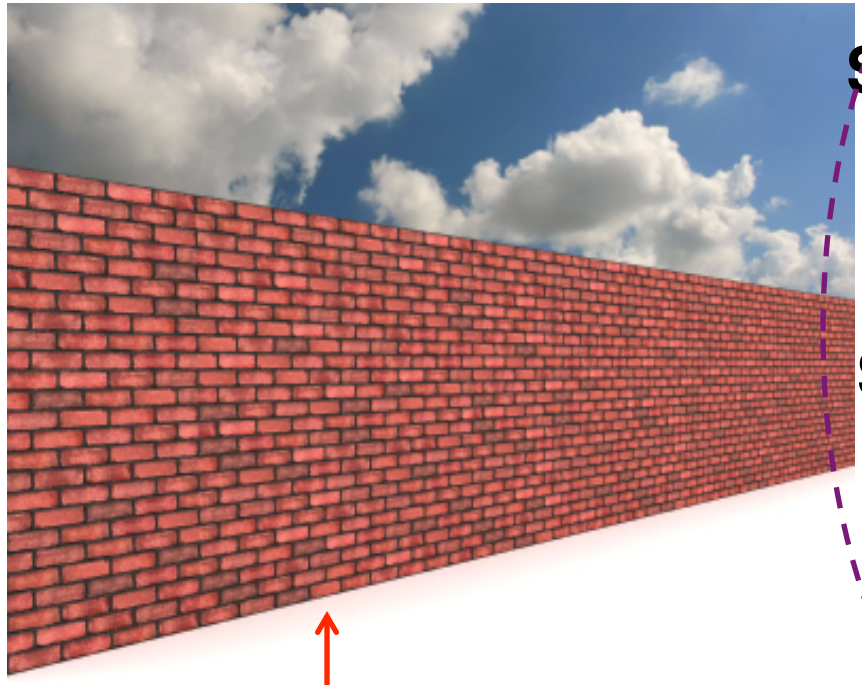
$$L_f = \frac{f_{12} - f^{1/2}}{f_2}; U_f = \frac{f_{12} + f^{1/2}}{f_2}$$

$$CI^{(1-\alpha)}_{\hat{\pi}_X} = (\max(L_f, 0), \min(U_f, 1))$$

$$SE^{C3} = \sqrt{CI^{(1-\alpha)}_{\hat{\pi}_X}}$$

Making Case II Comparable to Case I

Normalize to Make Binary Output Statistics Comparable (Back to Case I)



Sensor 1

Sensor 2

Sensor 3

Sensor 4

Sensor 5

Sensor 6

Sensor 7

y

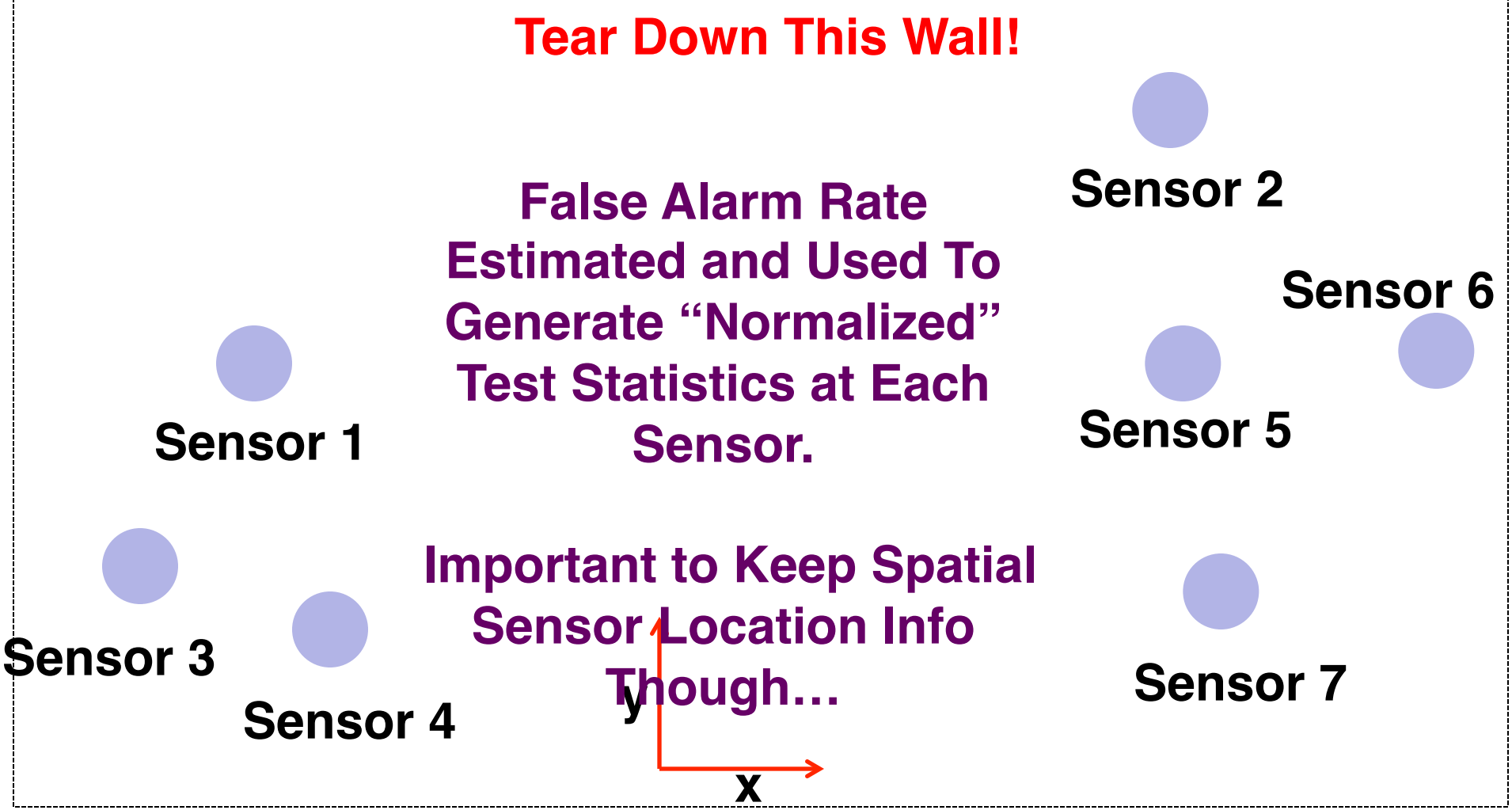
x

Making Case II Comparable to Case I

Tear Down This Wall!

False Alarm Rate Estimated and Used To Generate "Normalized" Test Statistics at Each Sensor.

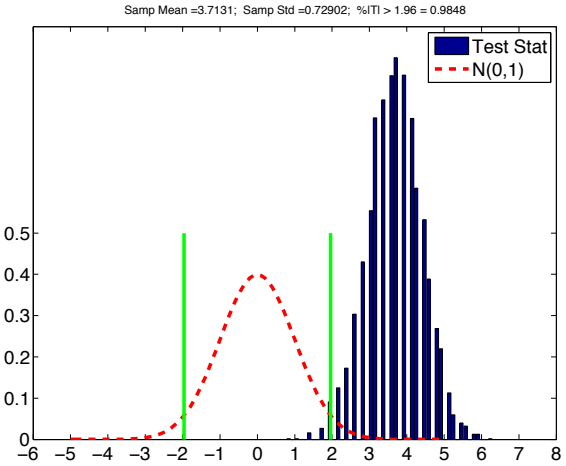
Important to Keep Spatial Sensor Location Info Though…



Test Statistic Distributions In Simulation

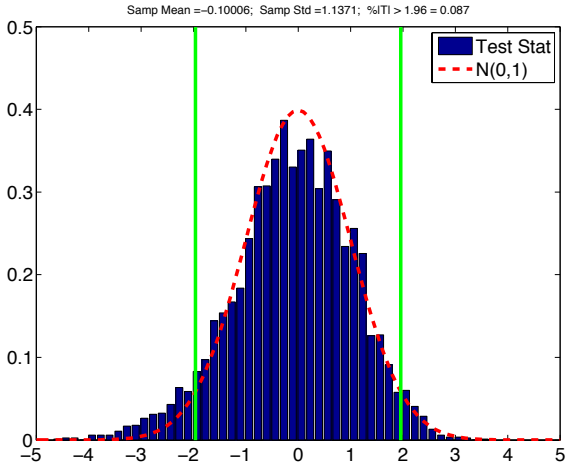
(True Rates Known But Only Nominal Provided to Estimators)

Naive



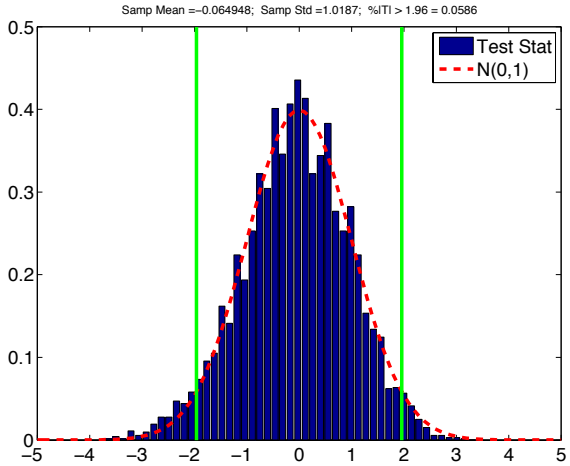
Test Stat Using Nominal False Alarm Rate

"Wald-Delta"



Test Stat Using "Field Estimate" False Alarm Rate and SE Est. 1

Fieller



Test Stat Using "Field Estimate" False Alarm Rate and SE Est. 2

Estimated Model With Uncertainty is Fairly Close to Large Sample $\mathcal{N}(0,1)$ Limit. However Finite Sampling Effects Still Detectable

Towards Multiple Testing (Fusion)

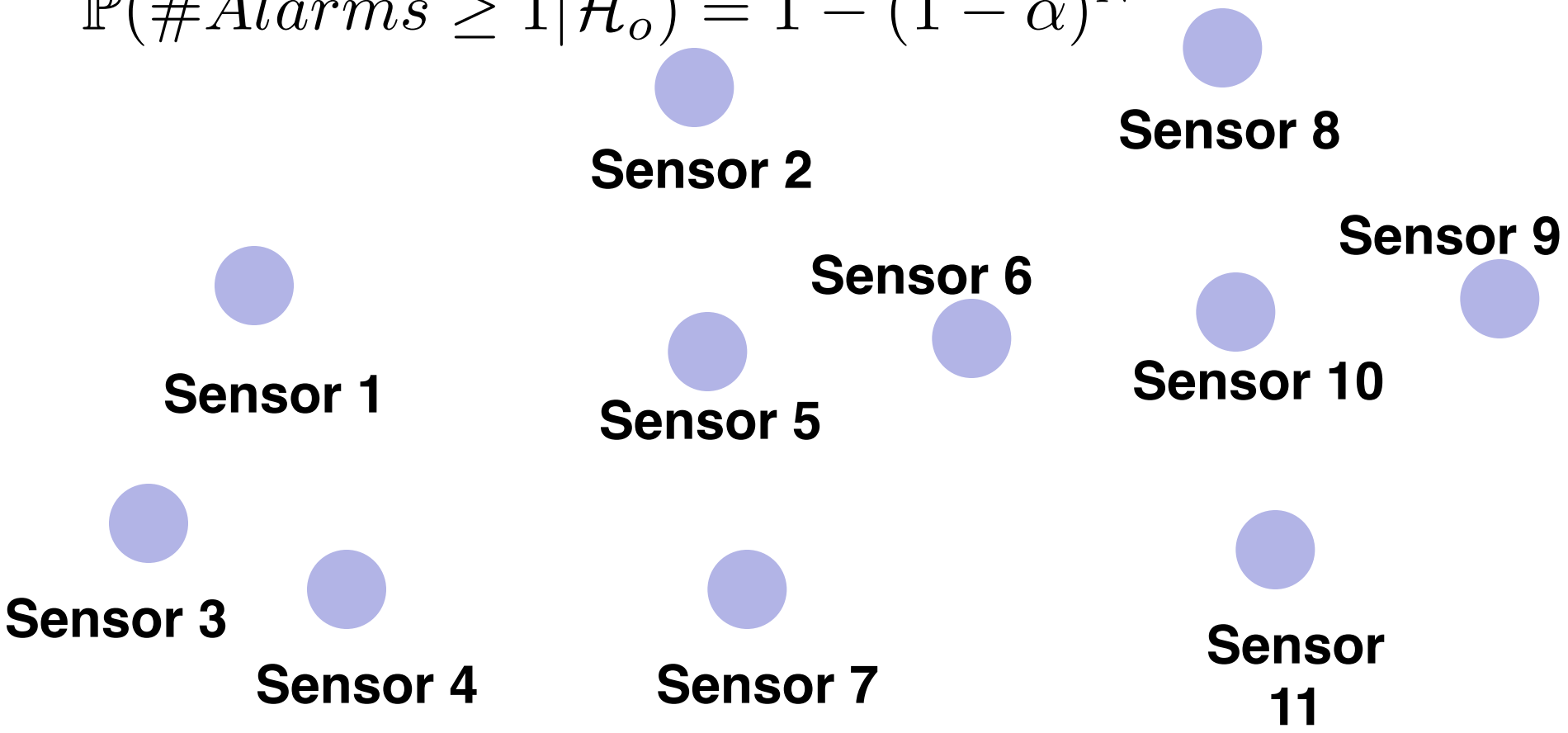
Calibrate at Each Sensor. Useful to Plot $\hat{\theta}_{0|0}$
As Function of Location for Each Sensor (Identify Potential Environment Queues Causing False Alarms)

Aim at Achieving an i.i.d. $\mathcal{N}(0,1)$ Test Statistic Distribution At Single Sensor Level. If Achieved, We Can Do Simultaneous Inference and More Readily Control the Error Rate, e.g. via Family Wise Error Rate [FWER] or False Discovery Rate [FDR] Methods

Recall the Standard Example

In Simple i.i.d. Case with (Uniform False Positive Rate)

$$\mathbb{P}(\#Alarms \geq 1 | \mathcal{H}_o) = 1 - (1-\alpha)^N$$



Multiple Testing Approaches to Problem

FWER Methods Address This But Are Not Designed To Control Our Lethal "Sin" (Type II Errors). Said Differently: "They Don't Scale Well". Though Defining "Large" is Nontrivial in Presence of Dependence

Efron, B. (w/ Discussion Contributed by Cai, T., Heller, R., Schwartzman, A. and Westfall, P.) *JASA* **105**(491), 1042–1067 (2010).

FDR Methods Commit Fewer Sins At Cost of Higher Type I Error Rate (Compared to FWER…BUT perhaps too liberal?)

Benjamini, Y. and Hochberg, Y. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**(1), 289–300 (1995).

We Feel FDR Shows Great Promise When Combined With Queues in Large Scale Surveillance Applications…BUT we need to test with field data.

We Are Now in the Data Deluge….

FDR Shows Great Promise When Combined With Queues (e.g., a Truck Passing) in Surveillance Applications.

Benjamini, Y. and Hochberg, Y. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**(1), 289–300 (1995).

Efron, B. *J. American Statistical Association* **105**(491), 1042–1055 (2010).

Muralidharan, O. *Annals of Applied Statistics* **4**(1), 422–438 (2010).

But It May Not Make "Best" Use of Data

(Depends Heavily on Optimality Criterion)

Westfall, P. H. *Statistica Sinica* **18**(3), 811–816 (2008).

Poor, H. V. and Hadjiliadis, O. *Quickest Detection.* Cambridge University Press, (2008).

Stand-off Sensors

We Illustrated Binary Sensor Calibration With "Point Sensors"

But Methods Also Applicable to "Stand-off Sensors"

Accurately Estimating Sensitivity in Environment Specific Context is Even More Important With These Sensor Types…… AND

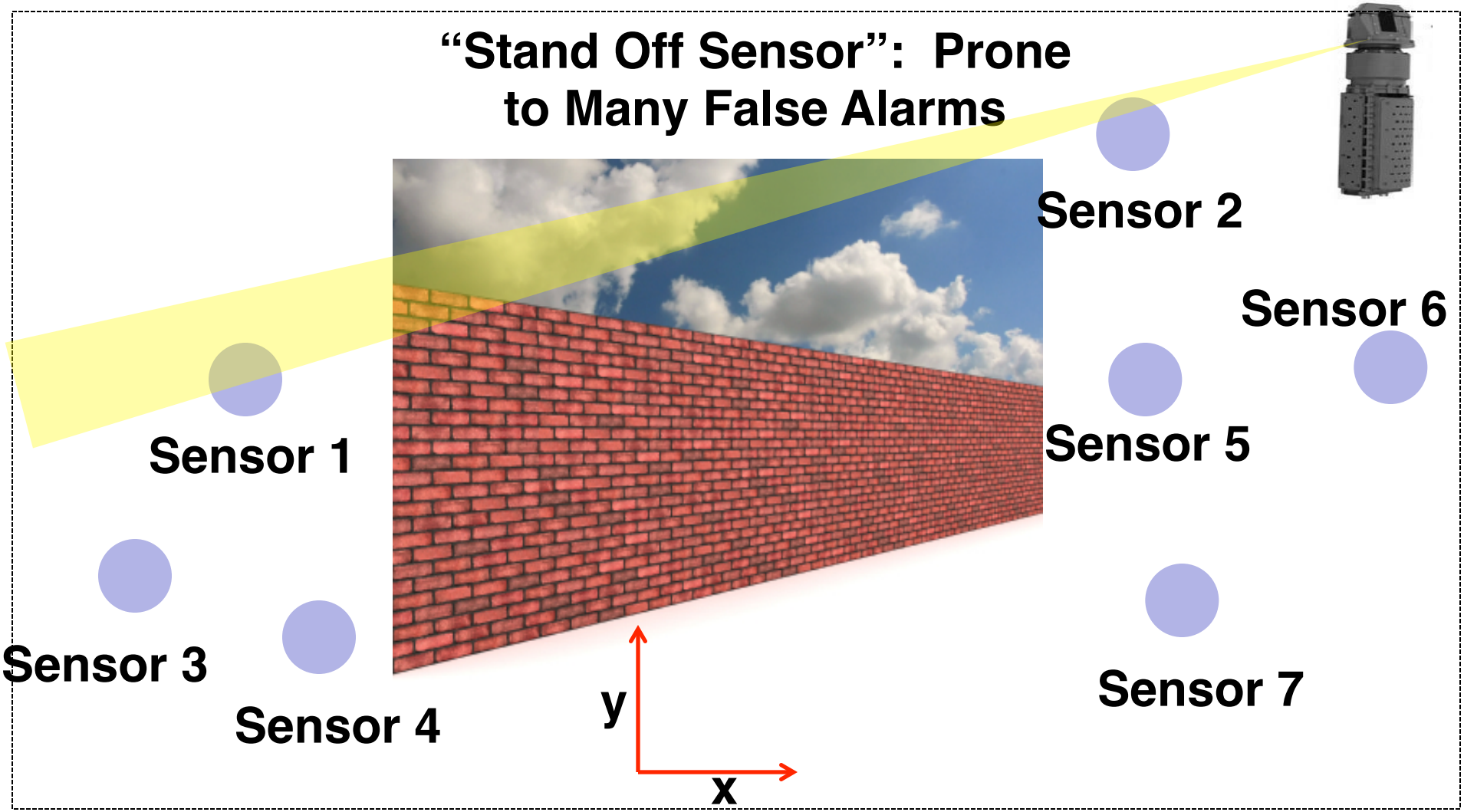
Stand-off Sensors

We Illustrated Binary Sensor Calibration With "Point Sensors"

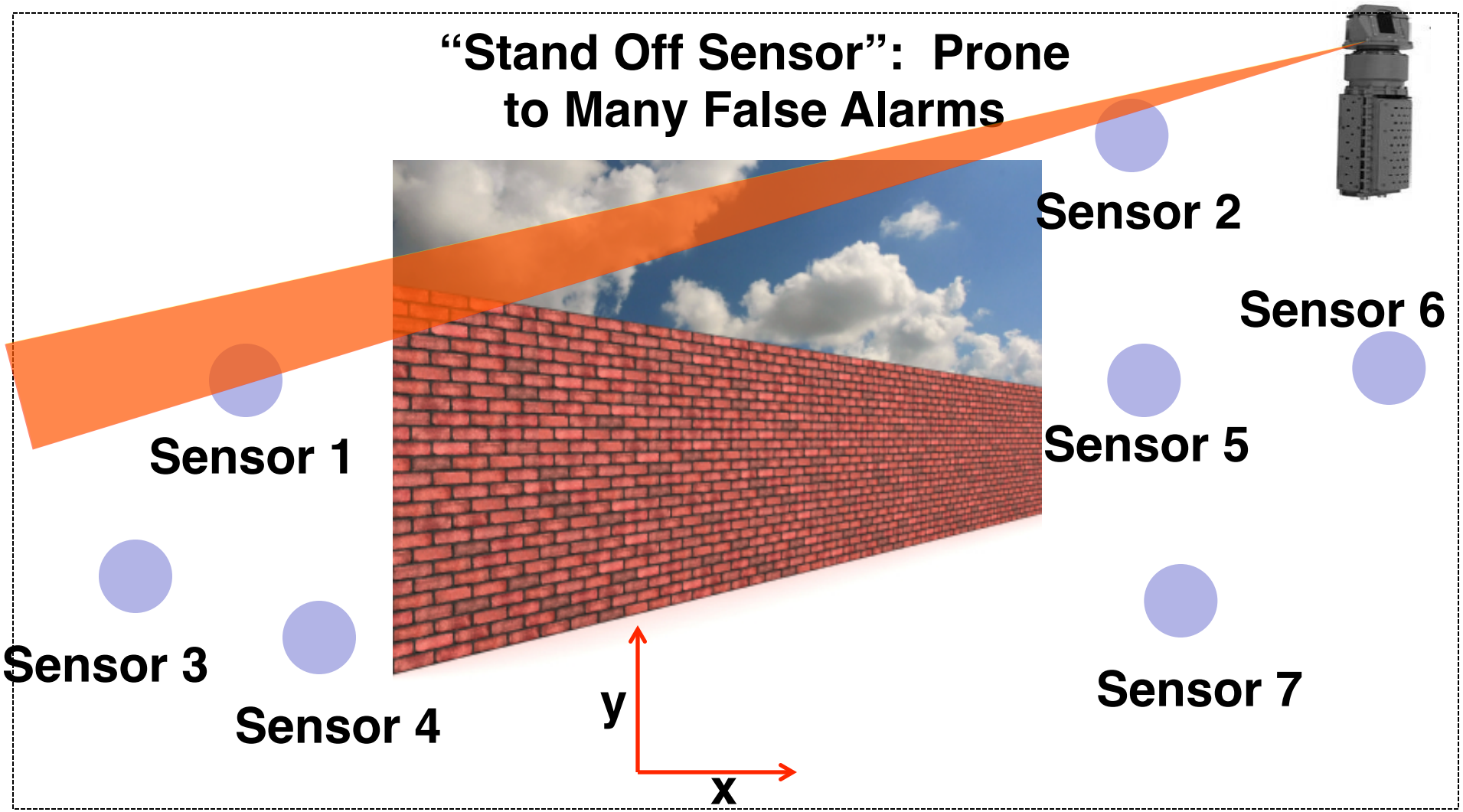
But Methods Also Applicable to "Stand-off Sensors"

Accurately Estimating Sensitivity in Environment Specific Context is Even More Important With These Sensor Types…… AND

Mixed Sensor Types

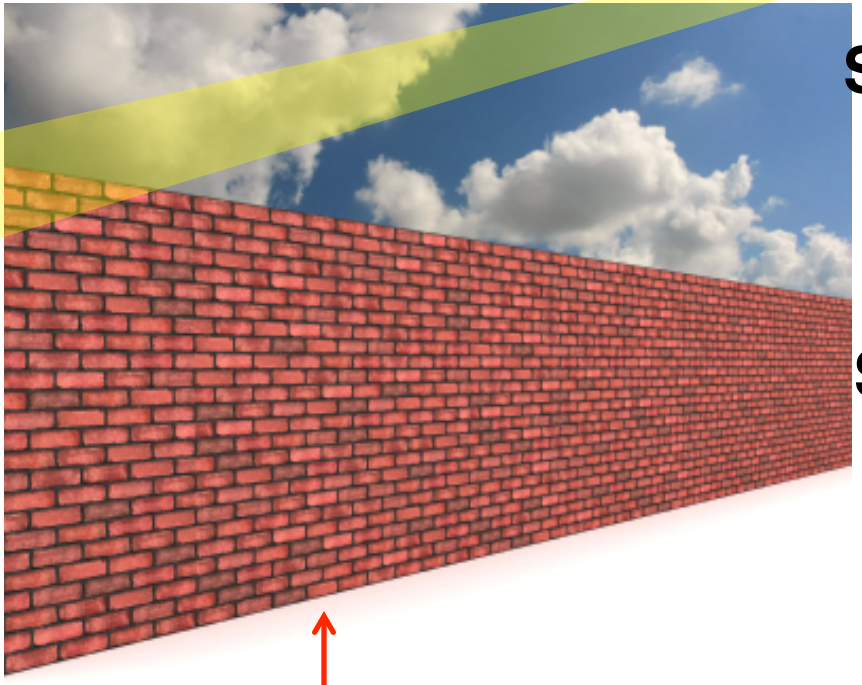


Mixed Sensor Types



Overlapping Coverage Regions

Fuse Sensors of Different Quality and Exploit Overlapping Coverage



Sensor 1

Sensor 2

Sensor 3

Sensor 4

Sensor 5

Sensor 6

Sensor 7

y

x

Mixed Sensor Types

Fuse Sensors of Different Quality and Exploit Overlapping Coverage

$$
\begin{aligned}
Z_1' &= (Z_1 + S)/\sqrt{2} \\
Z_2' &= (Z_2 + S)/\sqrt{2} \\
Z_3' &= Z_3 \\
Z_4' &= Z_4 \\
Z_5' &= Z_5 \\
Z_6' &= Z_6 \\
Z_7' &= Z_7
\end{aligned}
$$

Sensor 1

Sensor 2

Sensor 3

Sensor 4

Sensor 5

Sensor 6

Sensor 7

Mixed Sensor Types

Fuse Sensors of Different Quality and Exploit Overlapping Coverage

Sensor 1

Sensor 2

Sensor 3

Sensor 4

Sensor 5

Sensor 6

Sensor 7

$$
\begin{aligned}
Z_1' &= (Z_1 + S)/\sqrt{2} \\
Z_2' &= (Z_2 + S)/\sqrt{2} \\
Z_3' &= Z_3 \\
Z_4' &= Z_4 \\
Z_5' &= Z_5 \\
Z_6' &= Z_6 \\
Z_7' &= Z_7
\end{aligned}
$$

Mixed Sensor Types

Sensor 1

Sensor 2

Sensor 3

Sensor 4

Sensor 5

Sensor 6

Sensor 7

$$Z_1' = (Z_1 + S)/\sqrt{2}$$
$$Z_2' = (Z_2 + S)/\sqrt{2}$$
$$Z_3' = Z_3$$
$$Z_4' = Z_4$$
$$Z_5' = Z_5$$
$$Z_6' = Z_6$$
$$Z_7' = Z_7$$

"Signal Amplification" or "Enrichment"

Efron, B. *Annals of Applied Statistics* **2**(1), 197–223 (2008).

Efron, B. *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction.* Cambridge University Press, Cambridge, (2010).

Mixed Sensor Types

Multiple Comparisons with N(0,1) Distribution… but Independence Lost

$$Z_1' = (Z_1 + S)/\sqrt{2}$$
$$Z_2' = (Z_2 + S)/\sqrt{2}$$
$$Z_3' = Z_3$$
$$Z_4' = Z_4$$
$$Z_5' = Z_5$$
$$Z_6' = Z_6$$
$$Z_7' = Z_7$$

Sensor 1

Sensor 2

Sensor 3

Sensor 4

Sensor 5

Sensor 6

Sensor 7

"Signal Amplification" or "Enrichment"

Efron, B. *J. American Statistical Association* **105**(491), 1042–1055 (2010).

"Large" vs. "Small" Scale Inference

$$
\begin{aligned}
Z'_1 &= (Z_1 + S_2)/\sqrt{2} \\
Z'_2 &= (Z_2 + S_1 + S_2)/\sqrt{3} \\
Z'_3 &= Z_3 \\
Z'_4 &= (Z_4 + S_1)/\sqrt{2} \\
Z'_5 &= Z_5 \\
Z'_6 &= Z_6 \\
Z'_7 &= Z_7 \\
Z'_8 &= S_1 \\
Z'_9 &= S_2 \\
Z'_{10} &= S_3 \\
Z'_{11} &= S_4
\end{aligned}
$$

S_4

S_3

S_2

S_1

Sensor 1

Sensor 2

Sensor 3

Sensor 4

Sensor 5

Sensor 6

Sensor 7

Efron, B. *J. American Statistical Association* **105**(491), 1042–1055 (2010).

“Large” vs. “Small” Scale Inference

Fielded “Stand-off” sensors scan quickly over time & space.

Is it better to use a fine grid and count every unique spatial observation as a single observation and coarsely deal with dependence (e.g. root mean square correlation)?

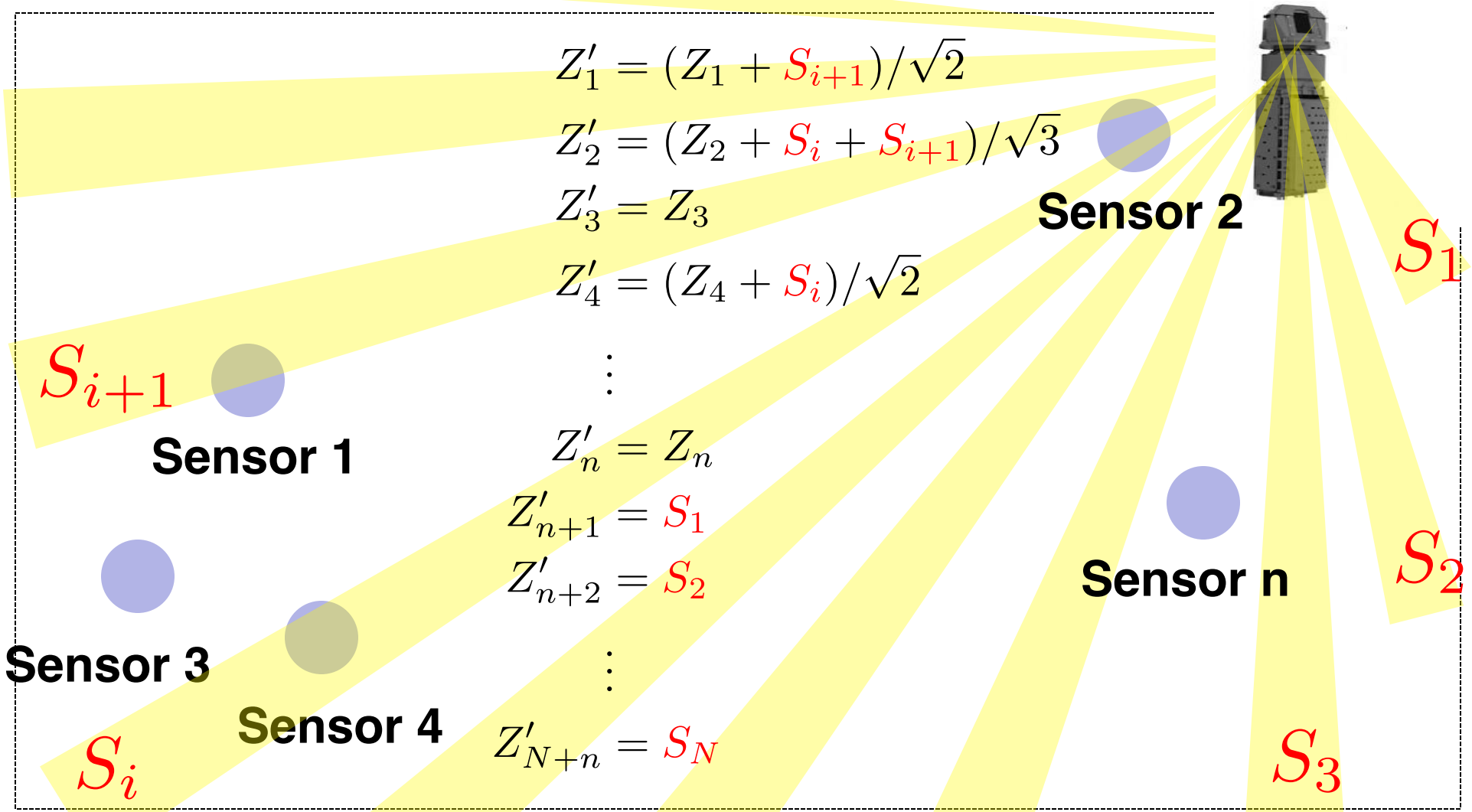
Efron, B. *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction.* Cambridge University Press, Cambridge, (2010).

OR

Is it better to only count “Stand-off” measurement overlapping with other sensors and lump everything other Stand-off into one coarse category (“small scale” inference and real-time resampling methods more accurately modeling correlation)

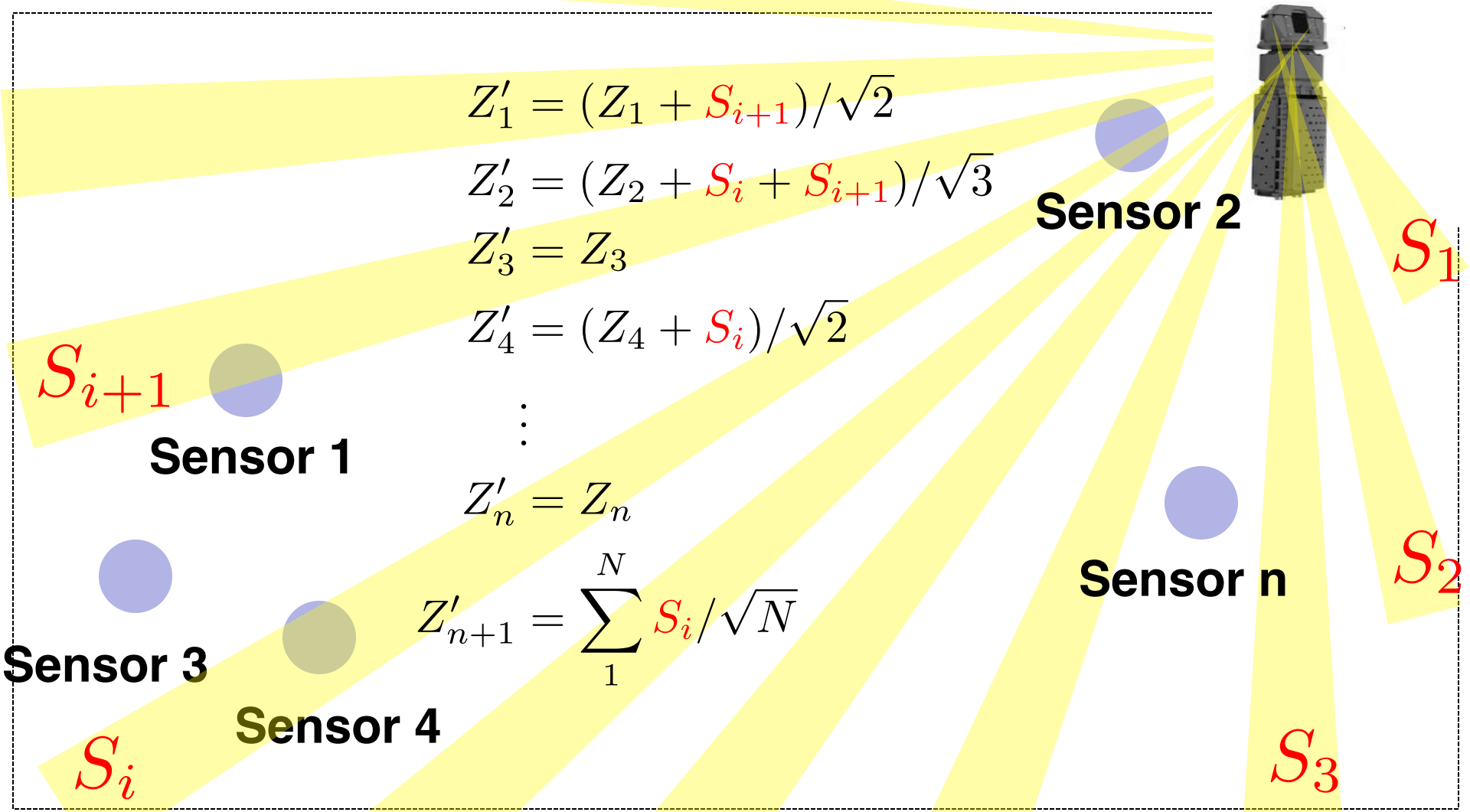
Westfall, P. H. and Young, S. S. *Resampling-based multiple testing: examples and methods for p-value adjustment.* Wiley, New York, (1993).

"Large" Scale Inference



Efron, B. *J. American Statistical Association* **105**(491), 1042–1055 (2010).

"Small" Scale Inference



Efron, B. *J. American Statistical Association* **105**(491), 1042–1055 (2010).

Conditioning on Events (or "Queues")

Empirical Bayes approach to "conditioning on events" (or observable covariates) is attractive in several "fusion" applications.

Genovese, C. R., Roeder, K., and Wasserman, L. *Biometrika* **93**(3), 509–524 (2006).

Efron, B. *Annals of Applied Statistics* **2**(1), 197–223 (2008).

Cai, T. T. and Sun, W. *Journal of the American Statistical Association* **104**(488), 1467–1481 (2009).

Hu, J. and Zhao, H. *Journal of the American Statistical Association* **06511**, 1–40 (2010).

However, other P-value weighting (e.g., "enriched" signals should be allowed "more say") schemes may assist

Westfall, P. H., Krishen, a., and Young, S. S. *Statistics in medicine* **17**(18), 2107–19 (1998).

Bottom Line: We want a *Systematic Method* for tuning sensor networks with minimal "knobs" but need to retain power in realistic scenarios where environment is not stationary (infrequent regime shifts).

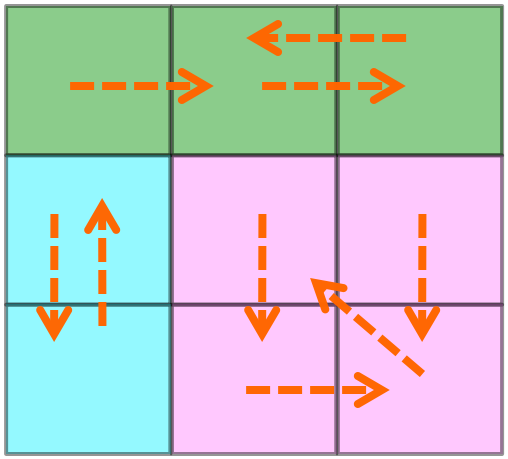
Summary Points

- Sensors Lie. Deal With It (Unrealistic to Assume They Won't "Lie").
- We Do Not Have the Luxury of Observing "Cases". Statistics of (Hopefully) Rare Events Unknown (Severely Complicates ROC and Methods Using Priors or Odds).
- By Training Different Categorical Sensors (Varying in Type and/or Quality). We Can Assist in Making Sensors Statistics Comparable.
- The Above Shows Promise in Heterogeneous Sensor Fusion (Queues and *A Priori* Known Spatial Information Can Assist This Task).
- Discussed Practical Complications Associated With Correlation and Conditioning BUT We Are Exploring Several Options. Opinions from Different "Sects" of the Statistics Community VERY Welcome!

Backup Slides

Should We Be More Ambitious About Known Spatial Information?

Form Clusters with the Hope of Increasing Power (Improve Signal to Noise Ratio of Test Stat). A Type of "Dimension Reduction" But Several Practical Questions Remain…



Heller, R., Stanley, D., Yekutieli, D., Rubin, N., and Benjamini, Y. Cluster-based analysis of FMRI data. *NeuroImage* **33**(2), 599–608, November (2006).

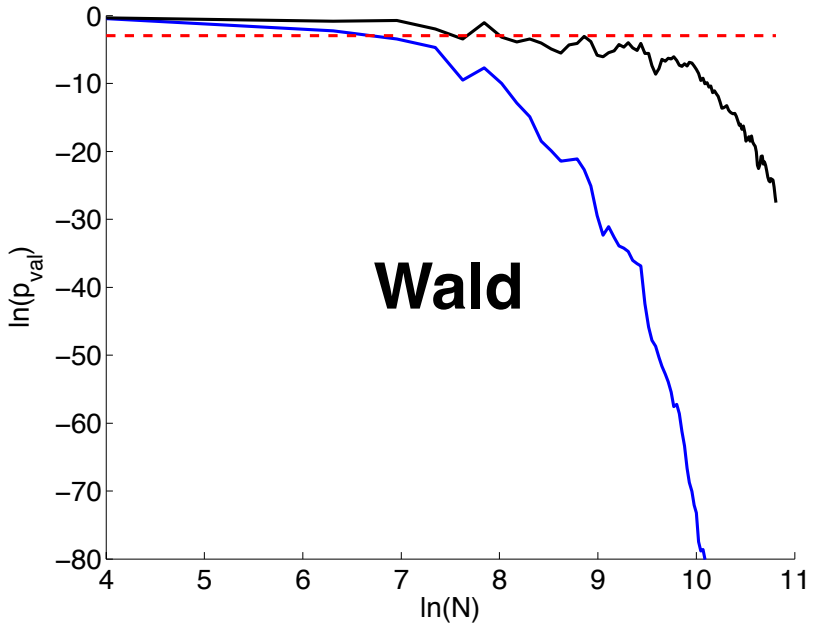
Benjamini, Y. and Heller, R. False Discovery Rates for Spatial Signals. *Journal of the American Statistical Association* **102**(480), 1272–1281, December (2007).

Chumbley, J., Worsley, K., Flandin, G., and Friston, K. Topological FDR for neuroimaging. *NeuroImage* **49**(4), 3057–64, February (2010).

Test Statistic Distributions In Simulation

(True Rates Known But Only Nominal Rates Provided to Algs)

Kolmogorov Smirnov Test for Assuming Normal Distribution



Estimated Model With Uncertainty is Fairly Close to Large Sample Limit. However Finite Sampling Effects Still Detectable with Large Enough MC Sample Size

(Test Stat Distribution Not Exactly Normal for Fixed Sample Size of 1's and 0's)

Mixed Sensor Types

"S" is Highly Simplified Here

$$S = h(Y) + \epsilon'$$

$$Y = \int_{\Omega} c(X)dX + \epsilon$$

$$Z_1' = (Z_1 + S)/\sqrt{2}$$

$$Z_2' = (Z_2 + S)/\sqrt{2}$$

$$Z_3' = Z_3$$

$$Z_4' = Z_4$$

$$Z_5' = Z_5$$

$$Z_6' = Z_6$$

$$Z_7' = Z_7$$

Sensor 1

Sensor 2

Sensor 3

Sensor 4

Sensor 5

Sensor 6

Sensor 7