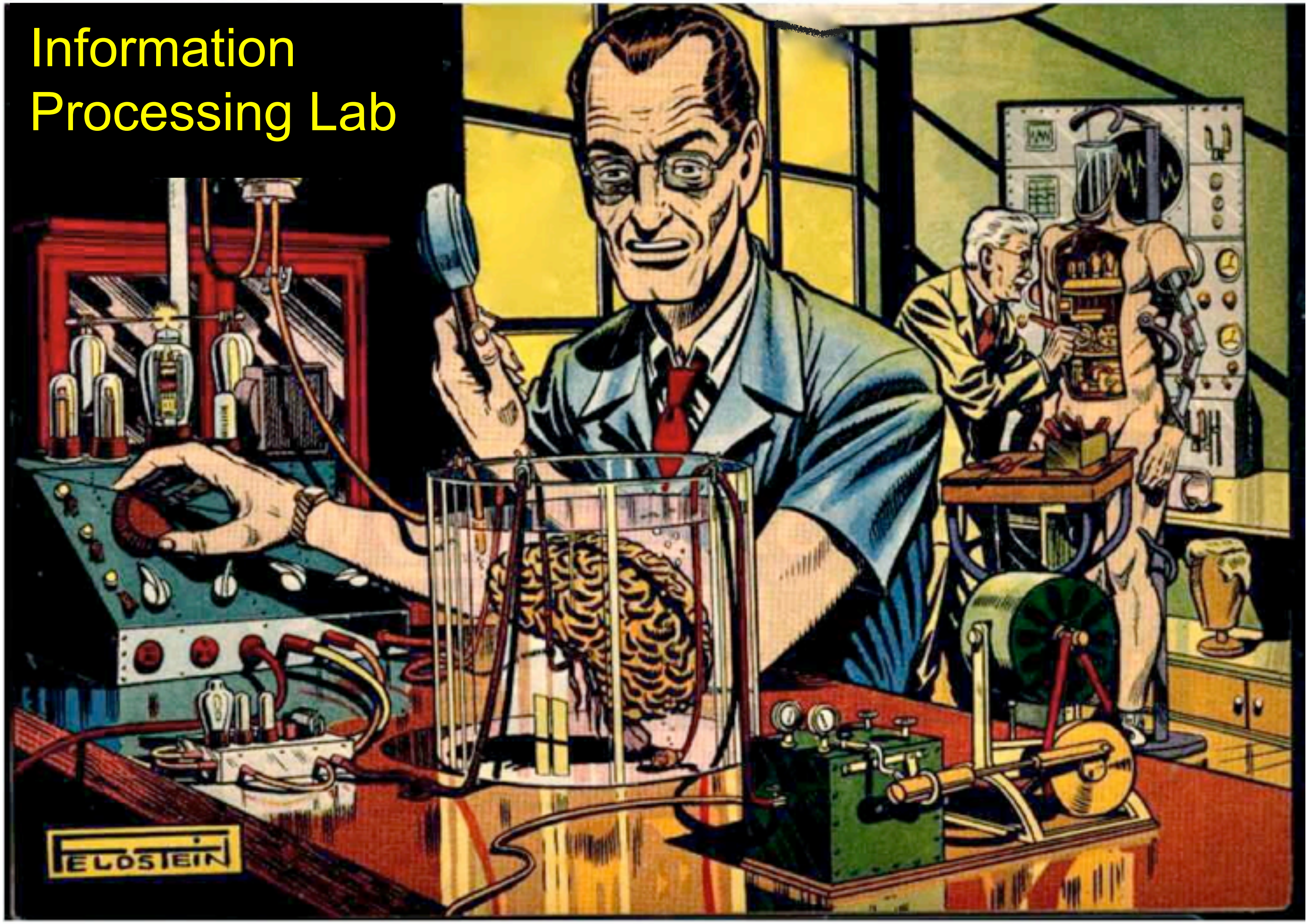
Sequential Analysis in High-Dimensional Multiple Testing and Sparse Recovery

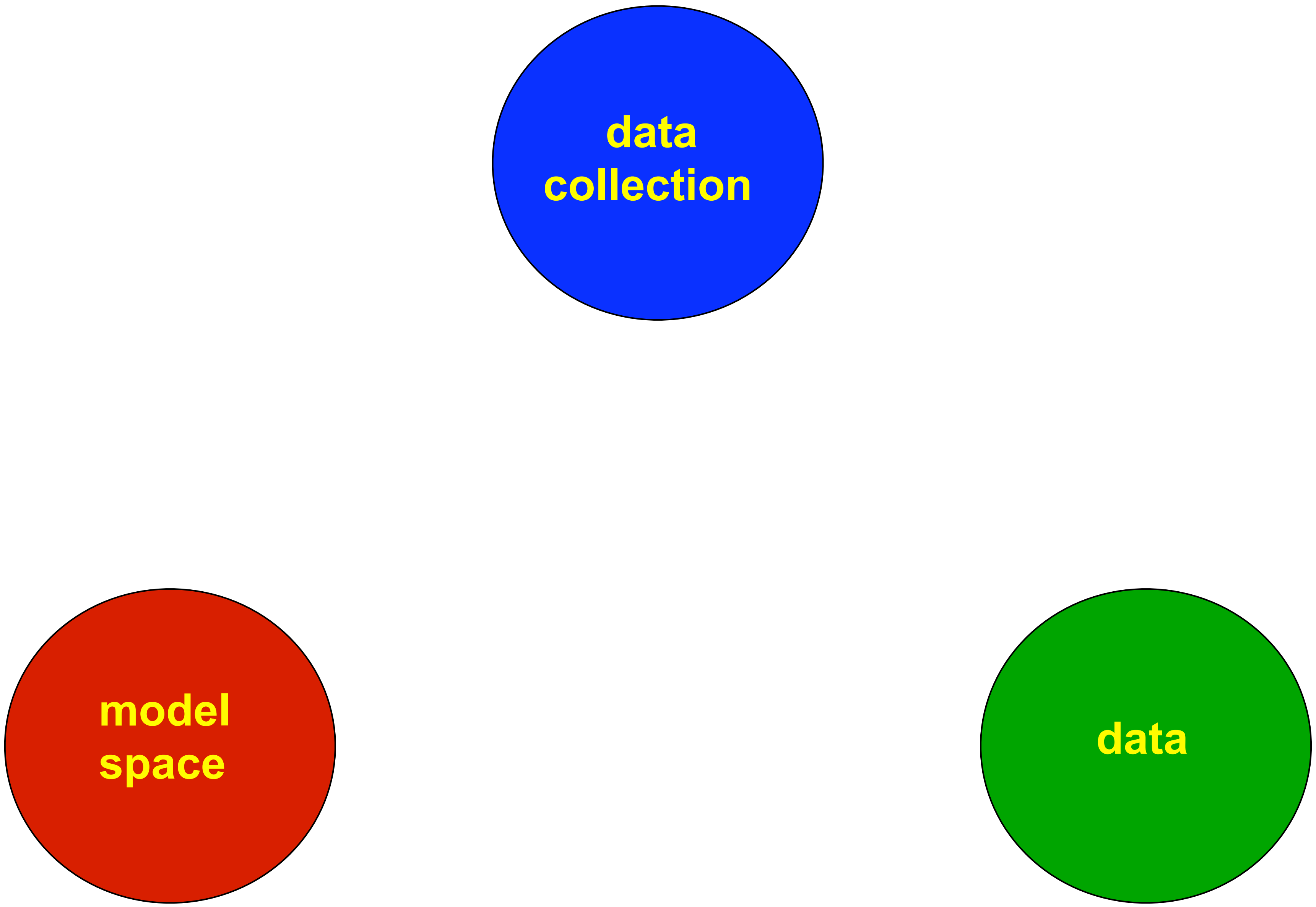


Duke SAHD Workshop, July 27, 2011

Rob Nowak www.ece.wisc.edu/~nowak
joint work with R. Castro, J. Haupt, & M. Malloy

Sequential Sensing and Experimental Design

$\mathcal{Y}$: possible measurements/experiments

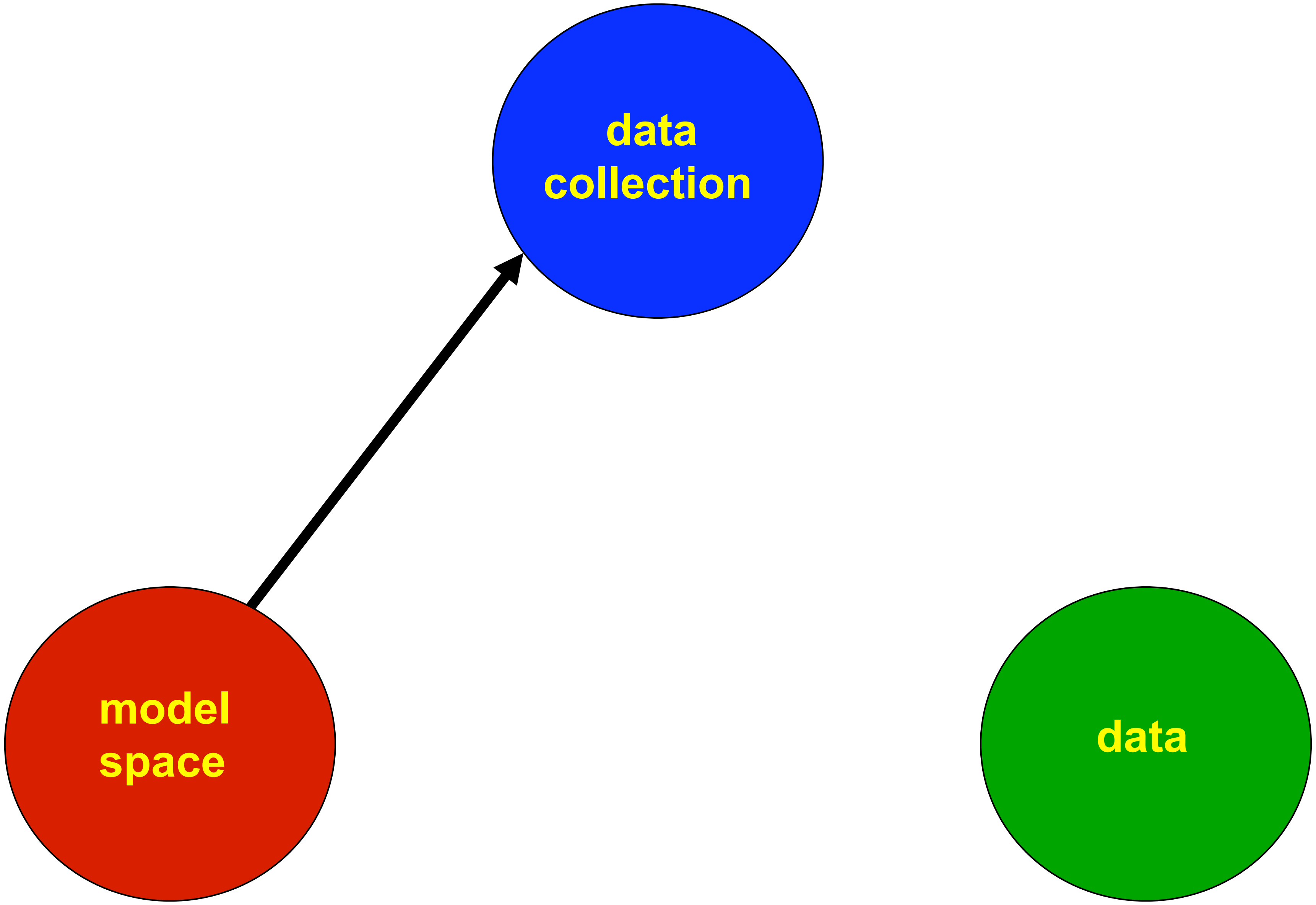


$\mathcal{X}$: models/hypotheses
under consideration

$y_1(x), y_2(x), \ldots$: information/data

Sequential Sensing and Experimental Design

$\mathcal{Y}$: possible measurements/experiments

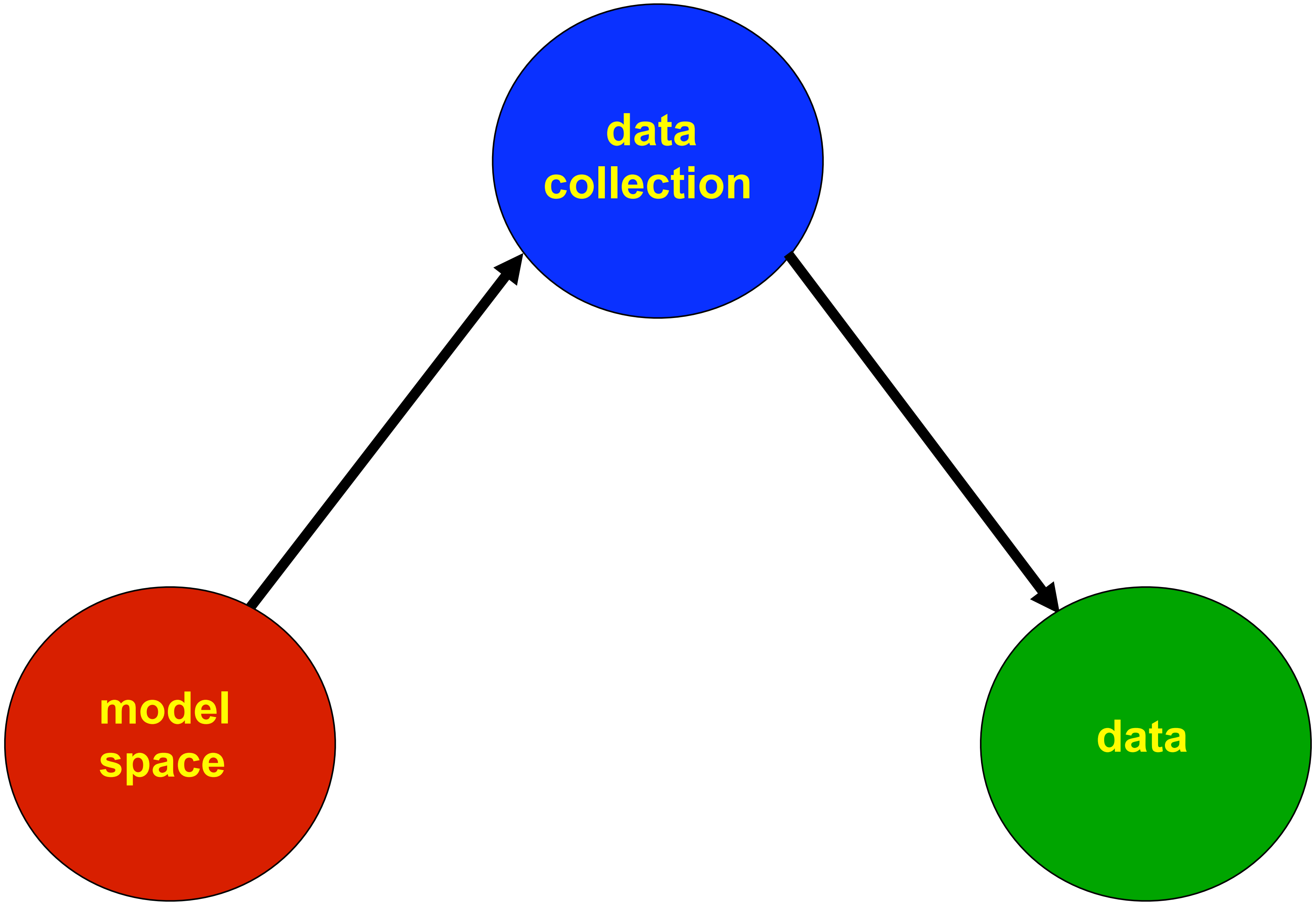


$\mathcal{X}$: models/hypotheses under consideration

$y_1(x), y_2(x), \ldots$: information/data

Sequential Sensing and Experimental Design

$\mathcal{Y}$: possible measurements/experiments

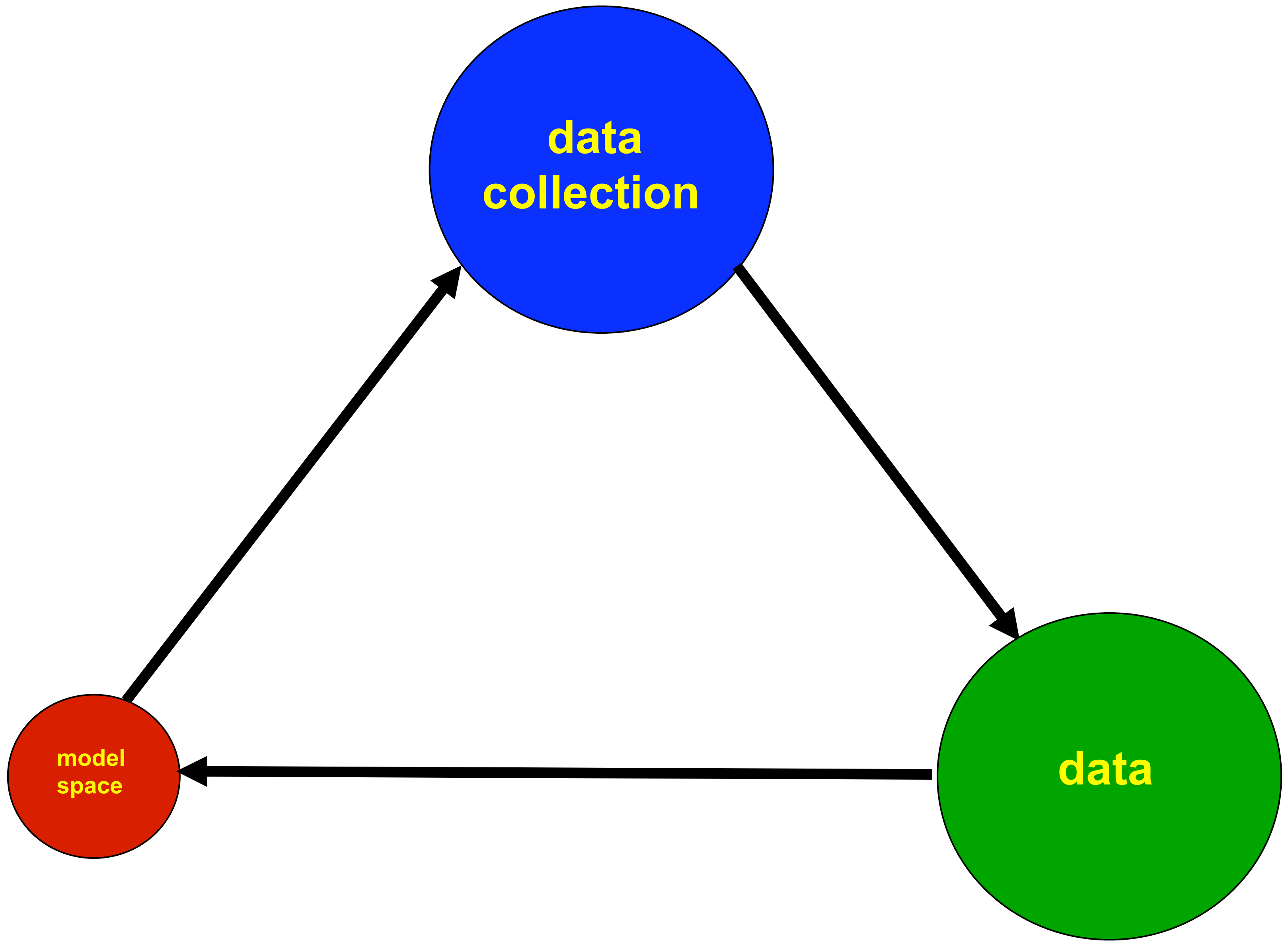


$\mathcal{X}$: models/hypotheses
under consideration

$y_1(x), y_2(x), \ldots$: information/data

Sequential Sensing and Experimental Design

$\mathcal{Y}$: possible measurements/experiments



$\mathcal{X}$: models/hypotheses under consideration

$y_1(x), y_2(x), \ldots$: information/data

Motivation: Inferring Biological Networks

Paul Alhquist
(Molecular Virology)

Motivation: Inferring Biological Networks


Paul Alhquist
(Molecular Virology)

virus

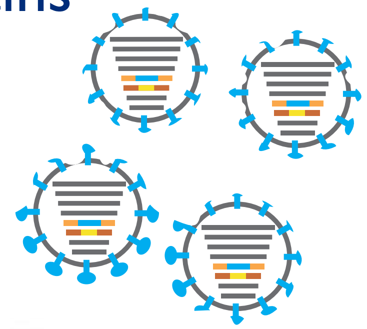
fruit fly

Motivation: Inferring Biological Networks

virus

fruit fly

13,071 single-gene
knock-down cell strains



Paul Alhquist
(Molecular Virology)

Motivation: Inferring Biological Networks

Paul Alhquist
(Molecular Virology)

virus

fruit fly

13,071 single-gene
knock-down cell strains

infect each strain
with fluorescing virus

microwell
array

Motivation: Inferring Biological Networks

Paul Alhquist
(Molecular Virology)

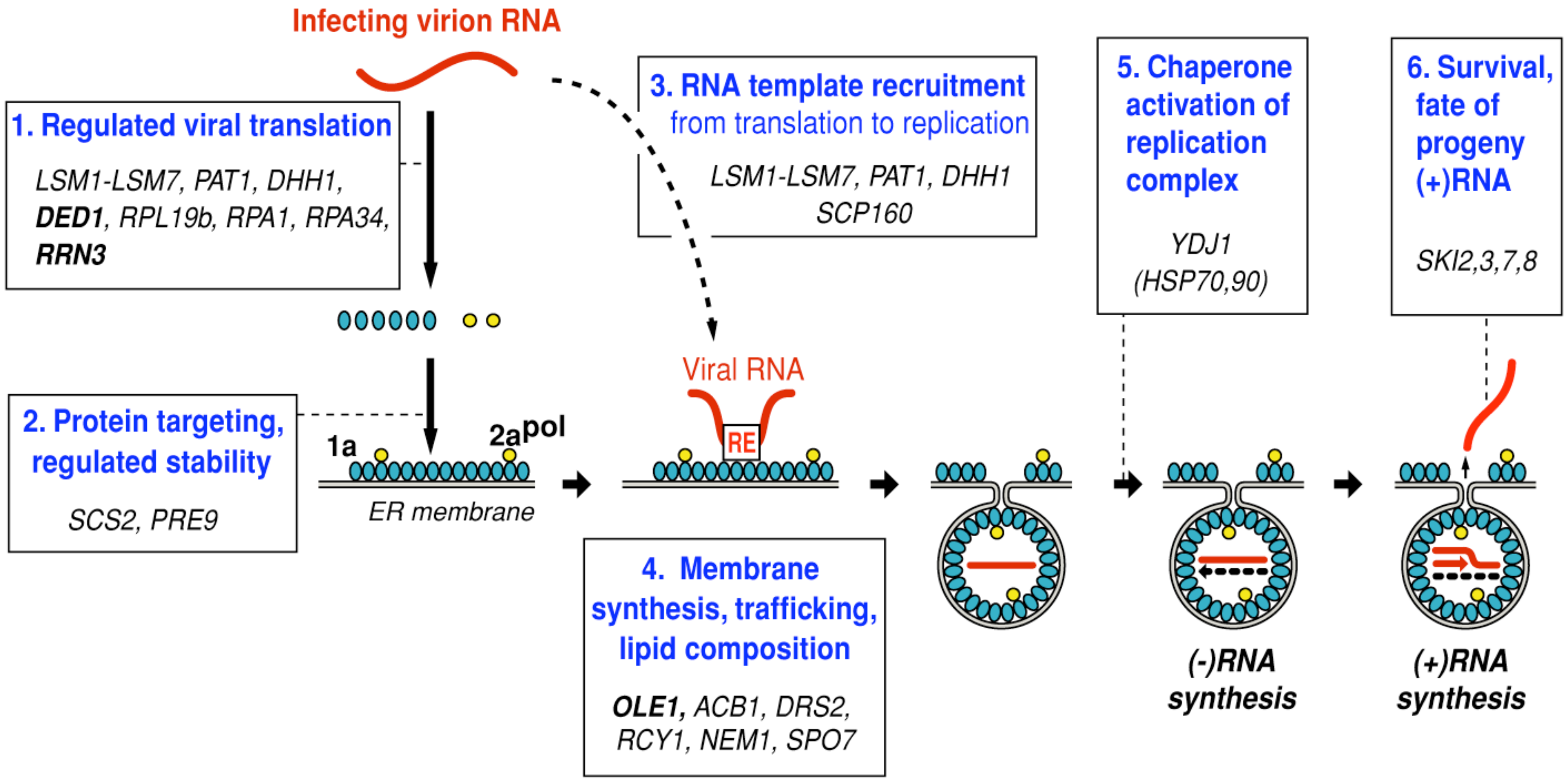
virus

13,071 single-gene
knock-down cell strains

fruit fly

infect each strain
with fluorescing virus

microwell
array



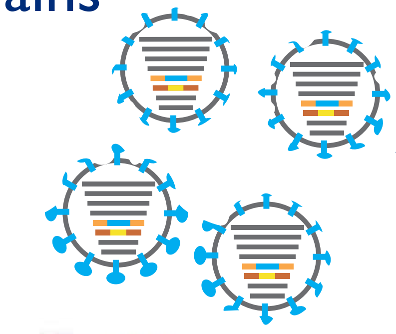
Motivation: Inferring Biological Networks

Paul Alhquist
(Molecular Virology)

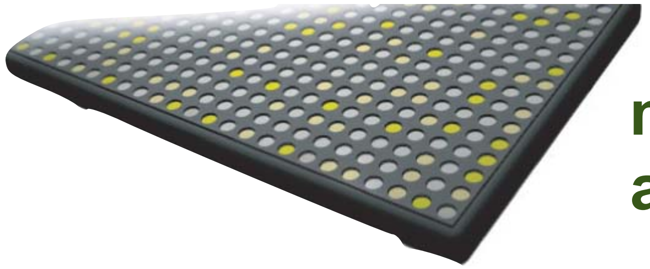
virus

fruit fly

13,071 single-gene
knock-down cell strains



infect each strain
with fluorescing virus



microwell
array

First question: Who are the players in the network?

"Drosophila RNAi screen identifies host genes important for influenza virus replication," Nature 2008. How do they confidently determine the ~100 out of 13K genes hijacked for virus replication from extremely noisy data?

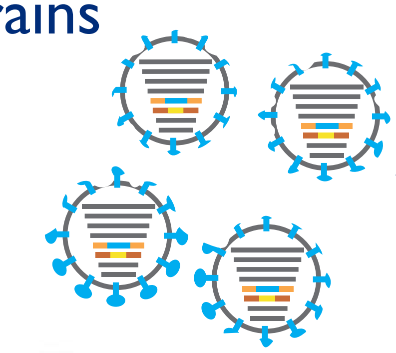
Motivation: Inferring Biological Networks

Paul Alhquist
(Molecular Virology)

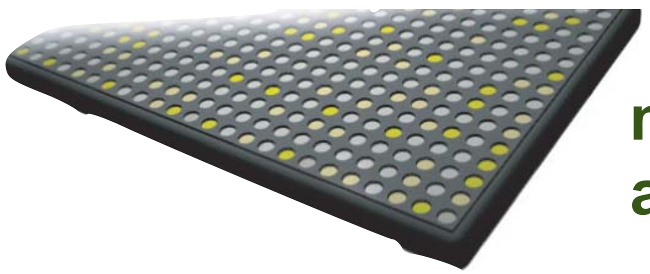
virus

fruit fly

13,071 single-gene
knock-down cell strains



infect each strain
with fluorescing virus



microwell
array

First question: Who are the players in the network?

“Drosophila RNAi screen identifies host genes important for influenza virus replication,” Nature 2008. How do they confidently determine the ~100 out of 13K genes hijacked for virus replication from extremely noisy data?

Sequential Experimental Design:

Stage 1: assay all 13K strains, twice; keep all with significant fluorescence in one or both assays for 2nd stage (13K → 1K)

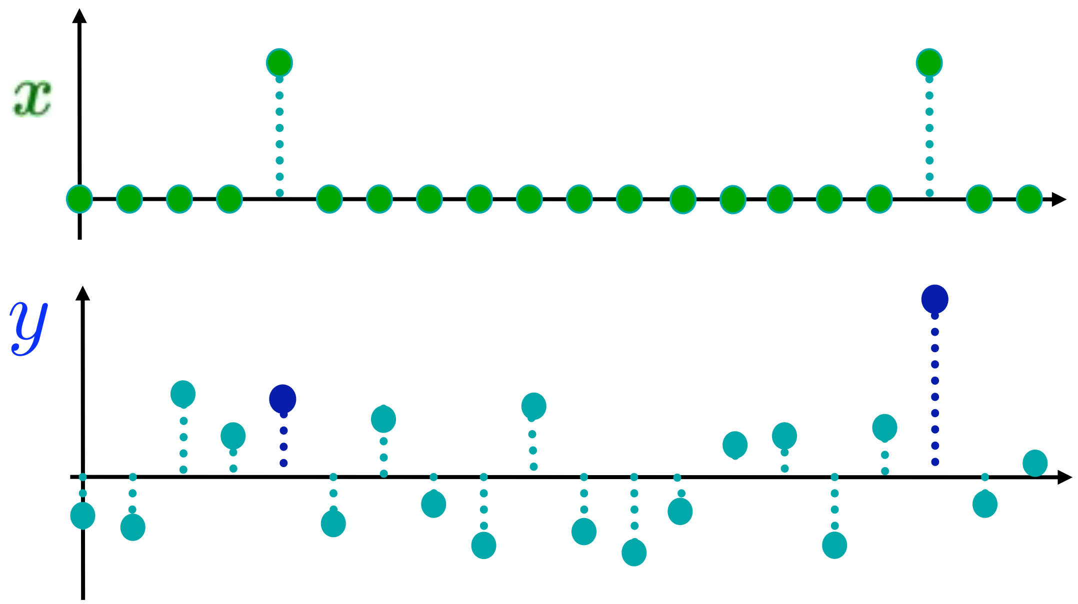
Stage 2: assay remaining 1K strains, 6-12 times; retain only those with statistically significant fluorescence (1K → 100)

vastly more efficient that replicating all 13K experiments many times

Idealized Example

non-sequential design

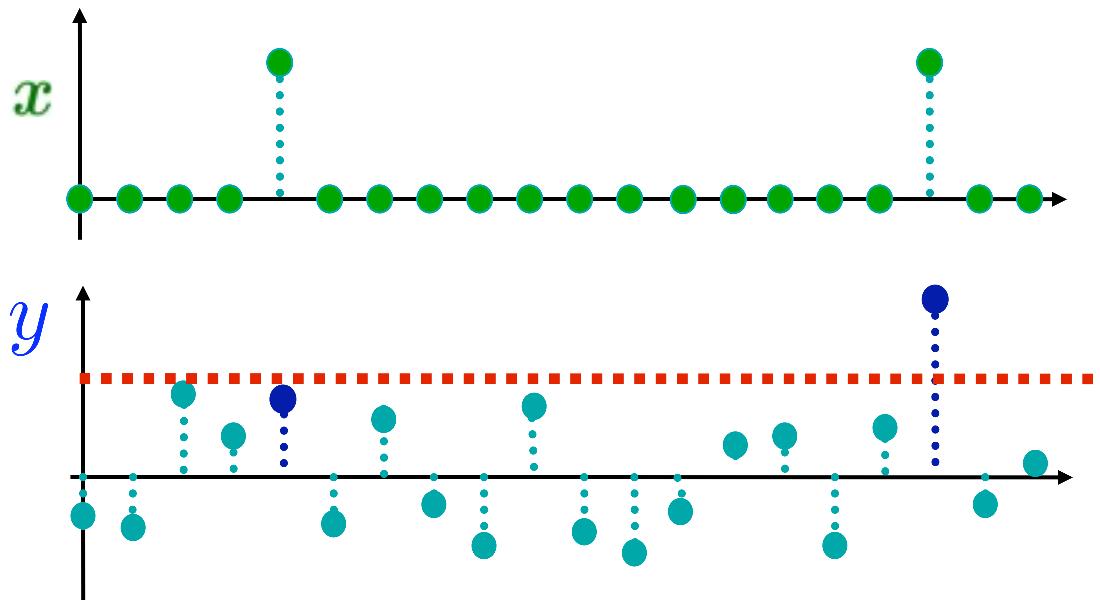
(n cell strains, 3 samples each)



Idealized Example

non-sequential design

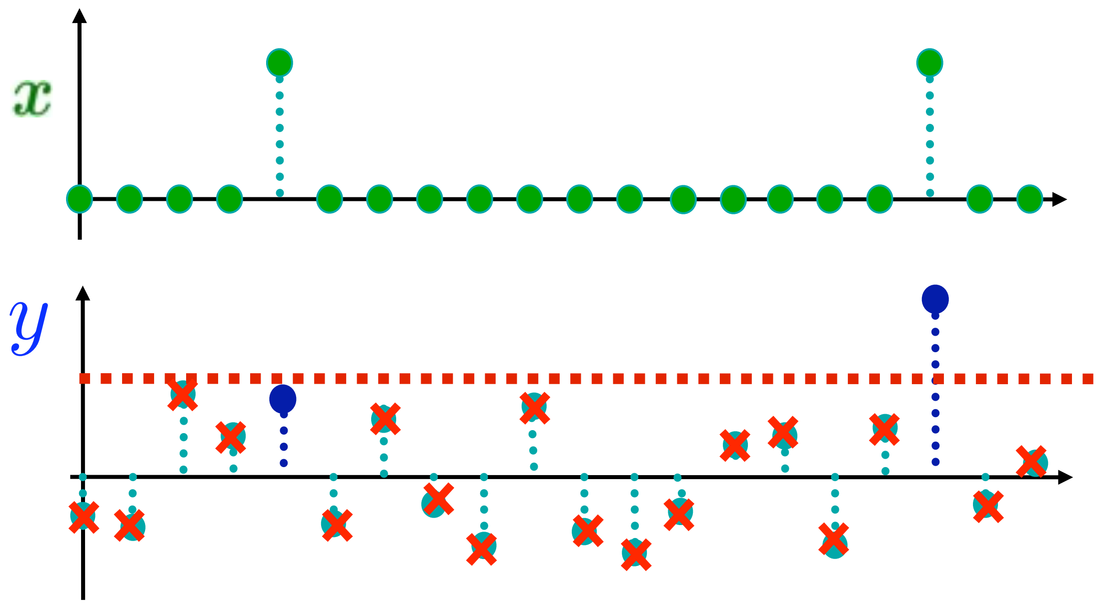
(n cell strains, 3 samples each)



measure each cell strain with equal precision/SNR, then threshold to control false-positive error

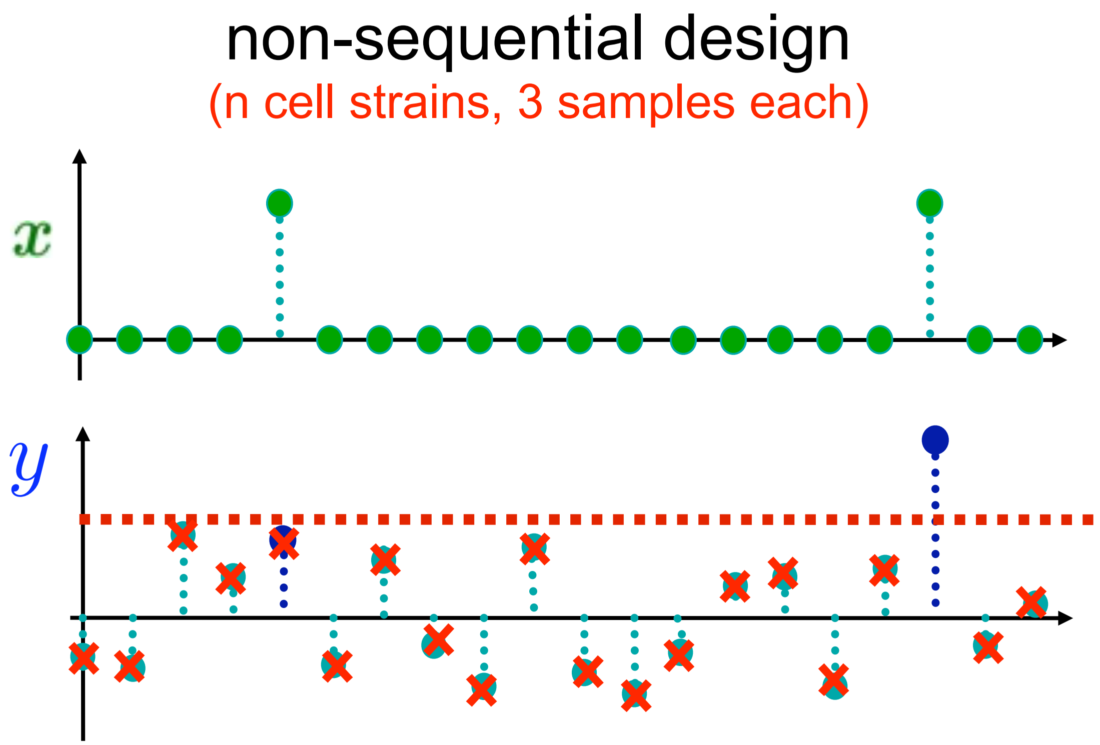
Idealized Example

non-sequential design
(n cell strains, 3 samples each)



measure each cell strain with equal precision/SNR, then threshold to control false-positive error

Idealized Example



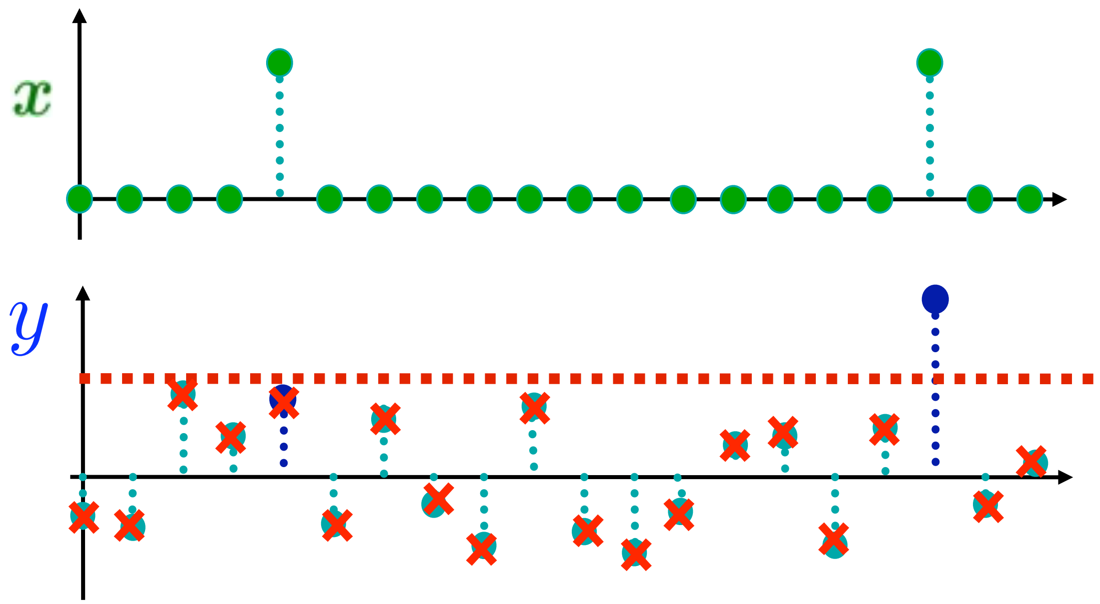
measure each cell strain with equal precision/SNR, then threshold to control false-positive error

may be impossible to reliably separate signals from noise

Idealized Example

non-sequential design

(n cell strains, 3 samples each)

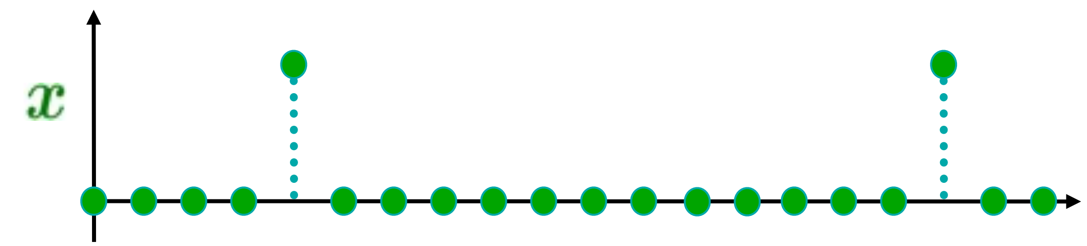


measure each cell strain with equal precision/SNR, then threshold to control false-positive error

may be impossible to reliably separate signals from noise

two-stage design

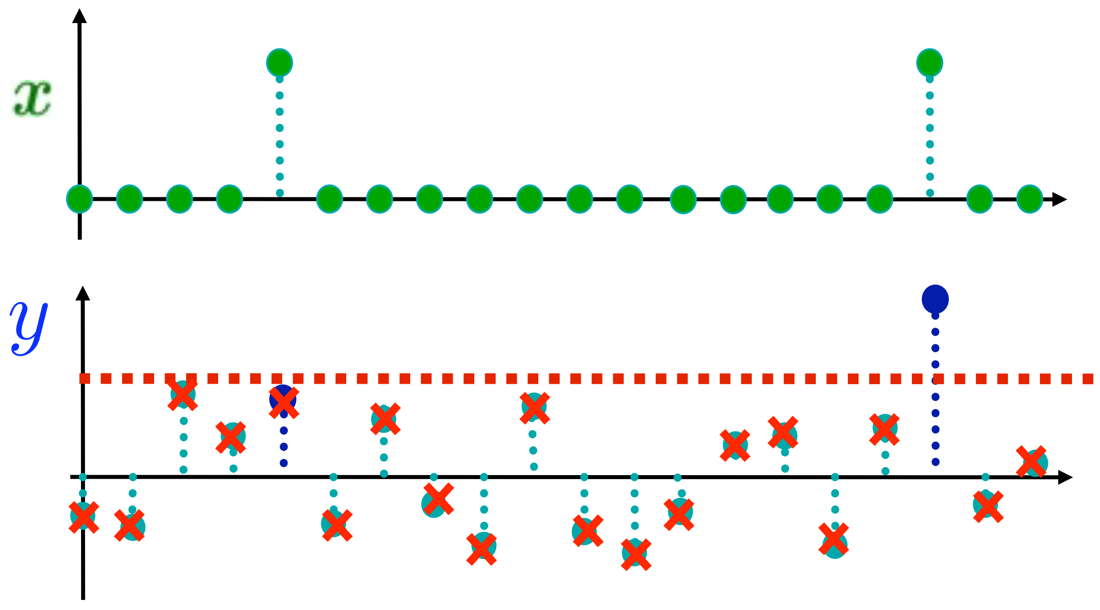
(adaptively allocate 3n samples)



Idealized Example

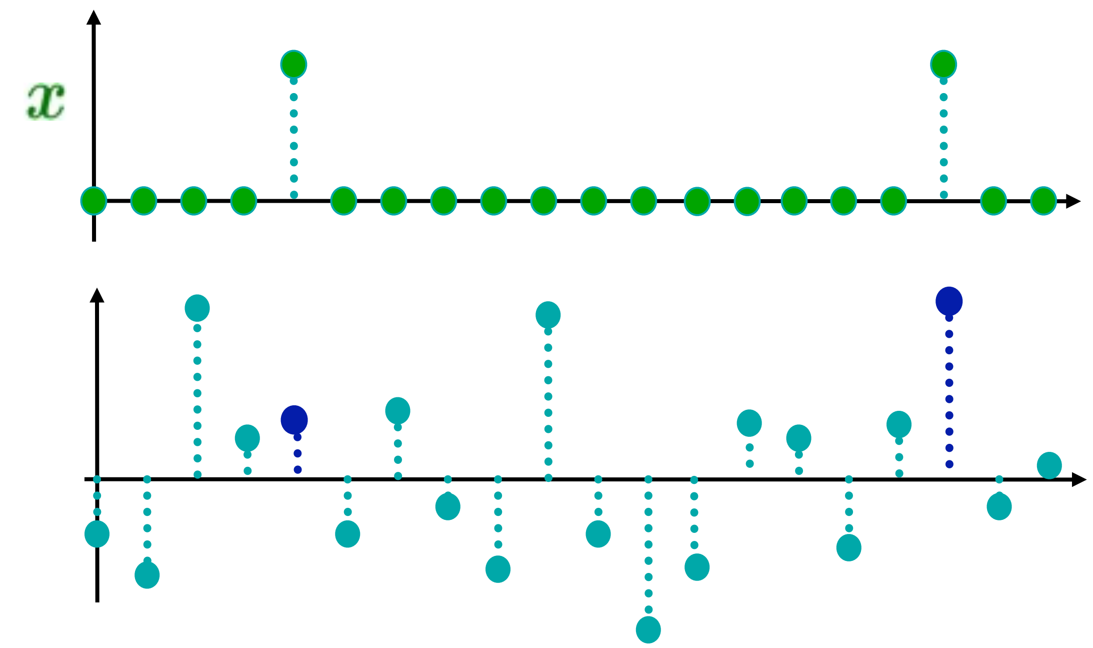
non-sequential design

(n cell strains, 3 samples each)



two-stage design

(adaptively allocate 3n samples)



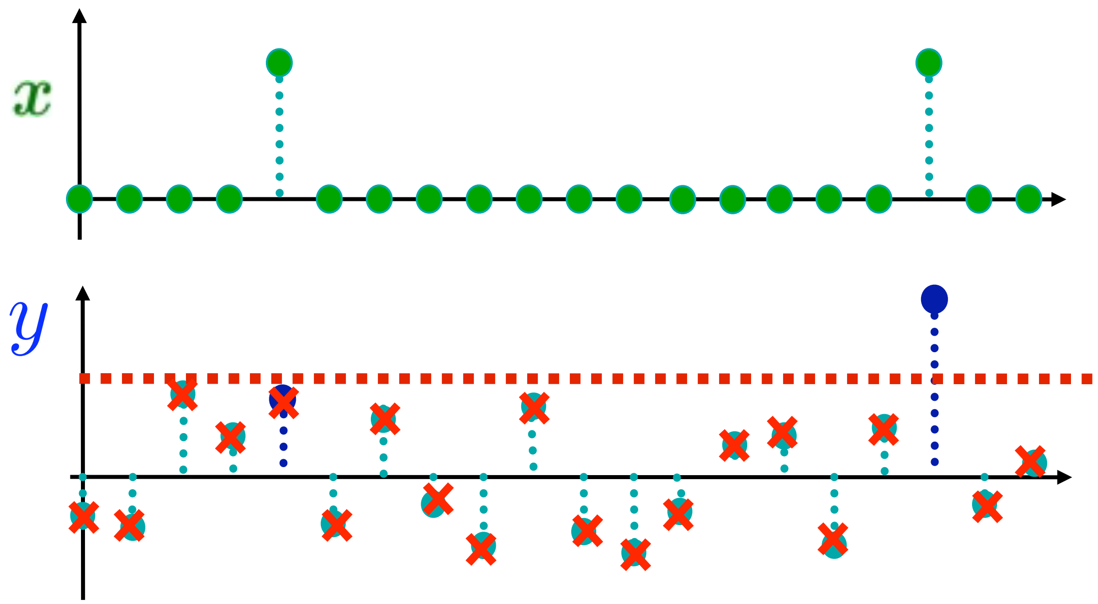
measure each cell strain with equal precision/SNR, then threshold to control false-positive error

may be impossible to reliably separate signals from noise

Idealized Example

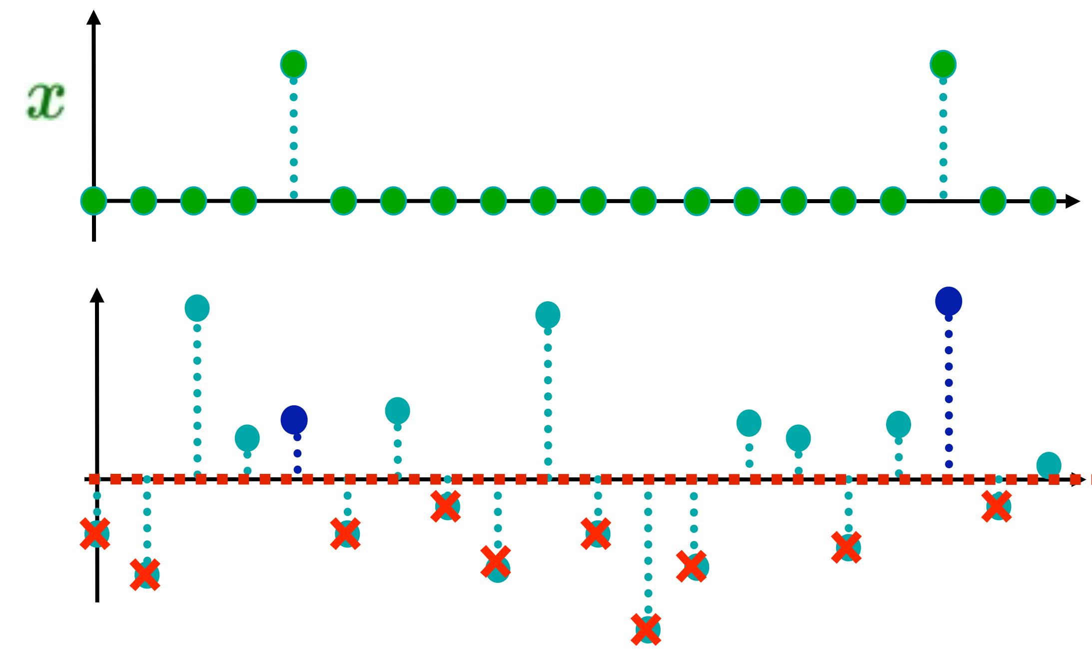
non-sequential design

(n cell strains, 3 samples each)



two-stage design

(adaptively allocate 3n samples)

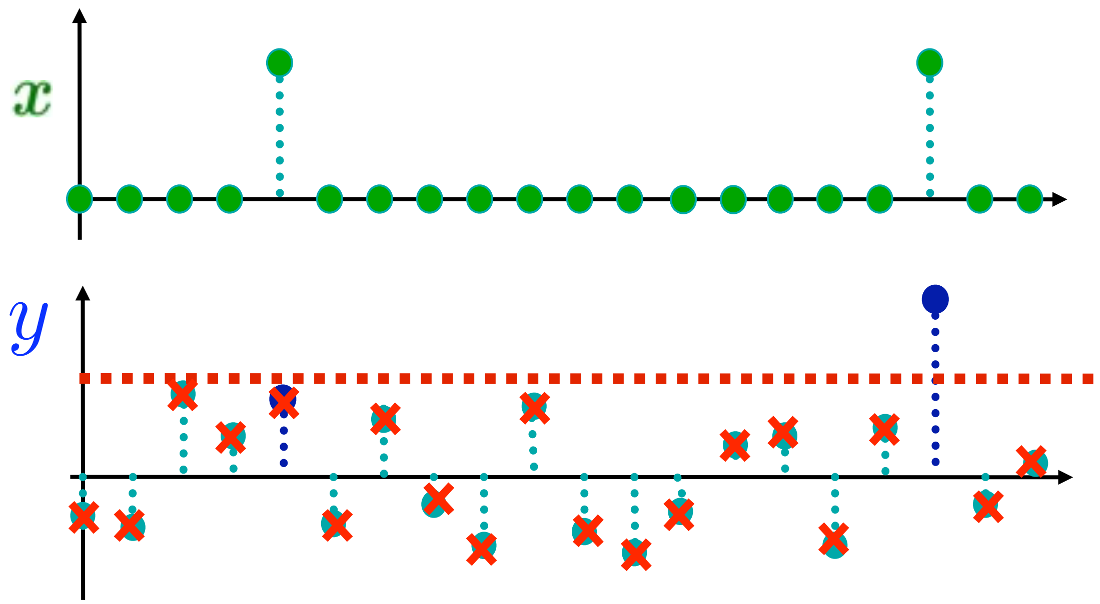


measure each cell strain with equal precision/SNR, then threshold to control false-positive error

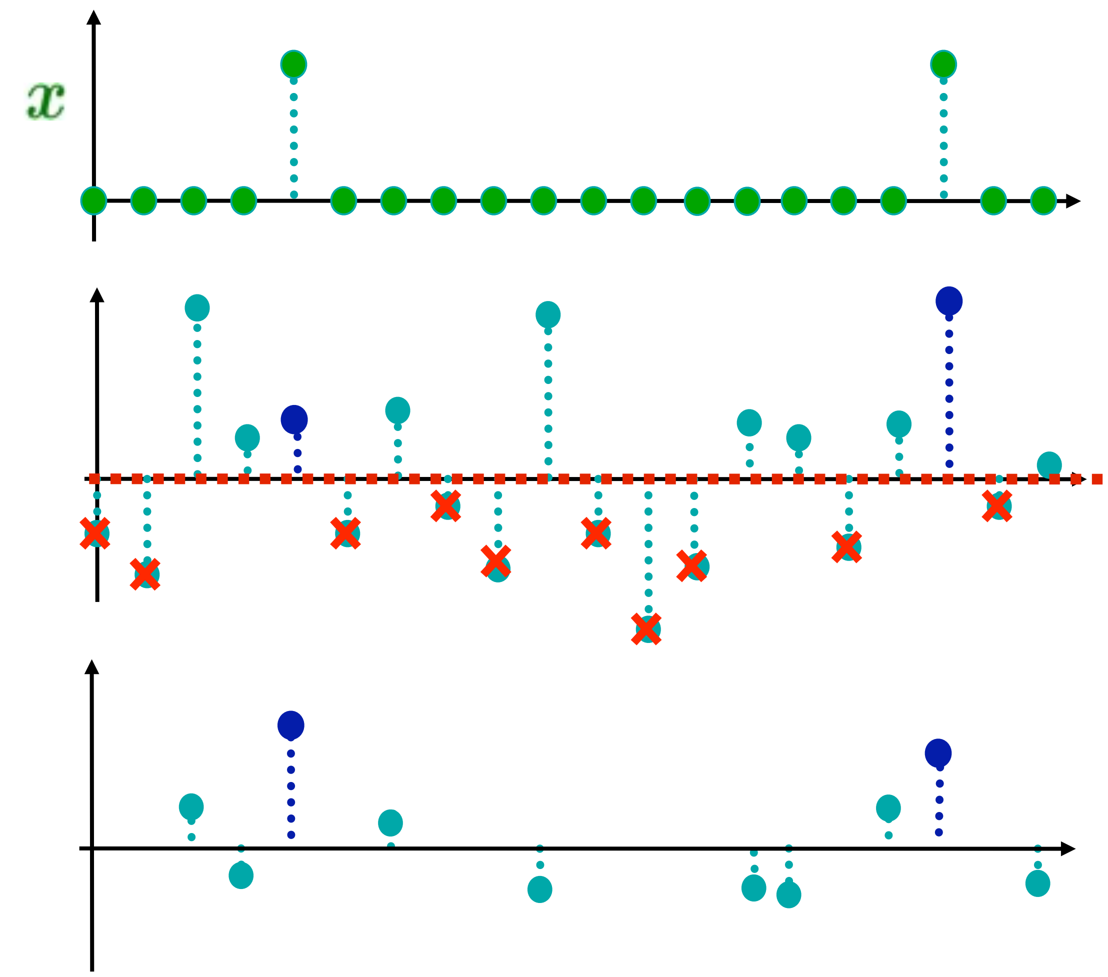
may be impossible to reliably separate signals from noise

Idealized Example

non-sequential design
(n cell strains, 3 samples each)



two-stage design
(adaptively allocate 3n samples)



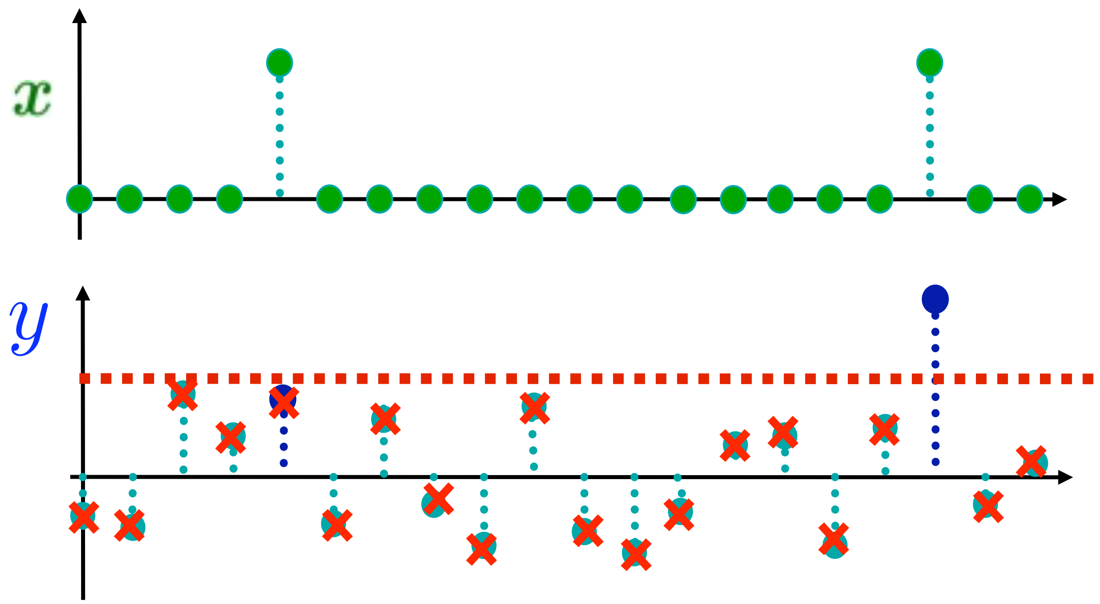
measure each cell strain with equal precision/SNR, then threshold to control false-positive error

may be impossible to reliably separate signals from noise

Idealized Example

non-sequential design

(n cell strains, 3 samples each)

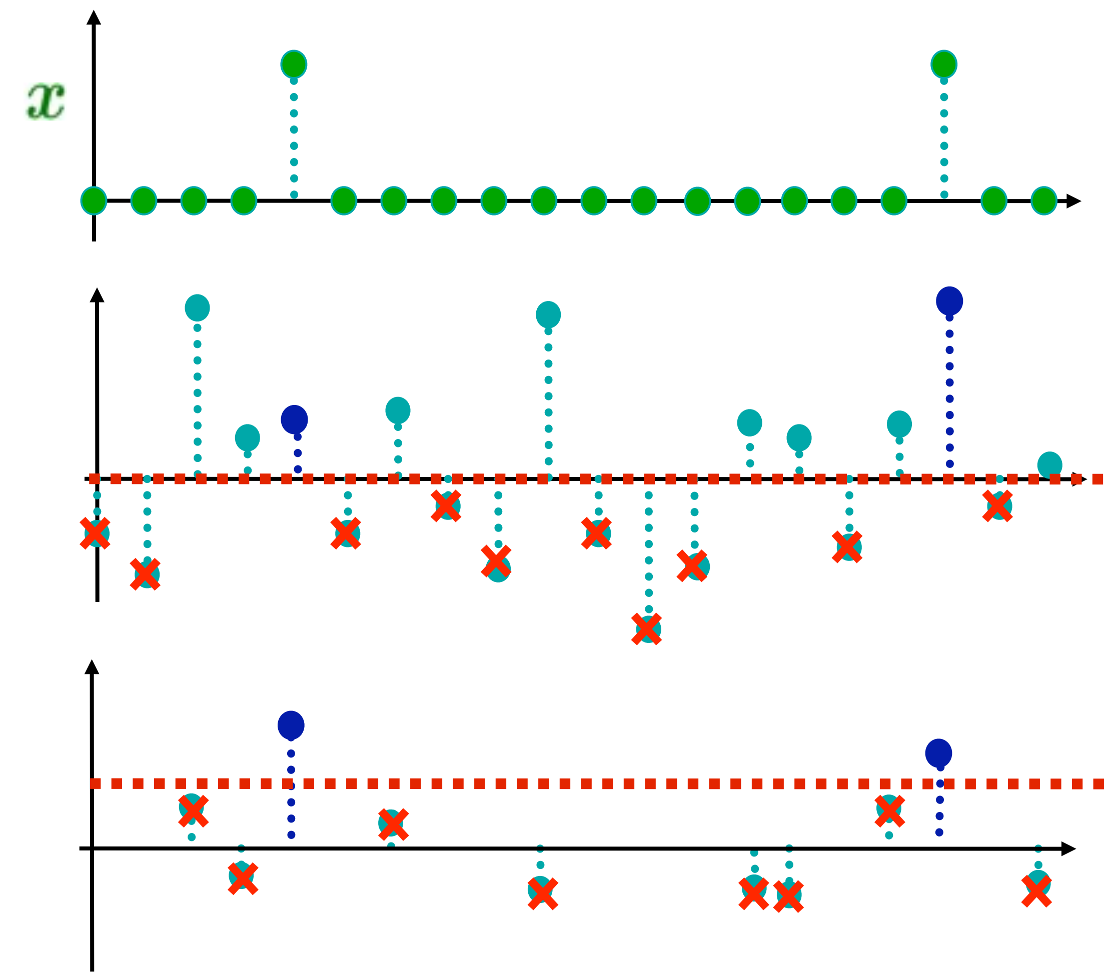


measure each cell strain with equal precision/SNR, then threshold to control false-positive error

may be impossible to reliably separate signals from noise

two-stage design

(adaptively allocate 3n samples)

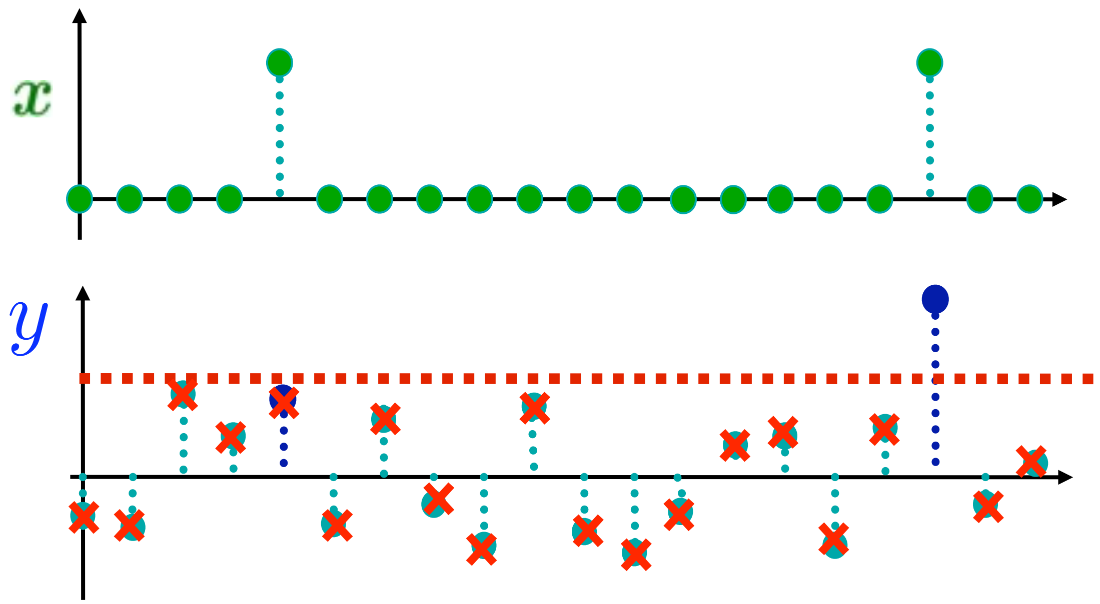


first stage has large false-positive rate, but low false-negative. larger SNR in second stage makes it easier to separate signals from noise.

Idealized Example

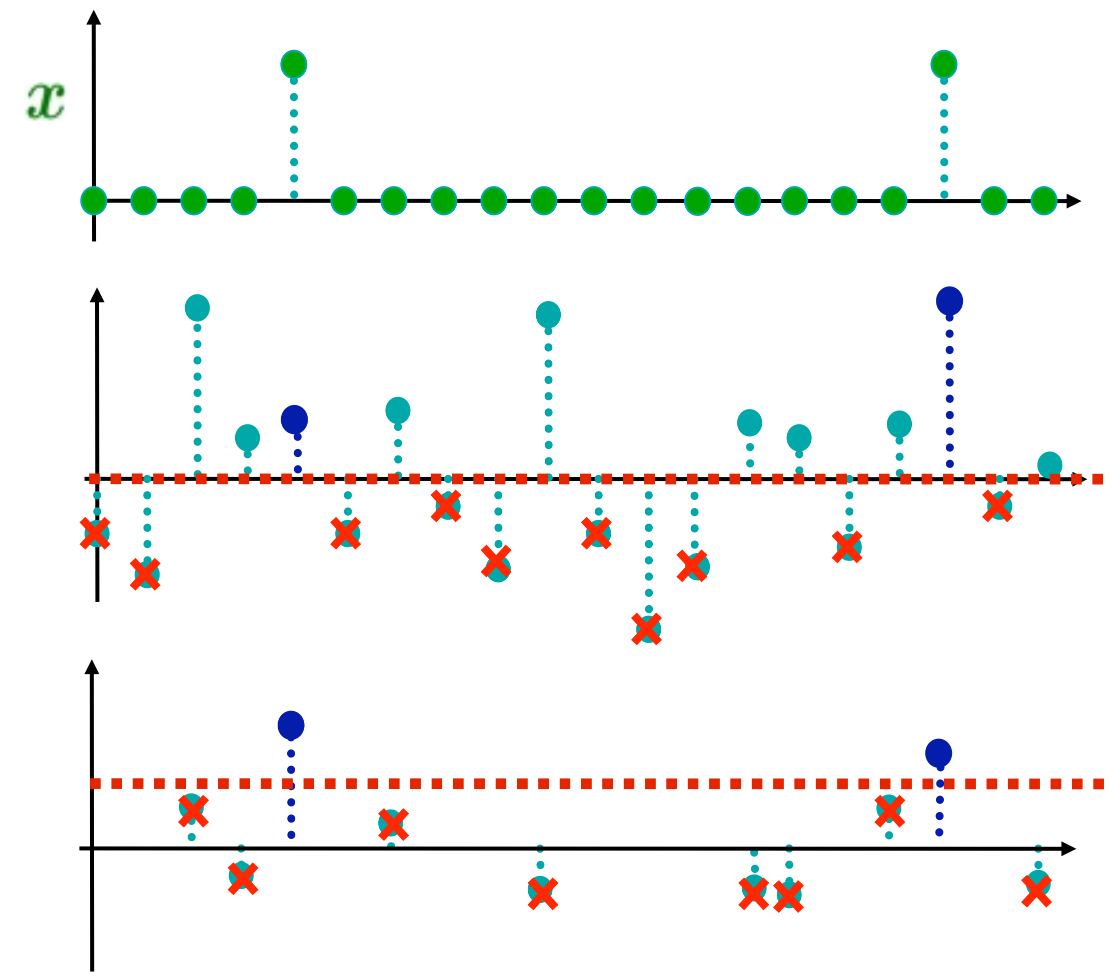
non-sequential design

(n cell strains, 3 samples each)



two-stage design

(adaptively allocate 3n samples)



measure each cell strain with equal precision/SNR, then threshold to control false-positive error

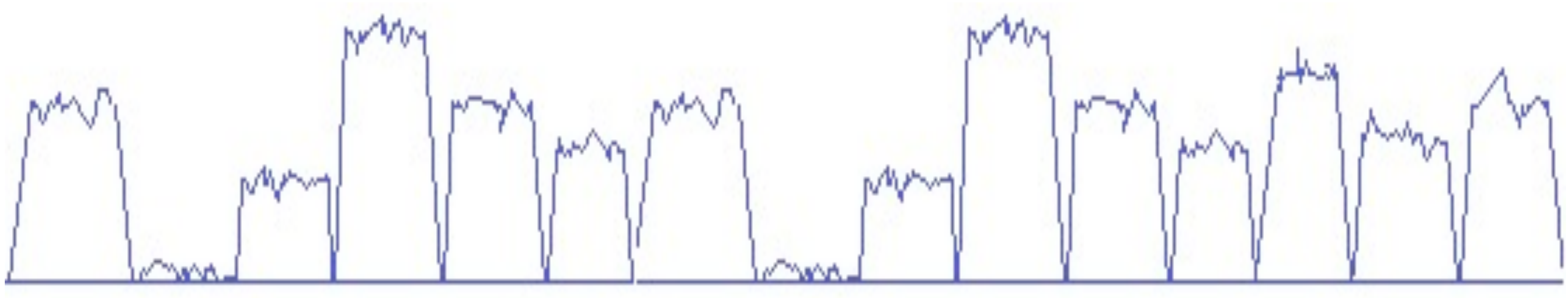
may be impossible to reliably separate signals from noise

first stage has large false-positive rate, but low false-negative. larger SNR in second stage makes it easier to separate signals from noise.

Under a fixed sensing/experimental budget, does this two-stage design (or some other sequential design) provide better error control than non-sequential design?

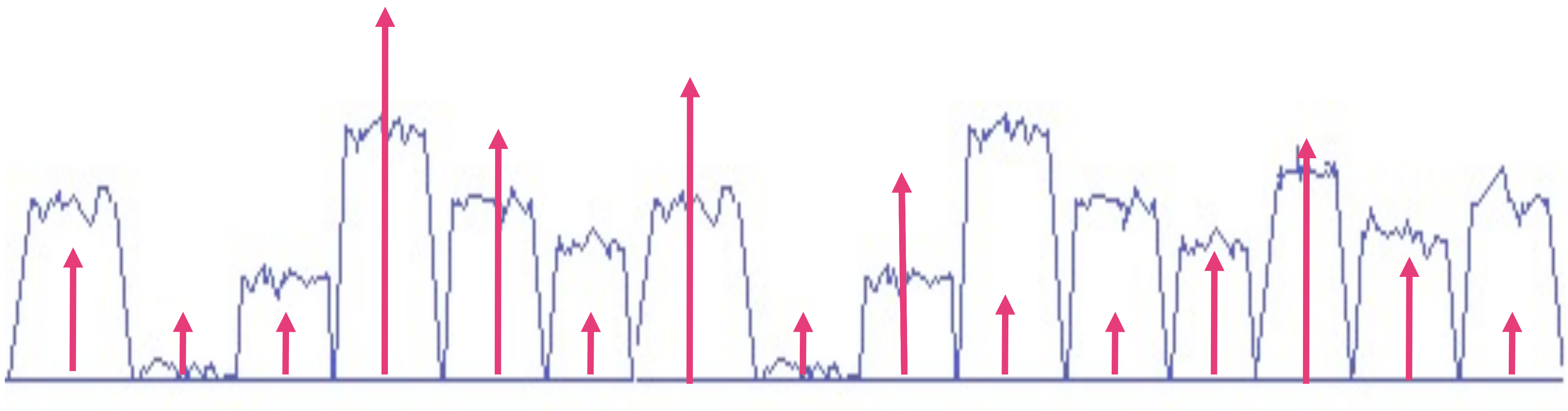
Cognitive Radio Spectrum Sensing

"primary" users have preference over "secondary" users



Cognitive Radio Spectrum Sensing

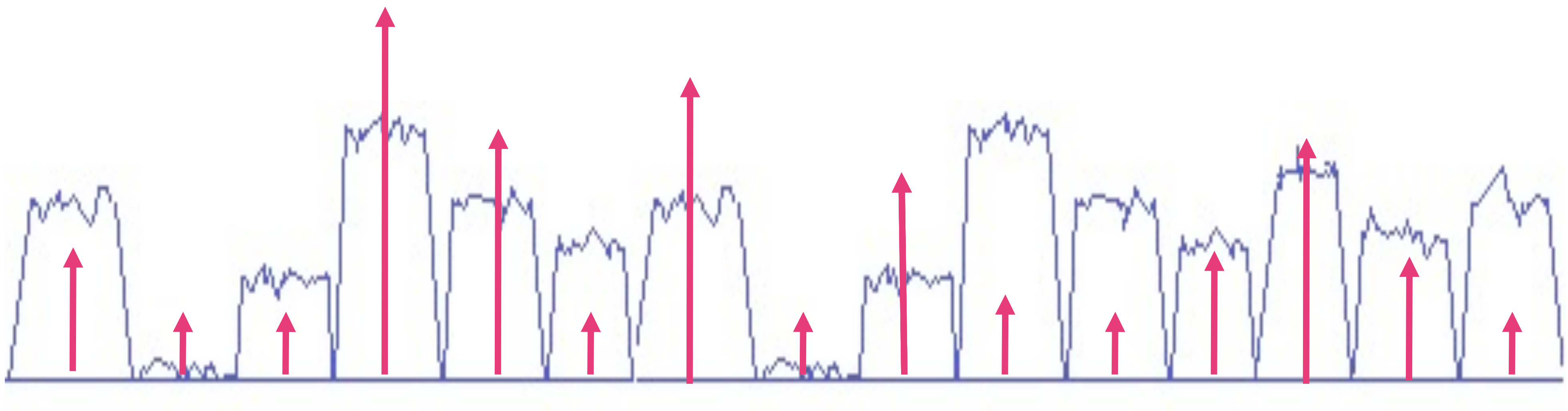
"primary" users have preference over "secondary" users



most channels occupied by primary users, but they come and go in unpredictable manner. Secondary users "sense" spectrum to find an unoccupied channel

Cognitive Radio Spectrum Sensing

"primary" users have preference over "secondary" users



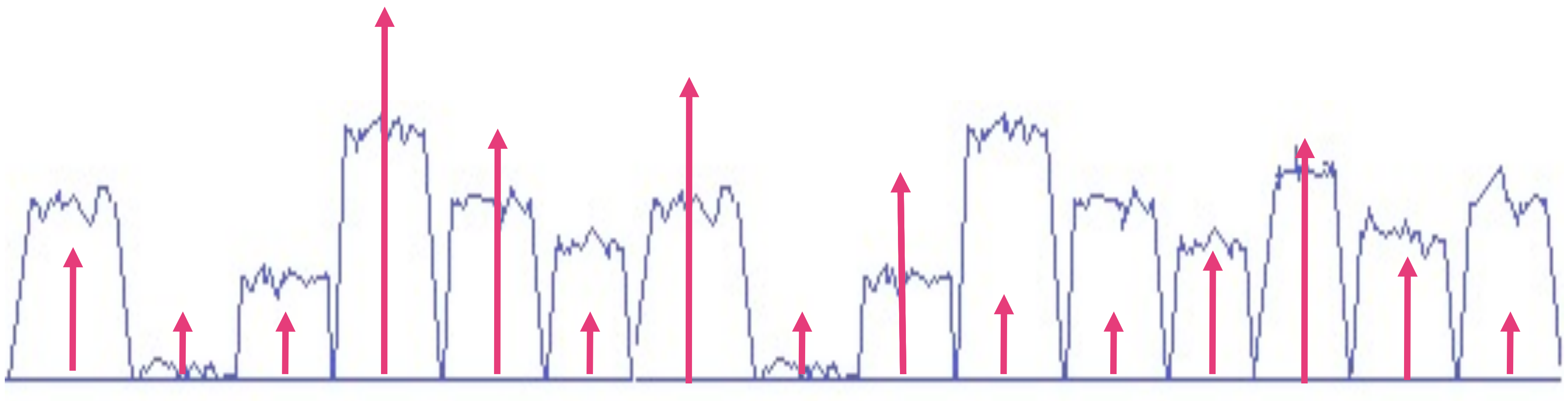
most channels occupied by primary users, but they come and go in unpredictable manner. Secondary users "sense" spectrum to find an unoccupied channel

Goal: Find open channel(s) as quickly as possible. Two approaches:

1) listen to each channel for a fixed amount of time and make decision
2) listen to each channel for a data-adaptive amount of time to make decisions as quickly as possible

Cognitive Radio Spectrum Sensing

"primary" users have preference over "secondary" users



most channels occupied by primary users, but they come and go in unpredictable manner. Secondary users "sense" spectrum to find an unoccupied channel

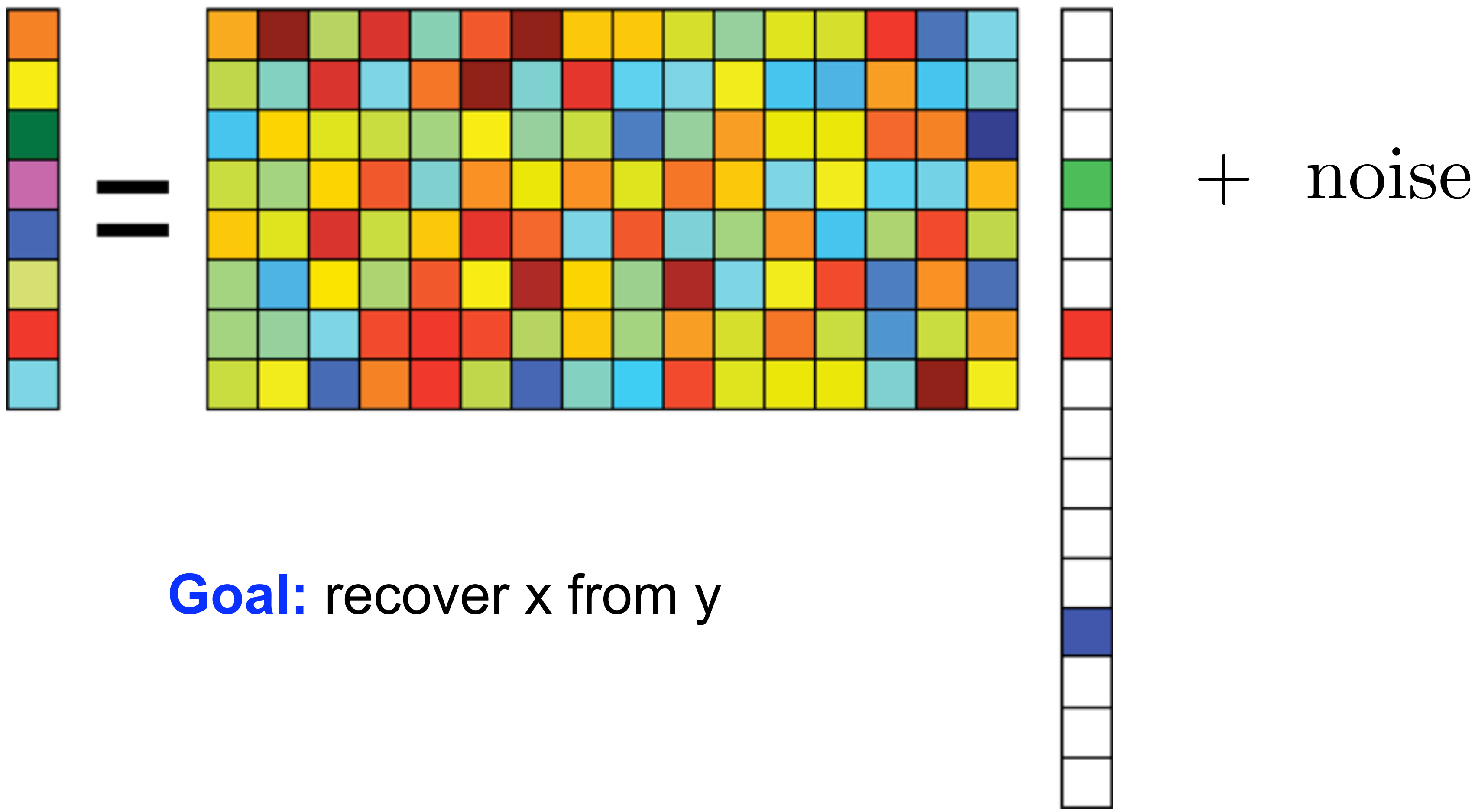
Goal: Find open channel(s) as quickly as possible. Two approaches:

1) listen to each channel for a fixed amount of time and make decision
2) listen to each channel for a data-adaptive amount of time to make decisions as quickly as possible

adaptive spectrum sensing can be significantly more time-efficient than fixed sensing

Sparse Recovery (image reconstruction, compressed sensing, inverse problems)

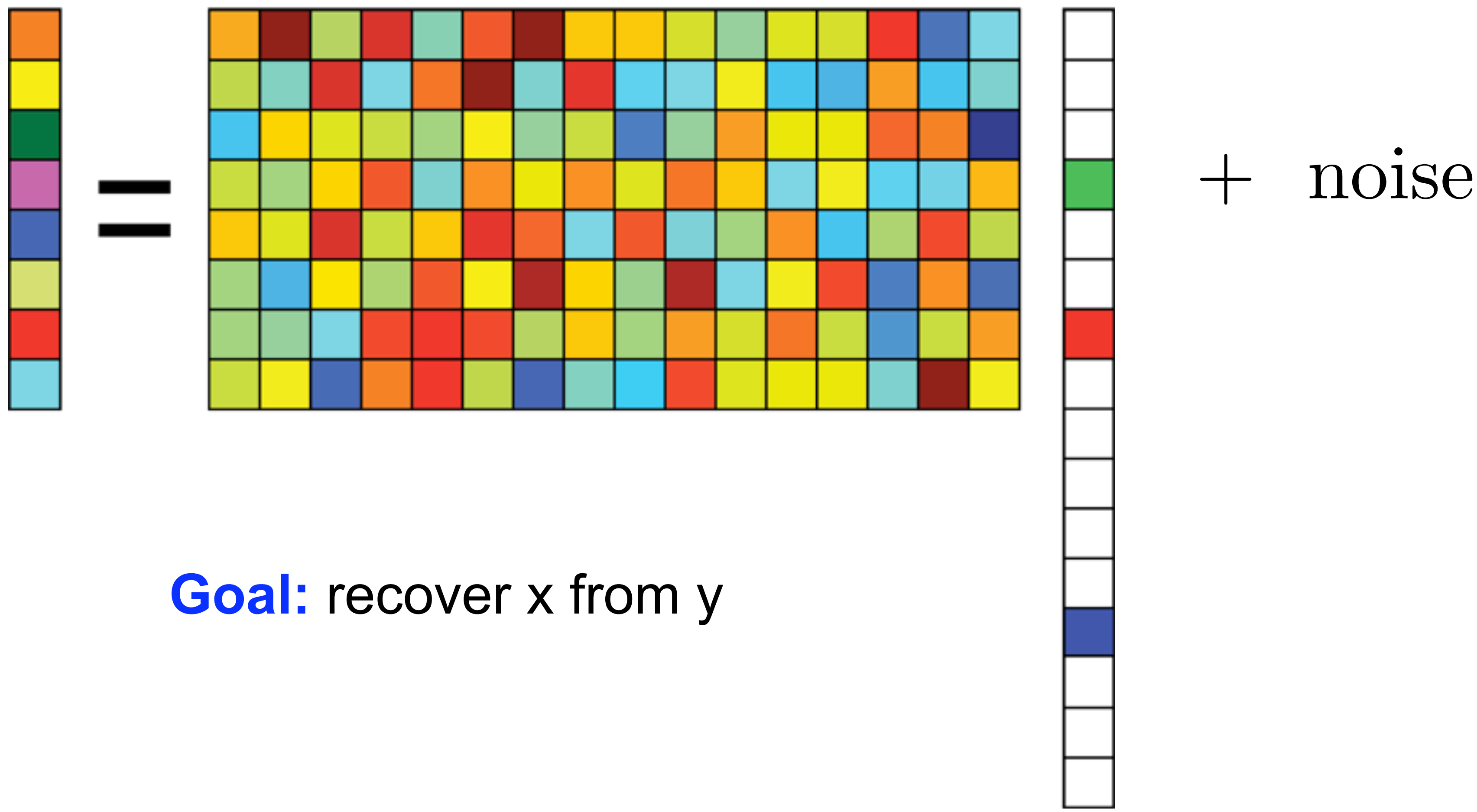
$$y \;=\; Ax \;+\; w \text{ , with } A \in \mathbb{R}^{m\times n},\ x \in \mathbb{R}^n \text{ (but sparse)},\ w \sim \mathcal{N}(0, I)$$



Goal: recover x from y

Sparse Recovery (image reconstruction, compressed sensing, inverse problems)

$$y \;=\; A\,x \;+\; w\;,\ \text{with } A \in \mathbb{R}^{m\times n},\ x \in \mathbb{R}^{n} \text{ (but sparse)},\ w \sim \mathcal{N}(0, I)$$



Goal: recover x from y

Is sequentially designing (rows of) A advantageous ?

Model Selection

Challenge: huge number of possible models, each with a different pattern of sparsity

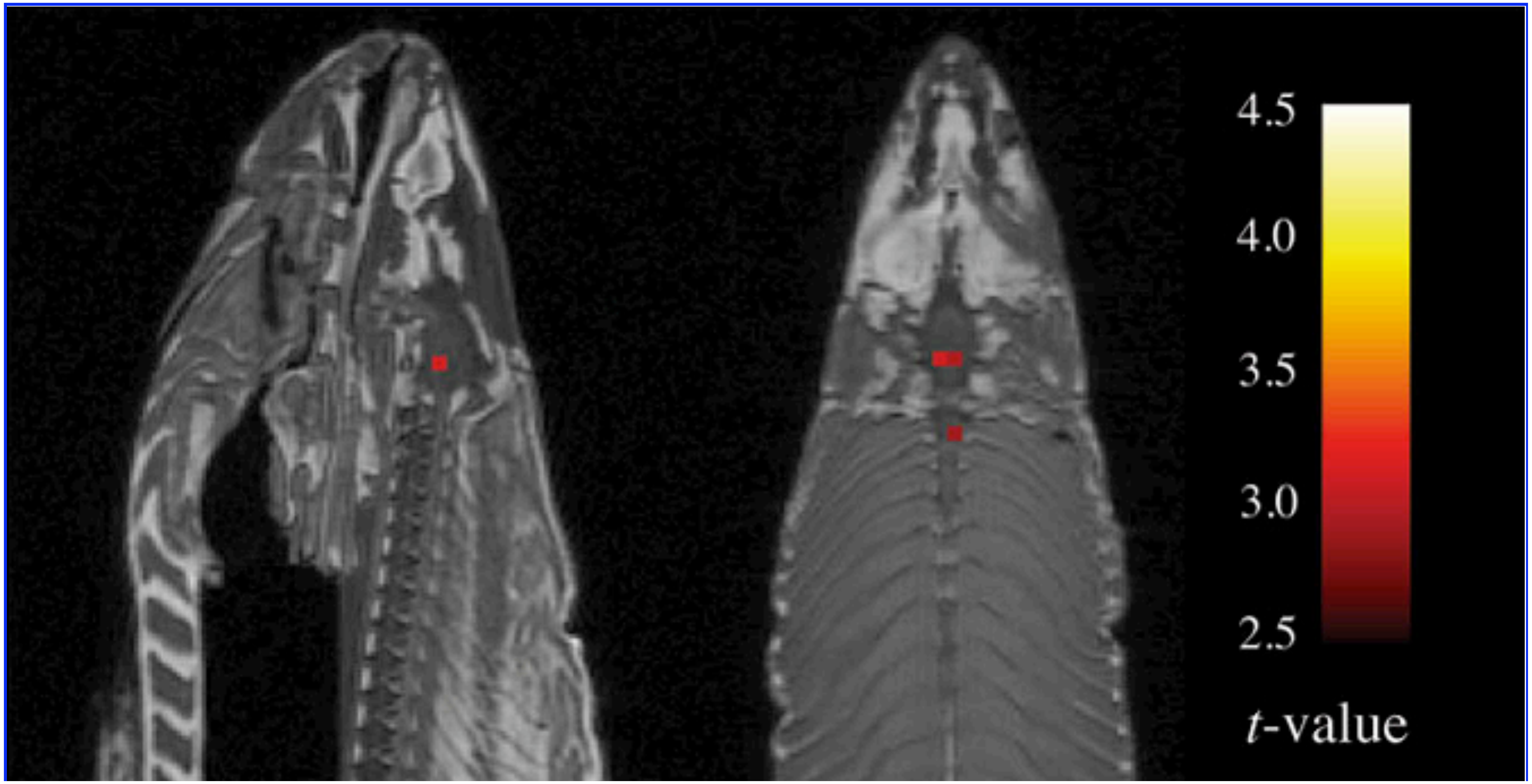
Model Selection

Wired Science
News for Your Neurons
Previous post
Next post

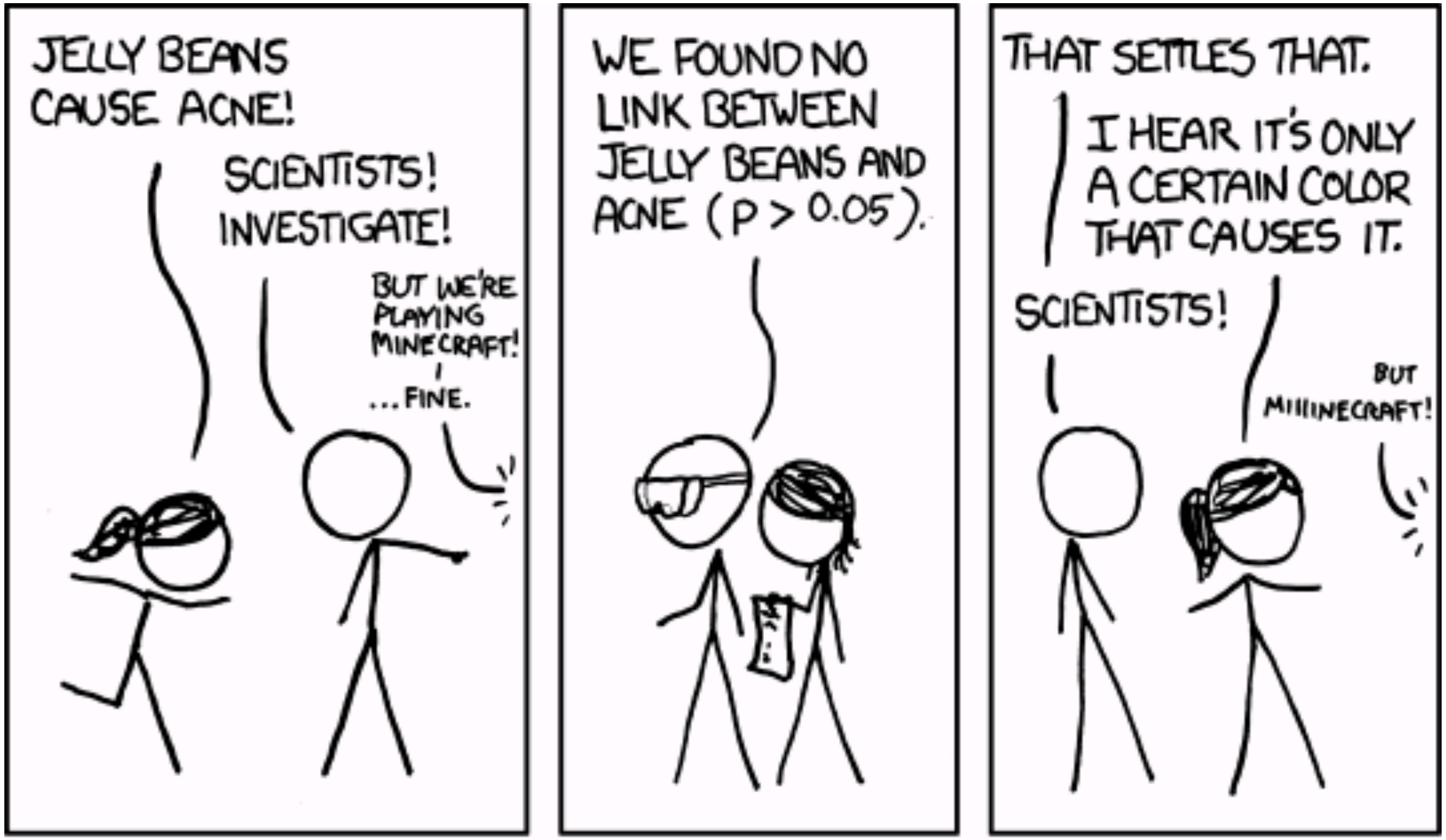
Challenge: huge number of possible models, each with a different pattern of sparsity

Scanning Dead Salmon in fMRI Machine Highlights Risk of Red Herrings

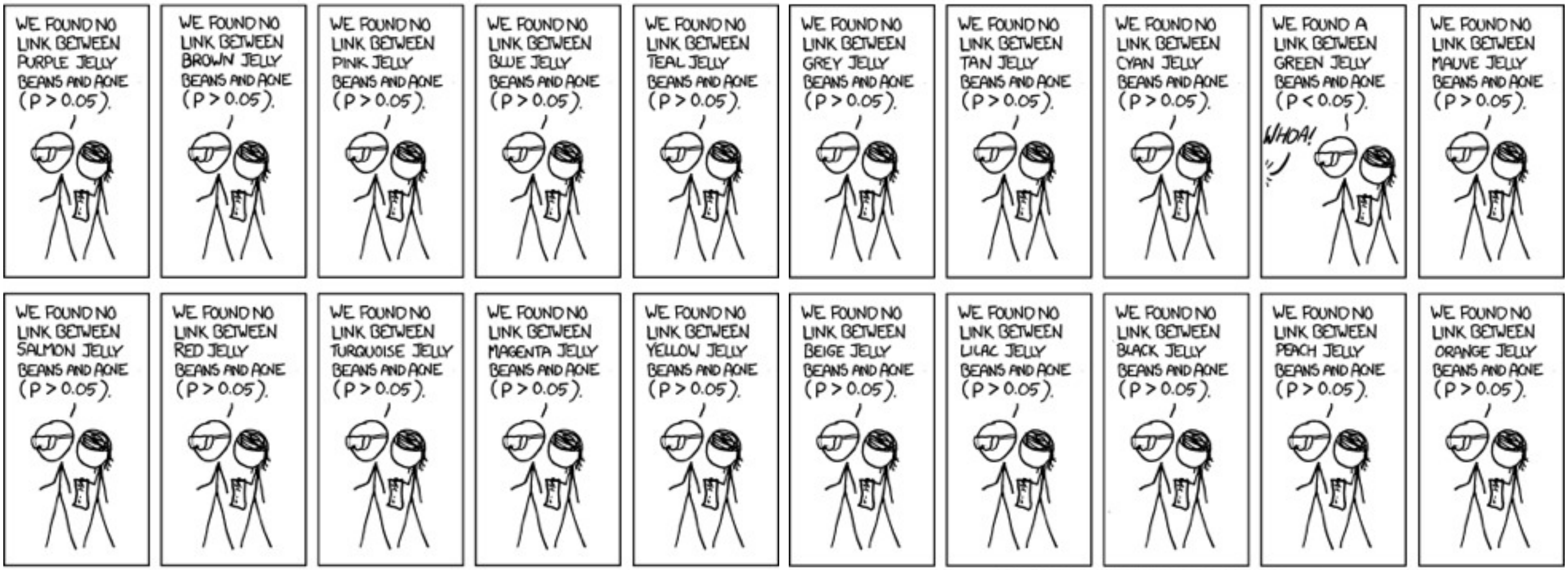
By Alexis Madrigal September 18, 2009 | 5:37 pm | Categories: Brains and Behavior

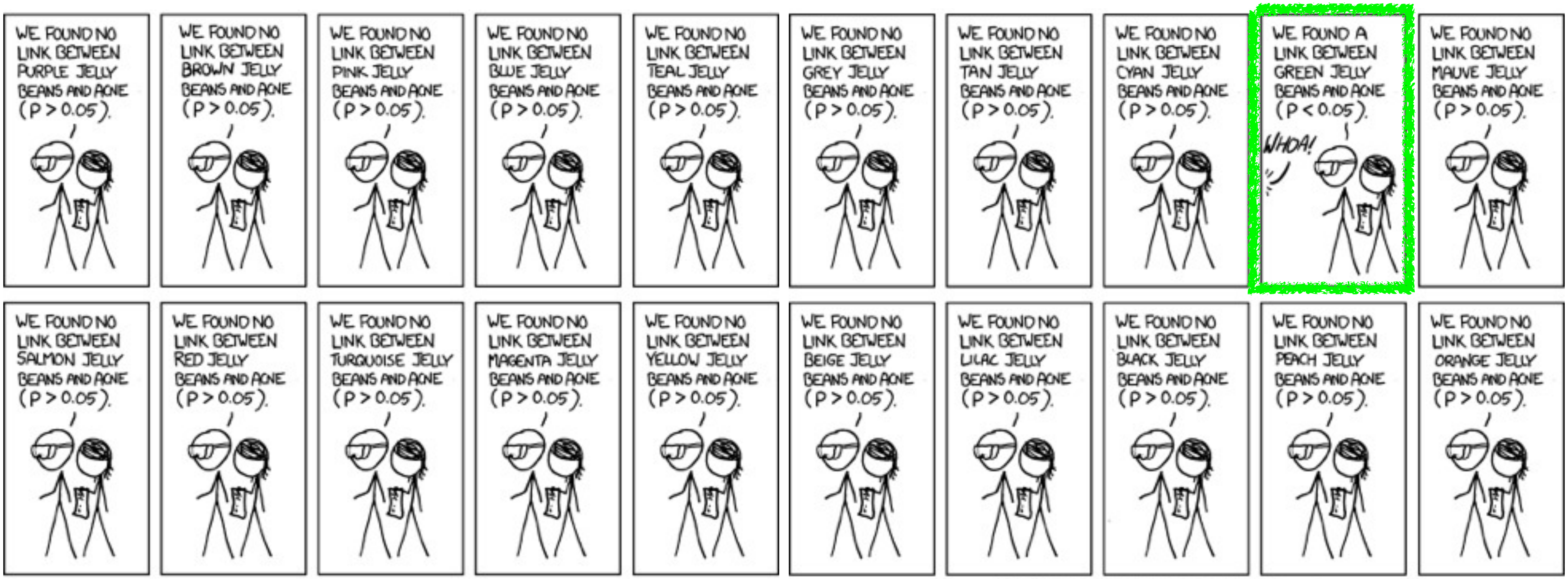


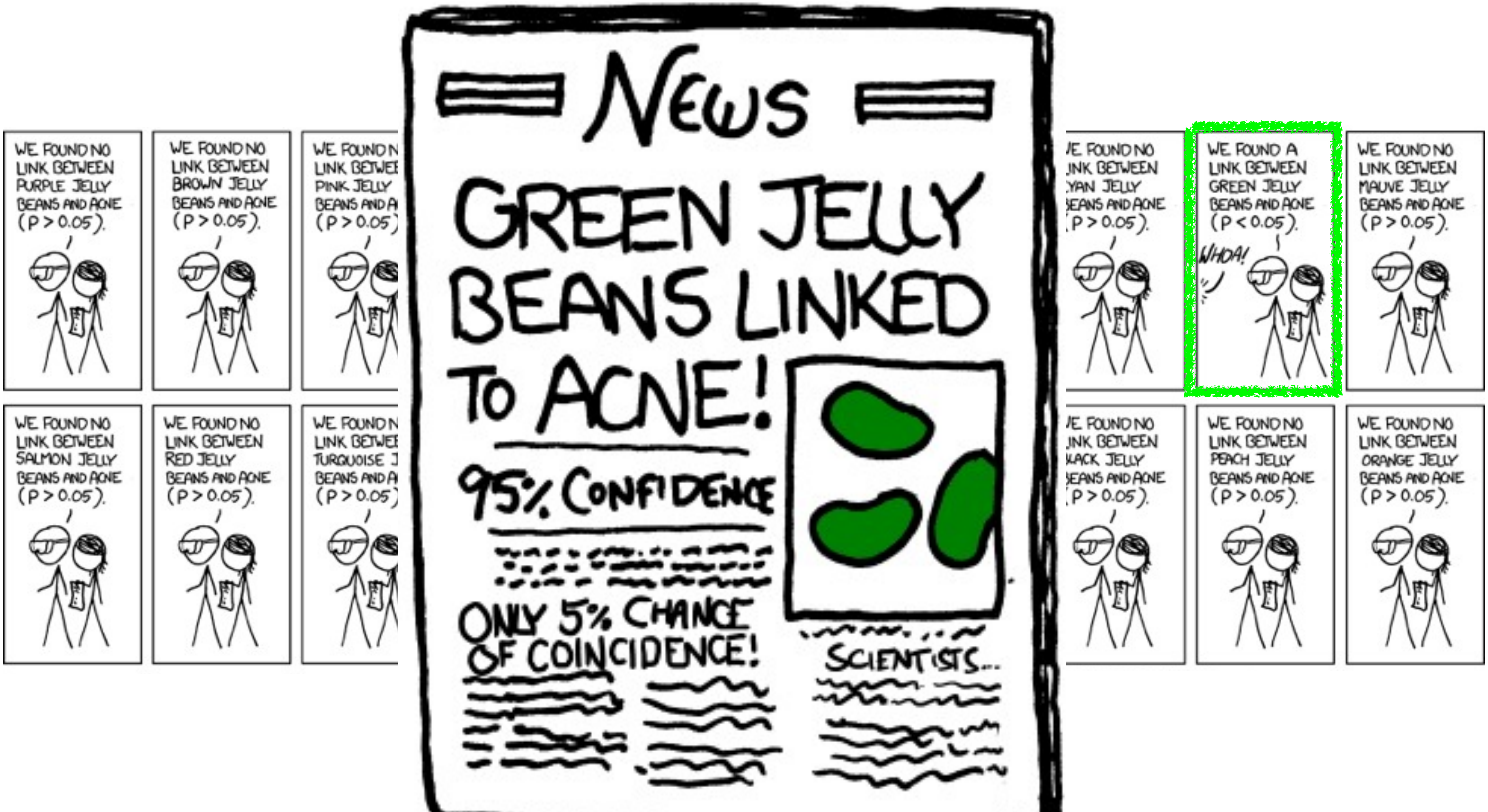
The Multiple Testing Problem



http://xkcd.com/882/


WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).


WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).


WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND N LINK BETWEE PINK JELLY BEANS AND A (P > 0.05
E FOUND NO INK BETWEEN YAN JELLY EANS AND ACNE P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND N LINK BETWEE TURQUOISE J BEANS AND A (P > 0.05
E FOUND NO INK BETWEEN LACK JELLY EANS AND ACNE P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).
NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

Single Experiment Model

$$y_i \; = \; x_i \; + \; z_i \, , \; i = 1, \ldots, n$$

$$z_i \; \overset{\text{iid}}{\sim} \; \mathcal{N}(0, 1)$$

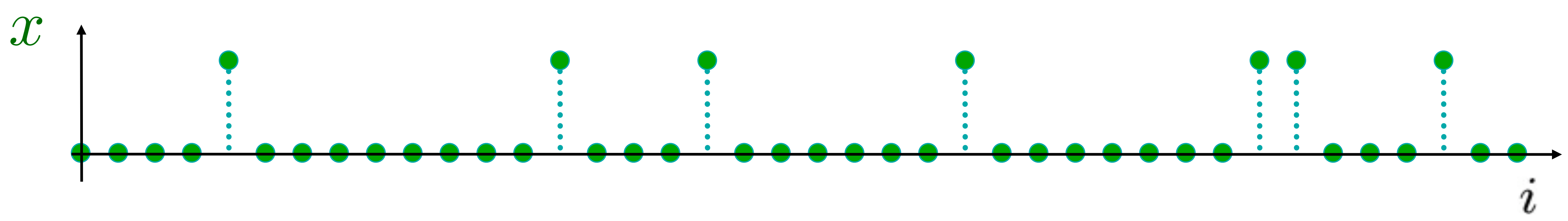
sparsity: $x_i = 0$ except on a small subset $\mathcal{S} \subset \{1, \ldots, n\}$ where $x_i = \mu > 0$

Single Experiment Model

$$y_i \;=\; x_i \;+\; z_i\,,\; i = 1, \ldots, n$$

$$z_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$

sparsity: $x_i = 0$ except on a small subset $\mathcal{S} \subset \{1, \ldots, n\}$ where $x_i = \mu > 0$

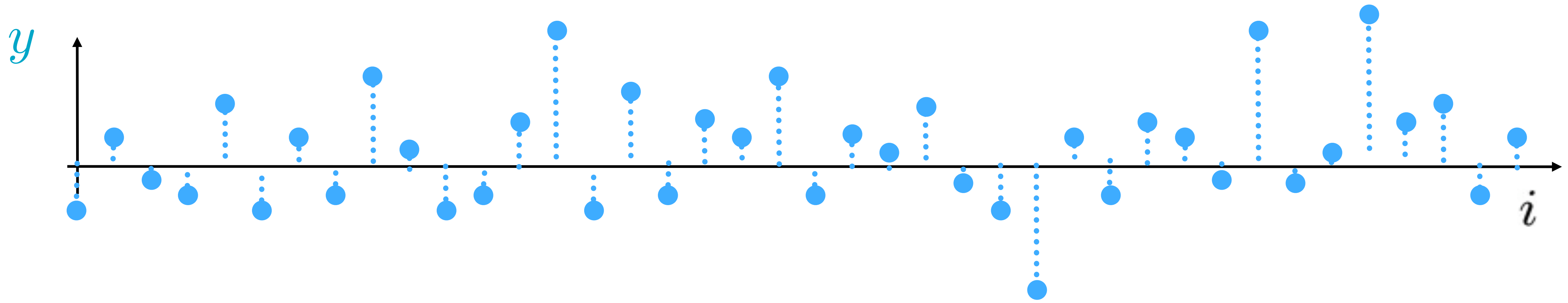


Single Experiment Model

$$y_i \;=\; x_i \;+\; z_i\,,\;\; i = 1, \ldots, n$$

$$z_i \;\overset{\text{iid}}{\sim}\; \mathcal{N}(0,1)$$

sparsity: $x_i = 0$ except on a small subset $\mathcal{S} \subset \{1, \ldots, n\}$ where $x_i = \mu > 0$

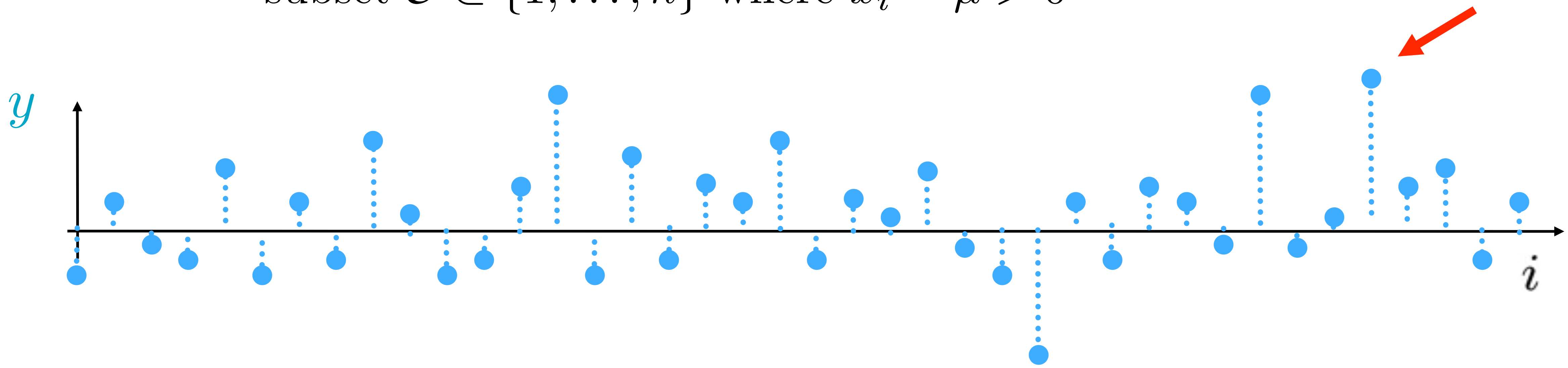


Single Experiment Model

$$y_i \; = \; x_i \; + \; z_i \, , \; i = 1, \ldots, n$$

$$z_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$

sparsity: $x_i = 0$ except on a small subset $\mathcal{S} \subset \{1, \ldots, n\}$ where $x_i = \mu > 0$



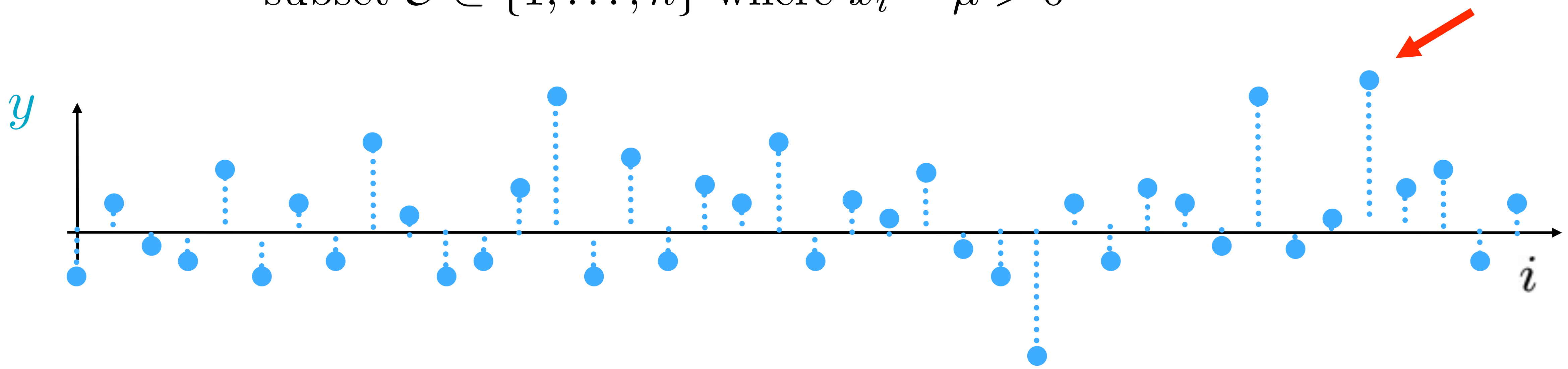
Suppose we want to locate just **one** signal component: $\widehat{i} \; = \; \arg\max_i y_i$

Single Experiment Model

$$y_i \;=\; x_i \;+\; z_i\,,\; i = 1, \ldots, n$$

$$z_i \;\overset{\text{iid}}{\sim}\; \mathcal{N}(0,1)$$

sparsity: $x_i = 0$ except on a small subset $\mathcal{S} \subset \{1, \ldots, n\}$ where $x_i = \mu > 0$



Suppose we want to locate just **one** signal component: $\widehat{i} \;=\; \arg\max_i y_i$

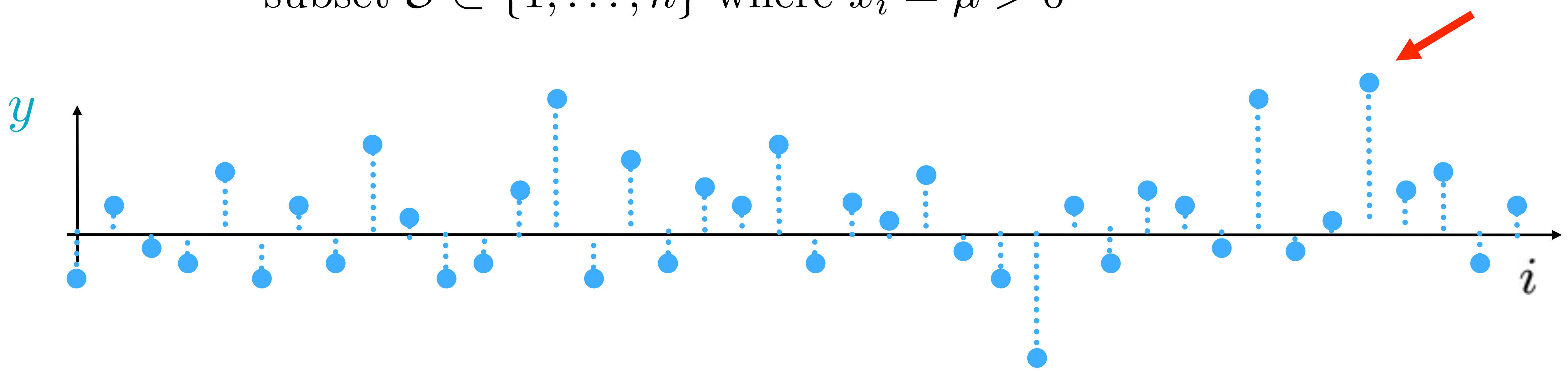
Even if no signal is present, $\max_i y_i \;\sim\; \sqrt{2\log n}$

Single Experiment Model

$$y_i \;=\; x_i \;+\; z_i\,,\;\; i = 1, \ldots, n$$

$$z_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$

sparsity: $x_i = 0$ except on a small subset $\mathcal{S} \subset \{1, \ldots, n\}$ where $x_i = \mu > 0$



Suppose we want to locate just **one** signal component: $\widehat{i} \;=\; \arg\max_i y_i$

Even if no signal is present, $\max_i y_i \;\sim\; \sqrt{2\log n}$

It is *impossible* to reliably detect signal components if $\mu < \sqrt{2\log n}$

An Alternative: Sequential Experimental Design

Instead of the usual non-adaptive observation model

$$y_i \ = \ x_i \ + \ z_i \,, \ i = 1, \dots, n$$

An Alternative: Sequential Experimental Design

Instead of the usual non-adaptive observation model

$$y_i \;=\; x_i \;+\; z_i\,,\; i = 1, \ldots, n$$

suppose we are able to sequentially collect several independent measurements of each component of x, according to

$$y_{i,j} \;=\; x_i \;+\; \gamma_{i,j}^{-1/2}\, z_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, k$$

where

j indexes the measurement steps

k denotes the total number of steps

$z_{i,j} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$

$\gamma_{i,j} \geq 0$ controls the precision of each measurement

An Alternative: Sequential Experimental Design

Instead of the usual non-adaptive observation model

$$y_i \;=\; x_i \;+\; z_i\,,\; i = 1, \ldots, n$$

suppose we are able to sequentially collect several independent measurements of each component of x, according to

$$y_{i,j} \;=\; x_i \;+\; \gamma_{i,j}^{-1/2}\, z_{i,j}\,, \quad i = 1, \ldots, n, \quad j = 1, \ldots, k$$

where

j indexes the measurement steps

k denotes the total number of steps

$z_{i,j} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$

$\gamma_{i,j} \geq 0$ controls the precision of each measurement

Total precision budget is constrained, but the choice of $\gamma_{i,j}$ can depend on past observations $\{y_{i,\ell}\}_{\ell<j}$.

Experimental (Precision) Budget

sequential measurement model

$$y_{i,j} \;=\; x_i \;+\; \gamma_{i,j}^{-1/2}\, z_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, k$$

Experimental (Precision) Budget

sequential measurement model

$$y_{i,j} \;=\; x_i \;+\; \gamma_{i,j}^{-1/2}\, z_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, k$$

The precision parameters $\{\gamma_{i,j}\}$ are required to satisfy

$$\sum_{j=1}^{k} \sum_{i=1}^{n} \gamma_{i,j} \;\leq\; n$$

For example, the usual non-adaptive, single measurement model corresponds to taking $k = 1$, and $\gamma_{i,1} = 1$, $i = 1, \ldots, n$. This baseline can be compared with adaptive procedures by allowing $k > 1$ and variable $\{\gamma_{i,j}\}$ satisfying budget.

Experimental (Precision) Budget

sequential measurement model

$$y_{i,j} \;=\; x_i \;+\; \gamma_{i,j}^{-1/2}\, z_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, k$$

The precision parameters $\{\gamma_{i,j}\}$ are required to satisfy

$$\sum_{j=1}^{k} \sum_{i=1}^{n} \gamma_{i,j} \;\leq\; n$$

For example, the usual non-adaptive, single measurement model corresponds to taking $k = 1$, and $\gamma_{i,1} = 1$, $i = 1, \ldots, n$. This baseline can be compared with adaptive procedures by allowing $k > 1$ and variable $\{\gamma_{i,j}\}$ satisfying budget.

Precision parameters control the SNR per component.

SNR is increased/decreased by

— more/fewer repeated samples or

— longer/shorter observation times

Experimental (Precision) Budget

sequential measurement model

$$y_{i,j} \;=\; x_i \;+\; \gamma_{i,j}^{-1/2}\, z_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, k$$

The precision parameters $\{\gamma_{i,j}\}$ are required to satisfy

$$\sum_{j=1}^{k} \sum_{i=1}^{n} \gamma_{i,j} \;\leq\; n$$

For example, the usual non-adaptive, single measurement model corresponds to taking $k = 1$, and $\gamma_{i,1} = 1$, $i = 1, \ldots, n$. This baseline can be compared with adaptive procedures by allowing $k > 1$ and variable $\{\gamma_{i,j}\}$ satisfying budget.

Precision parameters control the SNR per component.

SNR is increased/decreased by

— more/fewer repeated samples or

— longer/shorter observation times

allocate precision sequentially and adaptively

Sequential Thresholding

Sequential Thresholding

initialize: $\mathcal{S}_0 = \{1, \ldots, n\}$, $\gamma_{i,j}^{-1} = 2$

for $j = 1, \ldots, k$

1) measure: $y_{i,j} \sim \mathcal{N}(x_i, 2)$, $i \in \mathcal{S}_{j-1}$

2) threshold: $\mathcal{S}_j = \{i : y_{i,j} \geq 0\}$

end

output: $\mathcal{S}_k = \{i : y_{i,k} > 0\}$

Sequential Thresholding

precision $= \begin{cases} \frac{1}{2} & , \text{ if measured} \\ 0 & , \text{ otherwise} \end{cases}$

Sequential Thresholding

initialize: $\mathcal{S}_0 = \{1, \ldots, n\}$, $\gamma_{i,j}^{-1} = 2$

for $j = 1, \ldots, k$

1) measure: $y_{i,j} \sim \mathcal{N}(x_i, 2)$, $i \in \mathcal{S}_{j-1}$

2) threshold: $\mathcal{S}_j = \{i : y_{i,j} \geq 0\}$

end

output: $\mathcal{S}_k = \{i : y_{i,k} > 0\}$

Sequential Thresholding

Sequential Thresholding

initialize: $\mathcal{S}_0 = \{1, \ldots, n\}$, $\gamma_{i,j}^{-1} = 2$

for $j = 1, \ldots, k$

1) measure: $y_{i,j} \sim \mathcal{N}\left(x_i, 2\right) \;\; , \;\; i \in \mathcal{S}_{j-1}$

2) threshold: $\mathcal{S}_j = \{i : y_{i,j} \geq 0\}$

end

output: $\mathcal{S}_k \;\; = \;\; \{i \; : \; y_{i,k} > 0\}$

define the *sparsity level*

$s := |\mathcal{S}|$

Sequential Thresholding

Sequential Thresholding

initialize: $\mathcal{S}_0 = \{1, \dots, n\}$, $\gamma_{i,j}^{-1} = 2$

for $j = 1, \dots, k$

1) measure: $y_{i,j} \sim \mathcal{N}(x_i, 2)$, $i \in \mathcal{S}_{j-1}$

2) threshold: $\mathcal{S}_j = \{i : y_{i,j} \geq 0\}$

end

output: $\mathcal{S}_k = \{i : y_{i,k} > 0\}$

define the *sparsity level*

$$s := |\mathcal{S}|$$

total precision budget: $\mathbb{E}\left[\sum_{i,j} \gamma_{i,j}\right]$

$$
\begin{aligned}
&= \frac{1}{2} \sum_{j=1}^{k} \mathbb{E}|\mathcal{S}_{j-1}| \\
&\leq \frac{1}{2} \sum_{j=1}^{k} \left(\frac{n-s}{2^{j-1}} + s \right) \\
&\leq n - s + ks \approx n
\end{aligned}
$$

(when $n \gg s$)

Sequential Thresholding

Sequential Thresholding

initialize: $\mathcal{S}_0 = \{1, \ldots, n\}$, $\gamma_{i,j}^{-1} = 2$
for $j = 1, \ldots, k$
1) measure: $y_{i,j} \sim \mathcal{N}(x_i, 2)$, $i \in \mathcal{S}_{j-1}$
2) threshold: $\mathcal{S}_j = \{i : y_{i,j} \geq 0\}$
end
output: $\mathcal{S}_k = \{i : y_{i,k} > 0\}$

define the *sparsity level*
$s := |\mathcal{S}|$

total precision budget: $\mathbb{E}\left[\sum_{i,j} \gamma_{i,j}\right]$

$$= \frac{1}{2}\sum_{j=1}^{k} \mathbb{E}|\mathcal{S}_{j-1}|$$

$$\leq \frac{1}{2}\sum_{j=1}^{k}\left(\frac{n-s}{2^{j-1}} + s\right)$$

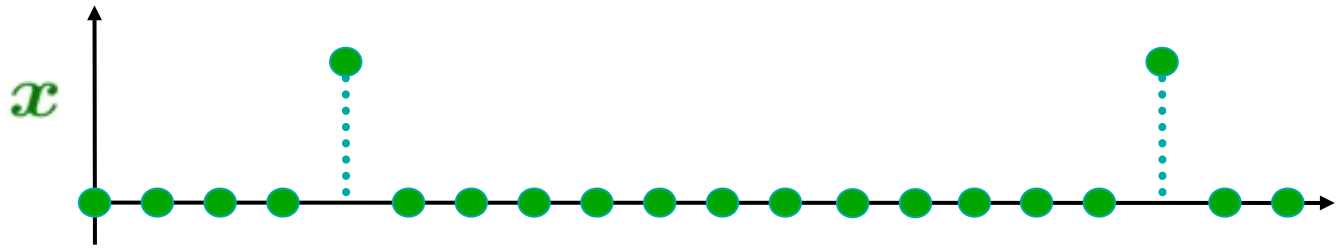
$$\leq n - s + ks \approx n$$

(when $n \gg s$)

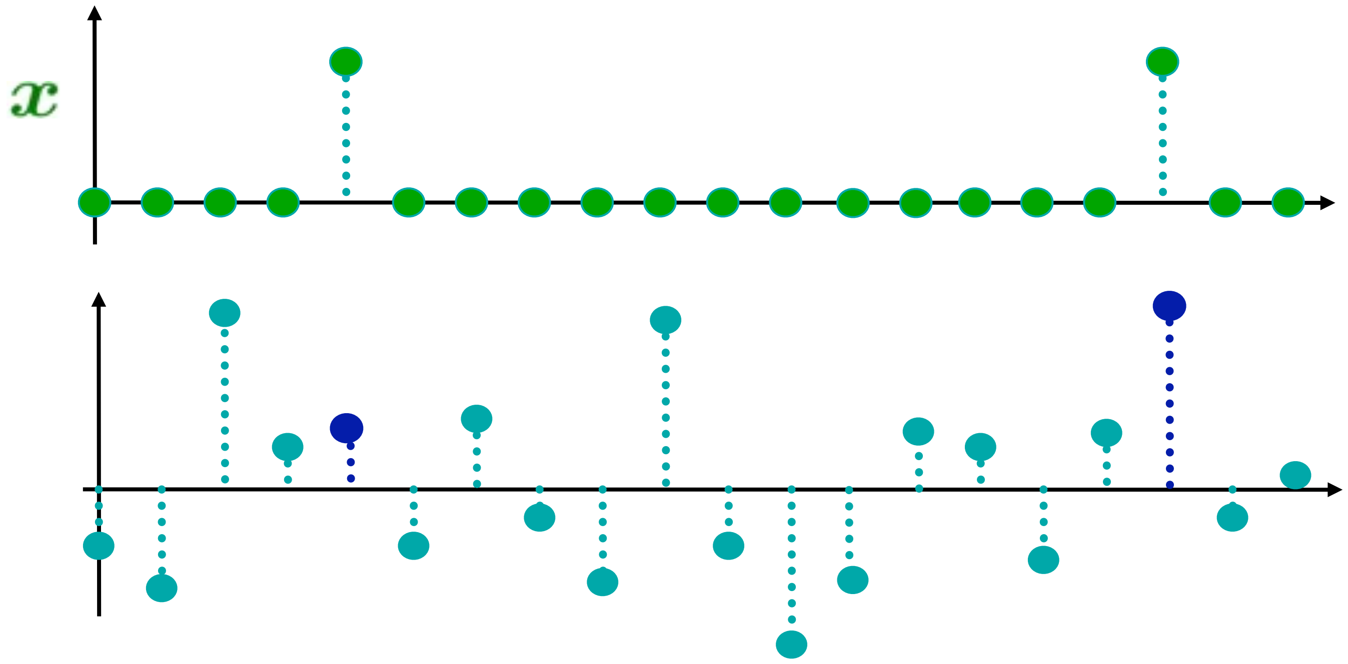
probability of error: $\mathbb{P}(\mathcal{S}_k \neq \mathcal{S}) = \mathbb{P}\left(\{\mathcal{S}^c \cap \mathcal{S}_k \neq \emptyset\} \cup \{\mathcal{S} \cap \mathcal{S}_k^c \neq \emptyset\}\right)$

$$\leq \mathbb{P}\left(\mathcal{S}^c \cap \mathcal{S}_k \neq \emptyset\right) + \mathbb{P}\left(\mathcal{S} \cap \mathcal{S}_k^c \neq \emptyset\right)$$

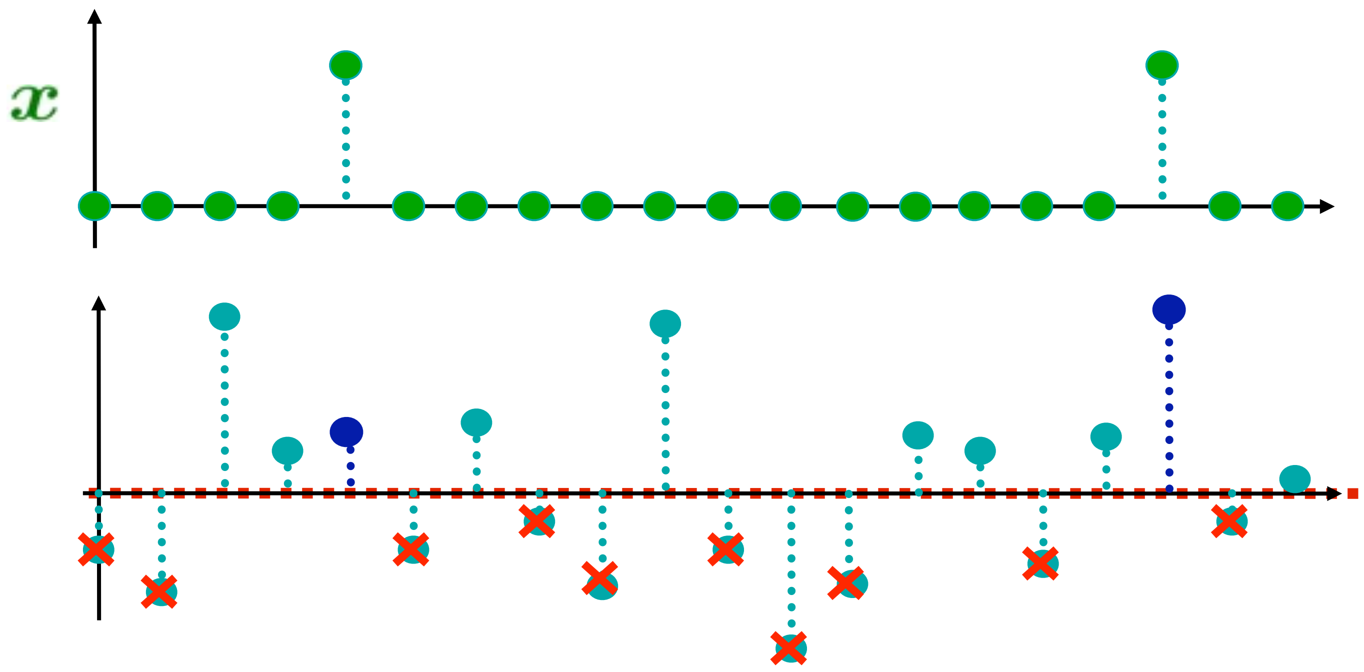
Idealized Example



Idealized Example

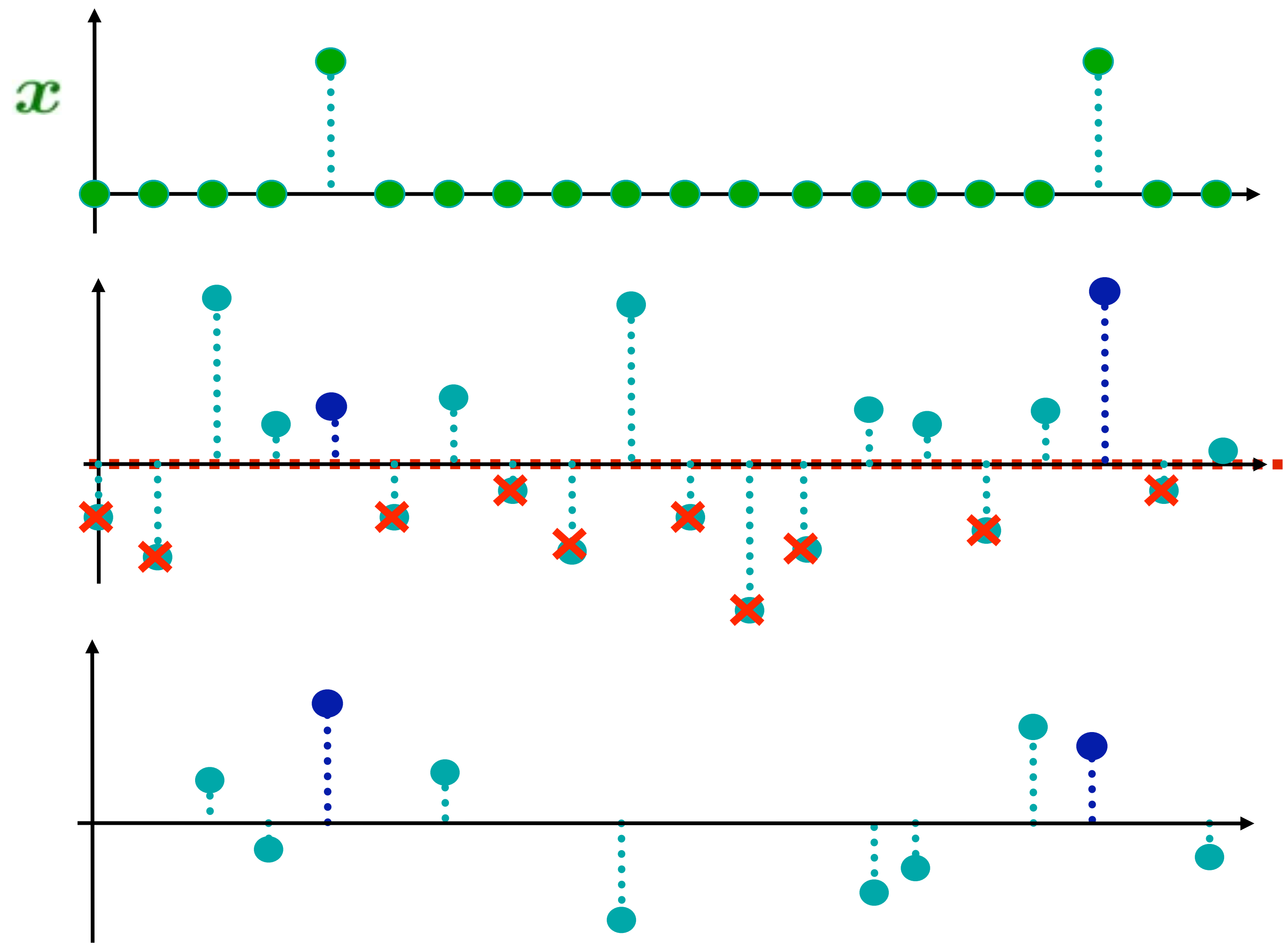


Idealized Example



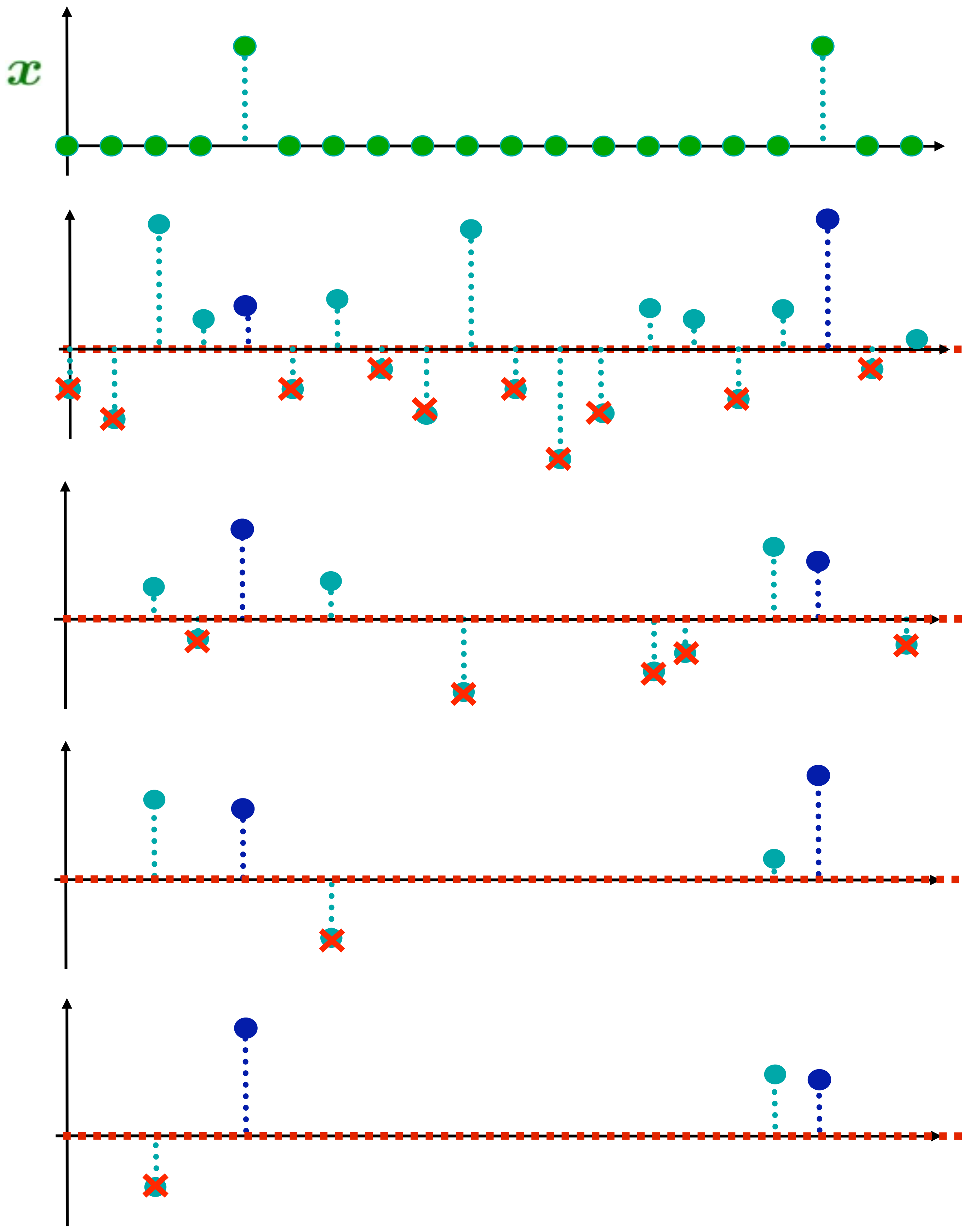
threshold at zero and re-measure
only those components that survive

Idealized Example



threshold at zero and re-measure
only those components that survive

Idealized Example

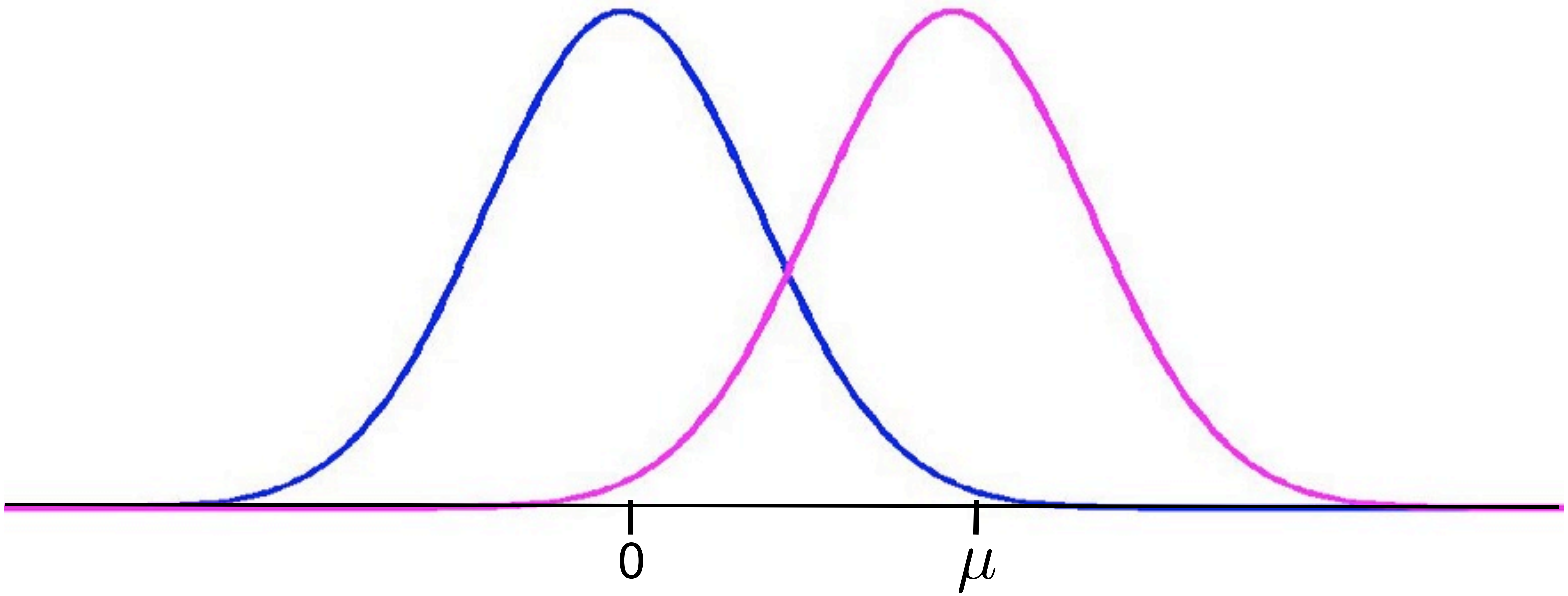


threshold at zero and re-measure only those components that survive

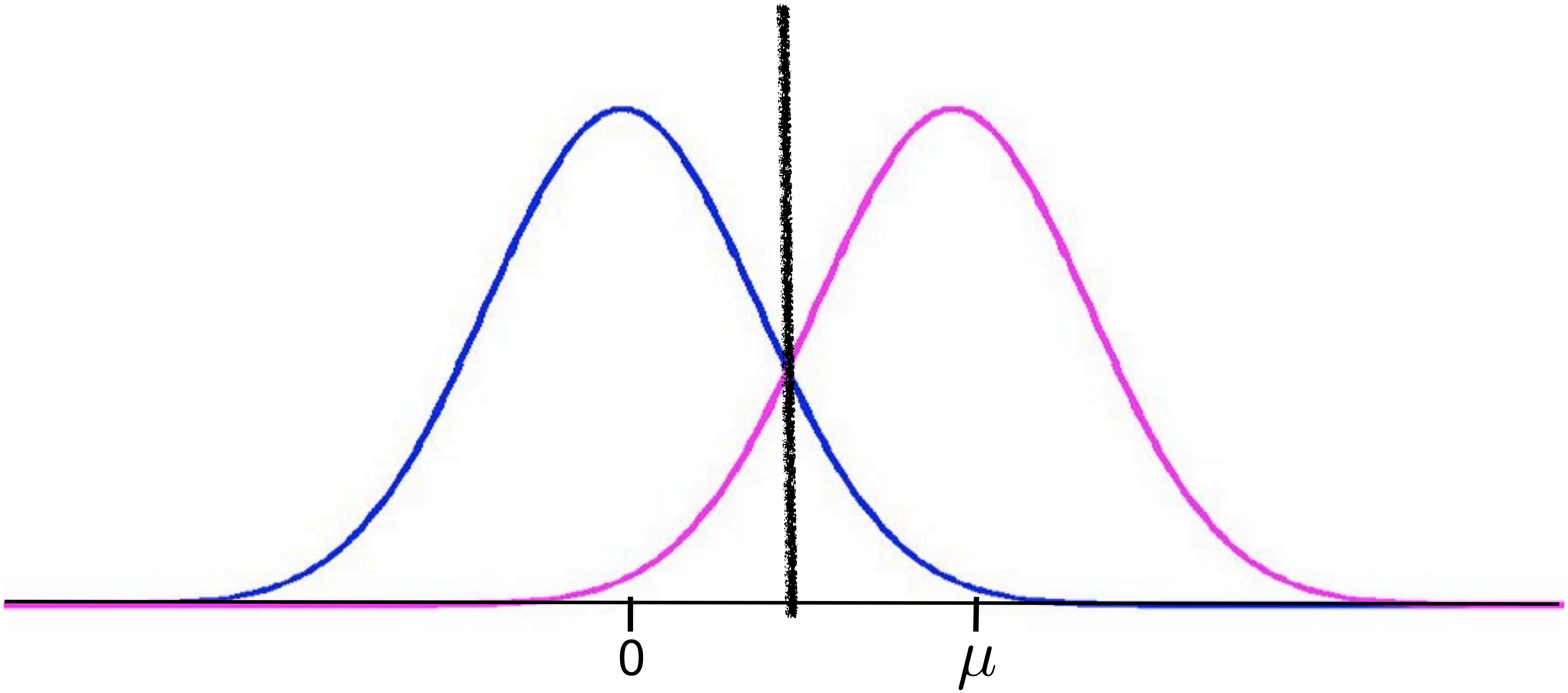
repeat several times

most of true signal components survive several thresholding steps, almost all of noise components do not

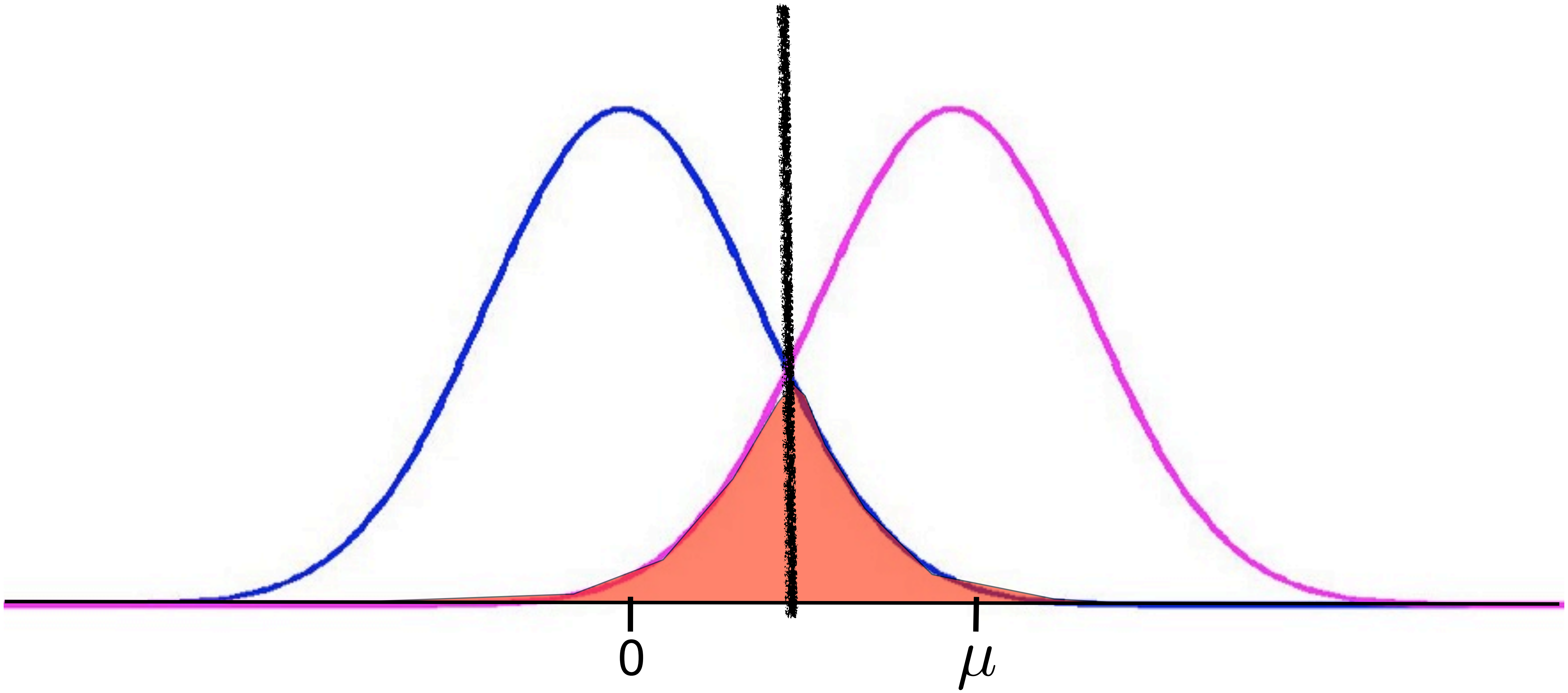
Low False-Negative, High False-Positive Rates



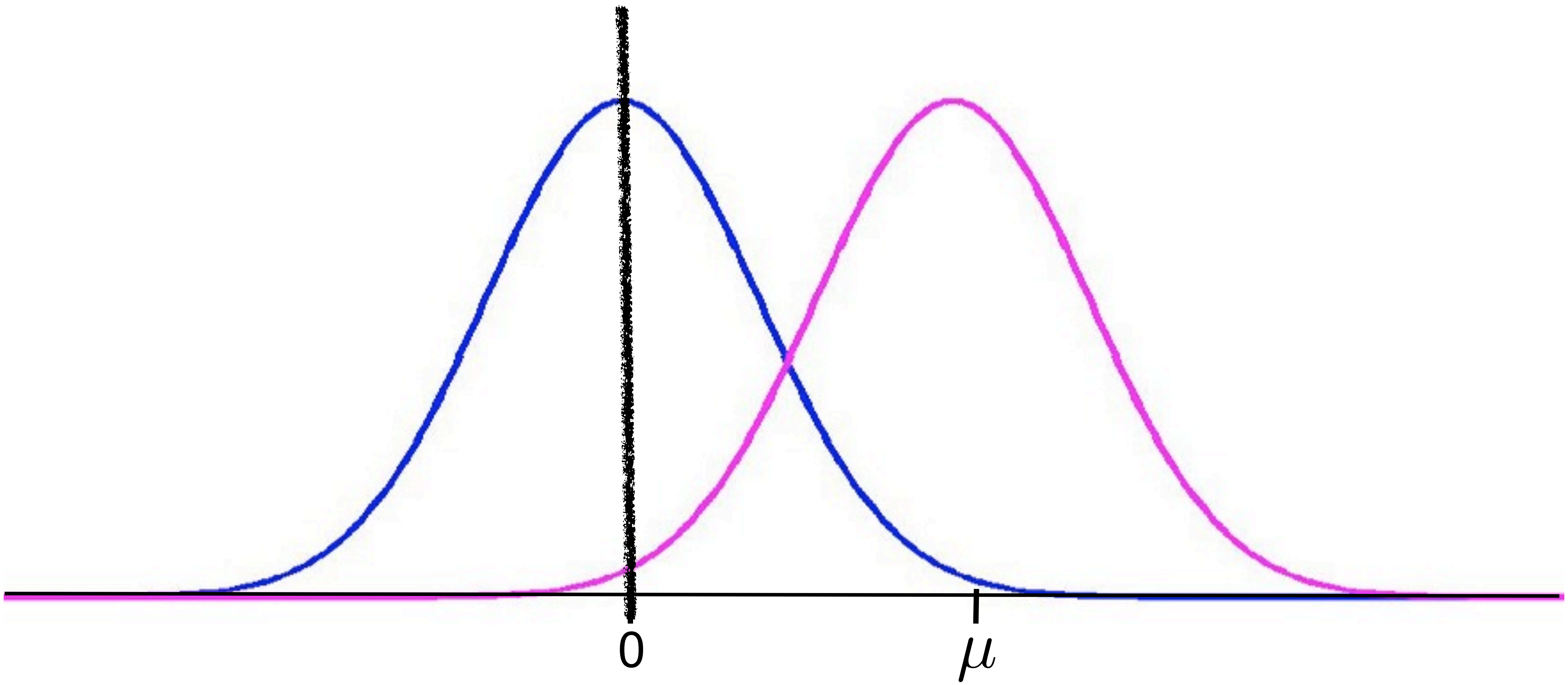
Low False-Negative, High False-Positive Rates



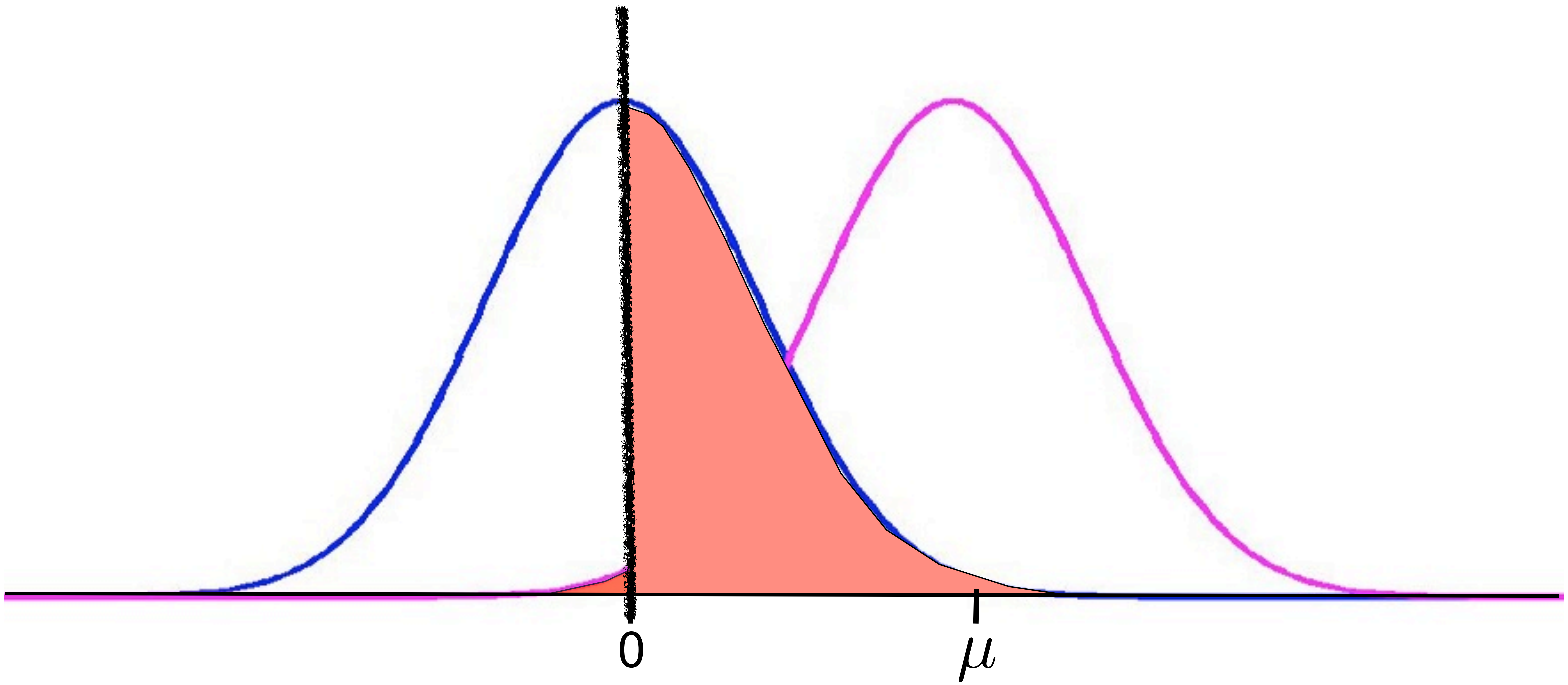
Low False-Negative, High False-Positive Rates



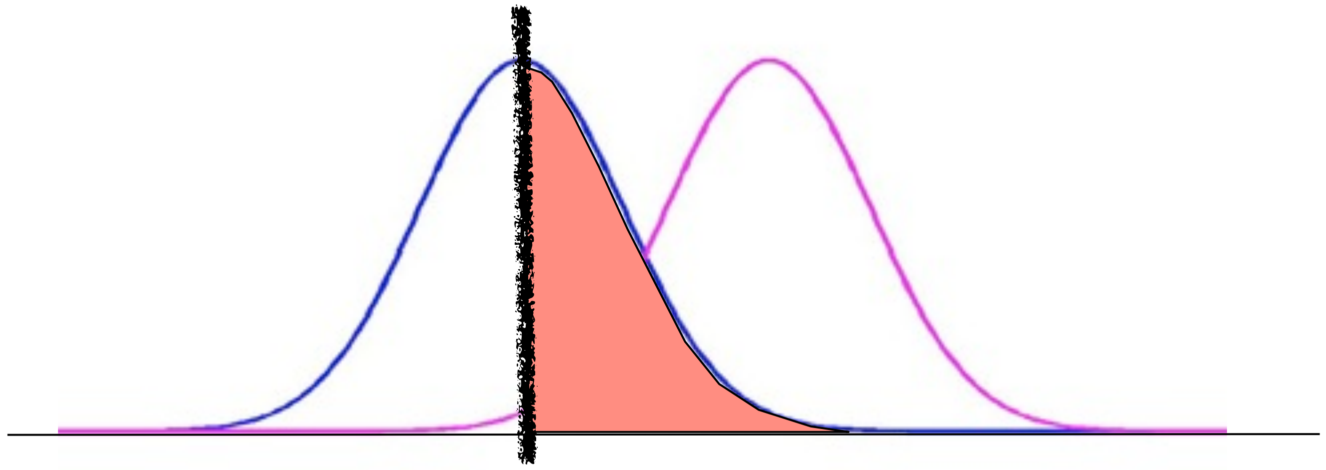
Low False-Negative, High False-Positive Rates



Low False-Negative, High False-Positive Rates



False Positives



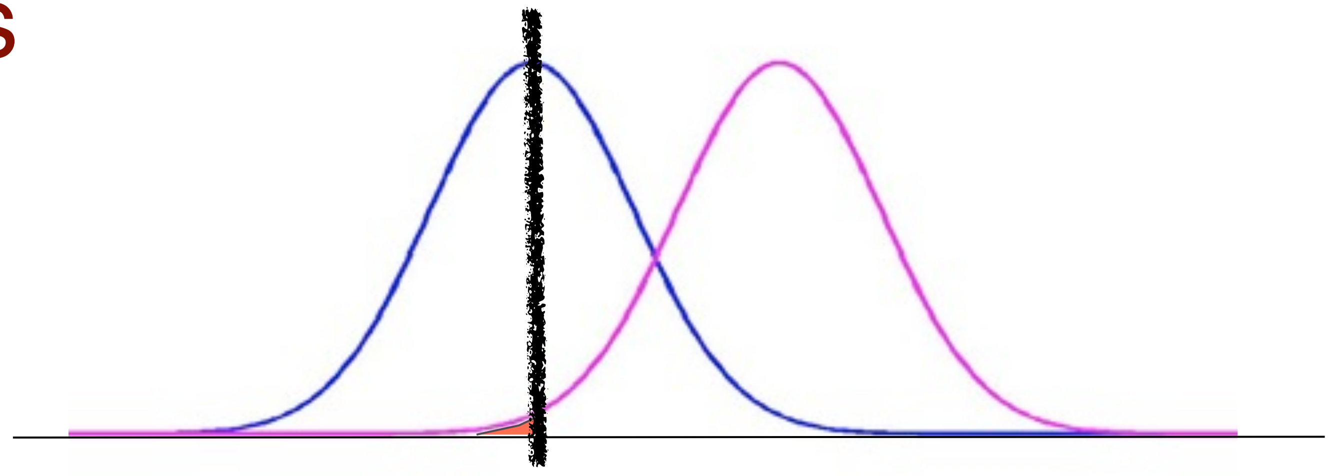
$$\mathbb{P}(\mathcal{S}_k \neq \mathcal{S}) \quad \leq \quad \mathbb{P}\left(\mathcal{S}^c \cap \mathcal{S}_k \neq \emptyset\right) \; + \; \mathbb{P}\left(\mathcal{S} \cap \mathcal{S}_k^c \neq \emptyset\right)$$

False Positives

$$\mathbb{P}(\mathcal{S}_k \neq \mathcal{S}) \quad \leq \quad \mathbb{P}\left(\mathcal{S}^c \cap \mathcal{S}_k \neq \emptyset\right) \; + \; \mathbb{P}\left(\mathcal{S} \cap \mathcal{S}_k^c \neq \emptyset\right)$$

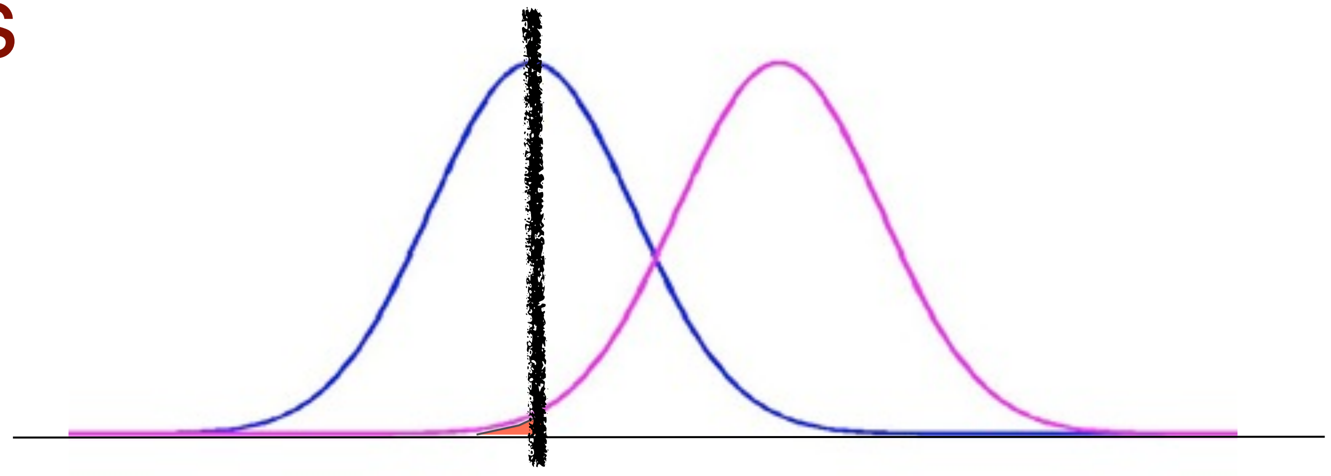
$$\begin{aligned}
\mathbb{P}\left(\mathcal{S}^c \cap \mathcal{S}_k \neq \emptyset\right) \quad &= \quad \mathbb{P}\left(\bigcup_{i \notin \mathcal{S}} \bigcap_{j=1}^{k} y_{i,j} > 0\right) \\
&\leq \quad \sum_{i \notin \mathcal{S}} \mathbb{P}\left(\bigcap_{j=1}^{k} y_{i,j} > 0\right) \\
&= \quad \sum_{i \notin \mathcal{S}} 2^{-k} \;=\; \frac{n-s}{2^k}
\end{aligned}$$

False Negatives



$$\mathbb{P}(\mathcal{S}_k \neq \mathcal{S}) \quad \leq \quad \mathbb{P}\left(\mathcal{S}^c \cap \mathcal{S}_k \neq \emptyset\right) \; + \; \mathbb{P}\left(\mathcal{S} \cap \mathcal{S}_k^c \neq \emptyset\right)$$

False Negatives



$$\mathbb{P}(\mathcal{S}_k \neq \mathcal{S}) \quad \leq \quad \mathbb{P}\left(\mathcal{S}^c \cap \mathcal{S}_k \neq \emptyset\right) \; + \; \mathbb{P}\left(\mathcal{S} \cap \mathcal{S}_k^c \neq \emptyset\right)$$

$$\begin{aligned}\mathbb{P}\left(\mathcal{S} \cap \mathcal{S}_k^c \neq \emptyset\right) \quad &= \quad \mathbb{P}\left(\bigcup_{j=1}^{k} \bigcup_{i \in \mathcal{S}} y_{i,j} < 0\right) \\ &\leq \quad \frac{ks}{2} \exp\left(-\frac{\mu^2}{4}\right)\end{aligned}$$

Probability of Error Bound

$$
\begin{aligned}
\mathbb{P}(\mathcal{S}_k \neq \mathcal{S}) &\leq \mathbb{P}\left(\mathcal{S}^c \cap \mathcal{S}_k \neq \emptyset\right) + \mathbb{P}\left(\mathcal{S} \cap \mathcal{S}_k^c \neq \emptyset\right) \\
&\leq \frac{n-s}{2^k} + \frac{ks}{2} \exp\left(-\frac{\mu^2}{4}\right) \\
&= \frac{n-s}{2^k} + \frac{1}{2} \exp\left(-\frac{(\mu^2 - 4\log(ks))}{4}\right)
\end{aligned}
$$

Probability of Error Bound

$$
\begin{aligned}
\mathbb{P}(\mathcal{S}_k \neq \mathcal{S}) &\leq \mathbb{P}\left(\mathcal{S}^c \cap \mathcal{S}_k \neq \emptyset\right) + \mathbb{P}\left(\mathcal{S} \cap \mathcal{S}_k^c \neq \emptyset\right) \\
&\leq \frac{n-s}{2^k} + \frac{ks}{2} \exp\left(-\frac{\mu^2}{4}\right) \\
&= \frac{n-s}{2^k} + \frac{1}{2} \exp\left(-\frac{(\mu^2 - 4\log(ks))}{4}\right)
\end{aligned}
$$

Consider high-dimensional limit as $n \to \infty$ and take $k = \log_2 n^{1+\epsilon}$

$$
\mathbb{P}(\mathcal{S}_k \neq \mathcal{S}) \leq \frac{n-s}{2^k} + \frac{1}{2} \exp\left(-\frac{(\mu^2 - 4\log(s(1+\epsilon)\log_2 n))}{4}\right)
$$

Probability of Error Bound

$$
\begin{aligned}
\mathbb{P}(\mathcal{S}_k \neq \mathcal{S}) \quad &\leq \quad \mathbb{P}\left(\mathcal{S}^c \cap \mathcal{S}_k \neq \emptyset\right) \; + \; \mathbb{P}\left(\mathcal{S} \cap \mathcal{S}_k^c \neq \emptyset\right) \\
&\leq \quad \frac{n-s}{2^k} \; + \; \frac{ks}{2} \exp\left(-\frac{\mu^2}{4}\right) \\
&= \quad \frac{n-s}{2^k} \; + \; \frac{1}{2} \exp\left(-\frac{(\mu^2 - 4\log(ks))}{4}\right)
\end{aligned}
$$

Consider high-dimensional limit as $n \to \infty$ and take $k = \log_2 n^{1+\epsilon}$

$$
\mathbb{P}(\mathcal{S}_k \neq \mathcal{S}) \quad \leq \quad \underbrace{\frac{n-s}{2^k}}_{0} \; + \; \frac{1}{2} \exp\left(-\frac{(\mu^2 - 4\log(s(1+\epsilon)\log_2 n))}{4}\right)
$$

Probability of Error Bound

$$
\begin{aligned}
\mathbb{P}(\mathcal{S}_k \neq \mathcal{S}) &\leq \mathbb{P}\left(\mathcal{S}^c \cap \mathcal{S}_k \neq \emptyset\right) + \mathbb{P}\left(\mathcal{S} \cap \mathcal{S}_k^c \neq \emptyset\right) \\
&\leq \frac{n-s}{2^k} + \frac{ks}{2} \exp\left(-\frac{\mu^2}{4}\right) \\
&= \frac{n-s}{2^k} + \frac{1}{2} \exp\left(-\frac{(\mu^2 - 4\log(ks))}{4}\right)
\end{aligned}
$$

Consider high-dimensional limit as $n \to \infty$ and take $k = \log_2 n^{1+\epsilon}$

$$
\mathbb{P}(\mathcal{S}_k \neq \mathcal{S}) \leq \cancelto{0}{\frac{n-s}{2^k}} + \frac{1}{2} \exp\left(-\frac{(\mu^2 - 4\log(s(1+\epsilon)\log_2 n))}{4}\right)
$$

Second term tends to zero if

$$
\mu \geq \sqrt{c \log(s \log_2 n)}\ , \ \text{for any } c > 4
$$

Gains of Sequential Design

Rui Castro
(Eindhoven)

Jarvis Haupt
(Minnesota)

Matt Malloy
(Madison)

non-sequential: $\mu > \sqrt{2 \log n}$ (necessary)

sequential thresholding: (sufficient)

$$\mu > \sqrt{4(\log s + \log \log_2 n)}$$

with a bit more work we can show

$$\mu > \sqrt{2(\log s + \log \log_2 \log n)}$$ suffices

Gains of Sequential Design

Rui Castro (Eindhoven) **Jarvis Haupt** (Minnesota) **Matt Malloy** (Madison)

non-sequential: $\mu > \sqrt{2 \log n}$ (necessary)

sequential thresholding: (sufficient)

$$\mu > \sqrt{4(\log s + \log \log_2 n)}$$

with a bit more work we can show

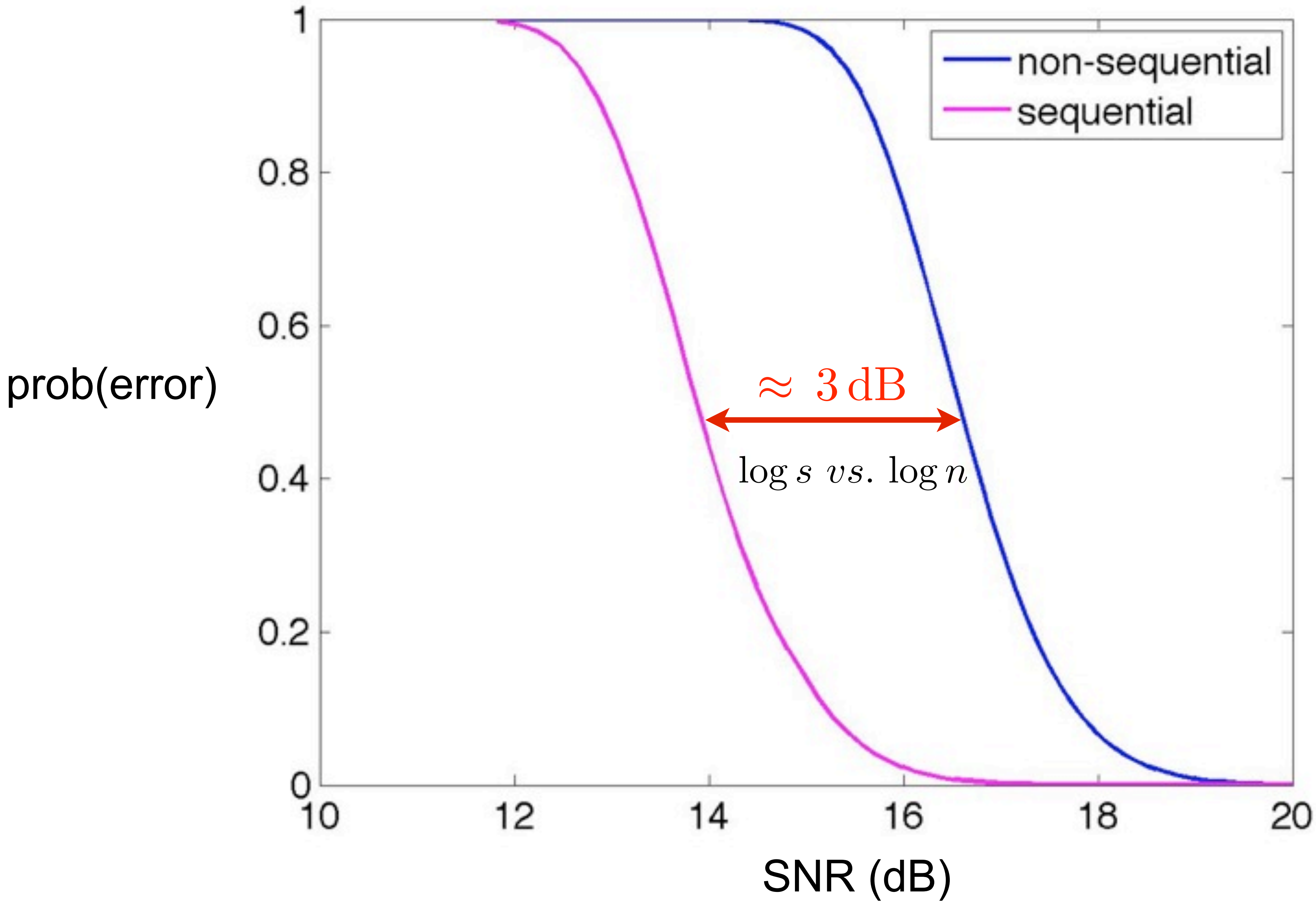
$$\mu > \sqrt{2(\log s + \log \log_2 \log n)} \quad \text{suffices}$$

significant gains when $s \ll n$

greater sensitivity for same precision budget or lower experimental requirements for equivalent sensitivity

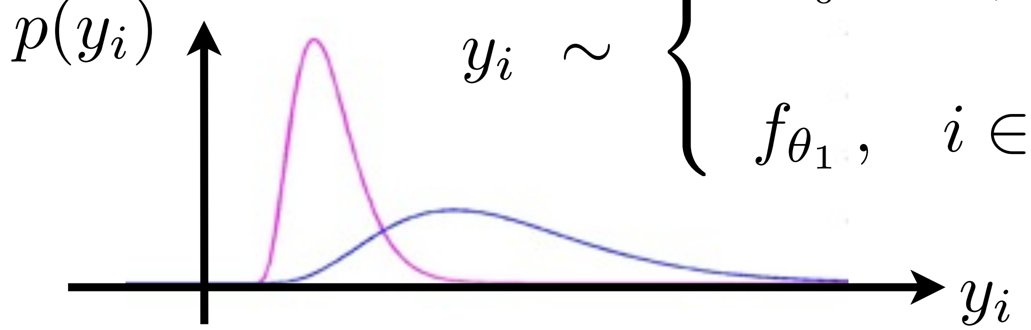
Biology Example

13,071 single-gene knock-down cell strains



sequential thresholding is about twice as sensitive (for equal experimental budget) or requires half the number experiments (for same sensitivity)

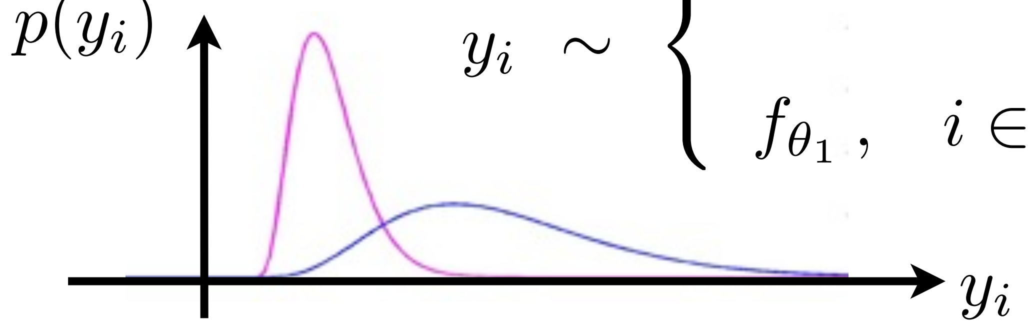
Lower Bound



$$y_i \sim \begin{cases} f_{\theta_0}\,, & i \notin \mathcal{S} \\ f_{\theta_1}\,, & i \in \mathcal{S} \end{cases}$$

assume f_{θ_0}, f_{θ_1} and sparsity level s are known

Lower Bound



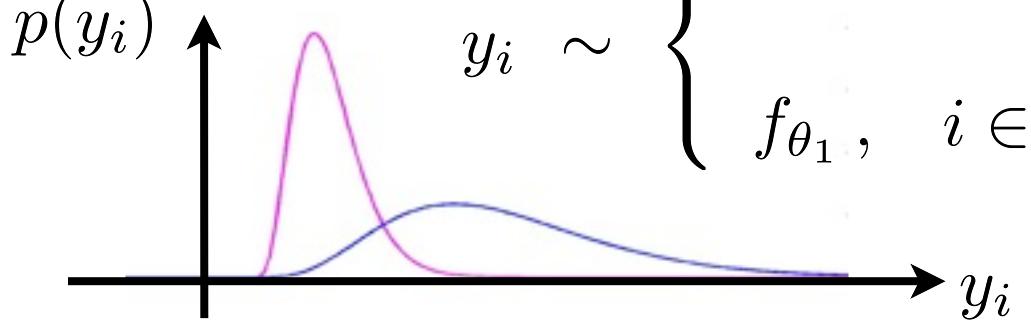
$$y_i \sim \begin{cases} f_{\theta_0}\,, & i \notin \mathcal{S} \\ f_{\theta_1}\,, & i \in \mathcal{S} \end{cases}$$

assume $f_{\theta_0}, f_{\theta_1}$ and sparsity level s are known

specify error rates (per-test):

$$\begin{aligned} \text{false-positive probability: } \alpha &= \epsilon/(n-s) \\ \text{false-negative probability: } \beta &= \epsilon/s \end{aligned}$$

Lower Bound



$$y_i \sim \begin{cases} f_{\theta_0}, & i \notin \mathcal{S} \\ f_{\theta_1}, & i \in \mathcal{S} \end{cases}$$

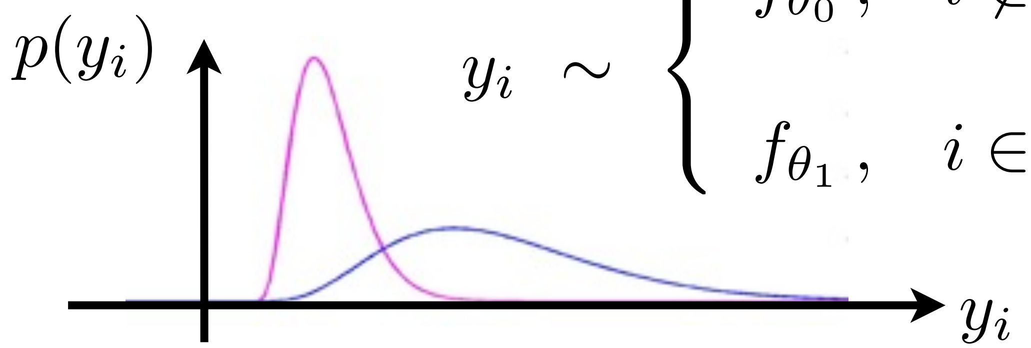
assume f_{θ_0}, f_{θ_1} and sparsity level s are known

specify error rates (per-test):

$$\text{false-positive probability:} \quad \alpha \;=\; \epsilon/(n-s)$$
$$\text{false-negative probability:} \quad \beta \;=\; \epsilon/s$$

expected number of errors: $\alpha\,(n-s) + \beta\,s \;=\; 2\epsilon$

Lower Bound



$$y_i \sim \begin{cases} f_{\theta_0}\,, & i \notin \mathcal{S} \\ f_{\theta_1}\,, & i \in \mathcal{S} \end{cases}$$

assume $f_{\theta_0}, f_{\theta_1}$ and sparsity level s are known

specify error rates (per-test):

$$\begin{aligned} \text{false-positive probability: } \alpha &= \epsilon/(n-s) \\ \text{false-negative probability: } \beta &= \epsilon/s \end{aligned}$$

expected number of errors: $\alpha\,(n-s) + \beta\, s \;=\; 2\epsilon$

expected number samples (precision per-test): for $\alpha, \beta \approx 0$

$$\mathbb{E}_0[M] \;\gtrsim\; D_0^{-1} \log \frac{1}{\beta} \;=\; D_0^{-1} \log \frac{s}{\epsilon}$$

$$\mathbb{E}_1[M] \;\gtrsim\; D_1^{-1} \log \frac{1}{\alpha} \;=\; D_1^{-1} \log \frac{n-s}{\epsilon}$$

where D_0, D_1 are KL divergences $D_0 := D(f_{\theta_0} \| f_{\theta_1})$ and $D_1 := D(f_{\theta_1} \| f_{\theta_0})$

Lower Bound

expected total sampling/precision:

$$\mathbb{E}[N] \quad = \quad (n-s)\mathbb{E}_0[M] + s\mathbb{E}_1[M] \;\gtrsim\; \frac{n}{D_0} \log \frac{s}{\epsilon} \;, \text{ when } s \ll n$$

Lower Bound

expected total sampling/precision:

$$\mathbb{E}[N] \quad = \quad (n-s)\mathbb{E}_0[M] + s\mathbb{E}_1[M] \gtrsim \frac{n}{D_0}\log\frac{s}{\epsilon}\ , \text{ when } s \ll n$$

sampling/precision budget: $\mathbb{E}[N] \leq n \;\Rightarrow\; D_0 \;\geq\; \log s/\epsilon$

minimum requirement for any testing scheme
with expected sample/precision budget m

Lower Bound

expected total sampling/precision:

$$\mathbb{E}[N] \quad = \quad (n-s)\mathbb{E}_0[M] + s\mathbb{E}_1[M] \; \gtrsim \; \frac{n}{D_0} \log \frac{s}{\epsilon} \;, \; \text{when } s \ll n$$

sampling/precision budget: $\mathbb{E}[N] \leq n \;\Rightarrow\; D_0 \;\geq\; \log s/\epsilon$

minimum requirement for any testing scheme
with expected sample/precision budget m

essentially achievable using sequential probability ratio test for each component
assuming sparsity level is known

Lower Bound

expected total sampling/precision:

$$\mathbb{E}[N] \quad = \quad (n-s)\mathbb{E}_0[M] + s\mathbb{E}_1[M] \;\gtrsim\; \frac{n}{D_0} \log \frac{s}{\epsilon} \;, \;\; \text{when } s \ll n$$

sampling/precision budget: $\mathbb{E}[N] \le n \;\Rightarrow\; D_0 \;\ge\; \log s/\epsilon$

minimum requirement for any testing scheme
with expected sample/precision budget m

essentially achievable using sequential probability ratio test for each component
assuming sparsity level is known

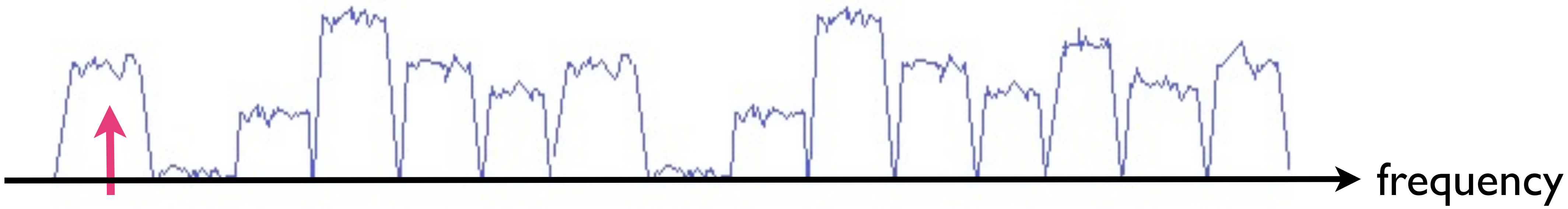
Gaussian case: $N(0,1)$ vs. $N(\mu,1) \;\Rightarrow\; D_0 = \mu^2/2$ and so prob(error) $\le \epsilon$ iff

$$\mu \;>\; \sqrt{2 \log s/\epsilon}$$

sequential thresholding: $\mu \;>\; \sqrt{2\,(\log s + \log\log_2 \log n)}$

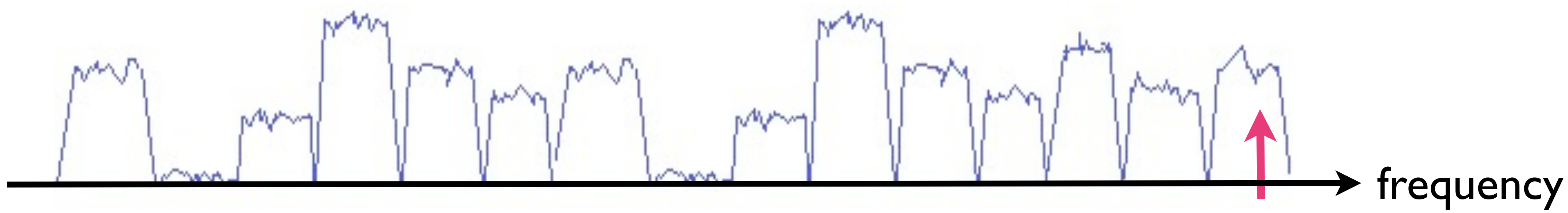
and is adaptive to sparsity level

Spectrum Sensing



goal: scan to find open channel(s) as quickly as possible

Spectrum Sensing



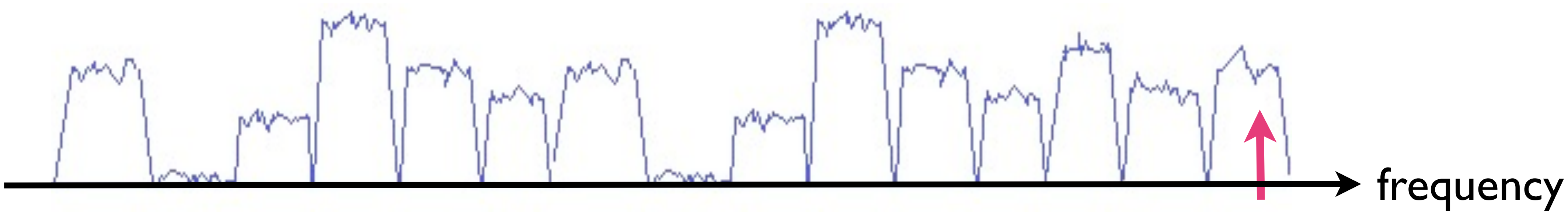
goal: scan to find open channel(s) as quickly as possible

Spectrum Sensing

goal: scan to find open channel(s) as quickly as possible

channel samples: $y_{i,j} \overset{\text{iid}}{\sim} \mathcal{CN}(0, \theta)\ , \quad \theta_0 > \theta_1 = 1$

Spectrum Sensing

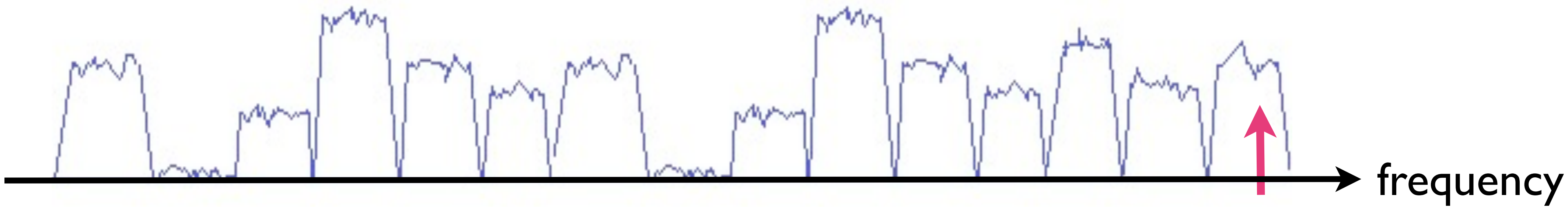


goal: scan to find open channel(s) as quickly as possible

channel samples: $y_{i,j} \overset{\text{iid}}{\sim} \mathcal{CN}(0,\theta)\ , \quad \theta_0 > \theta_1 = 1$

test statistic: $t_i(m_i) \ = \ \sum_{j=1}^{m_i} |y_{i,j}|^2 \ \sim \begin{cases} \Gamma(m_i,\theta_0)\,, & i \notin \mathcal{S} \\ \Gamma(m_i,1)\,, & i \in \mathcal{S} \end{cases}$

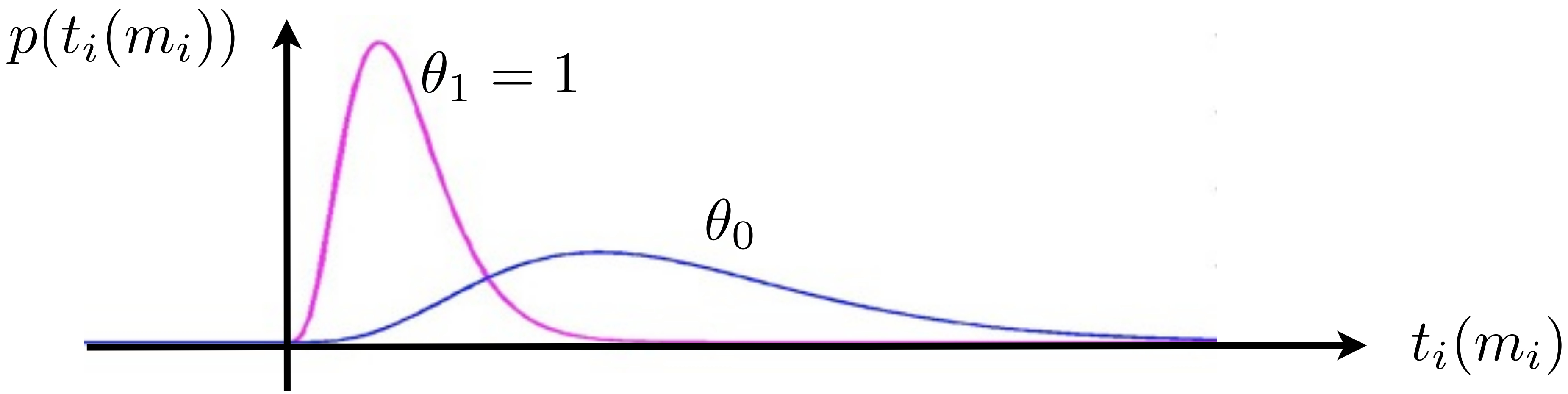
Spectrum Sensing



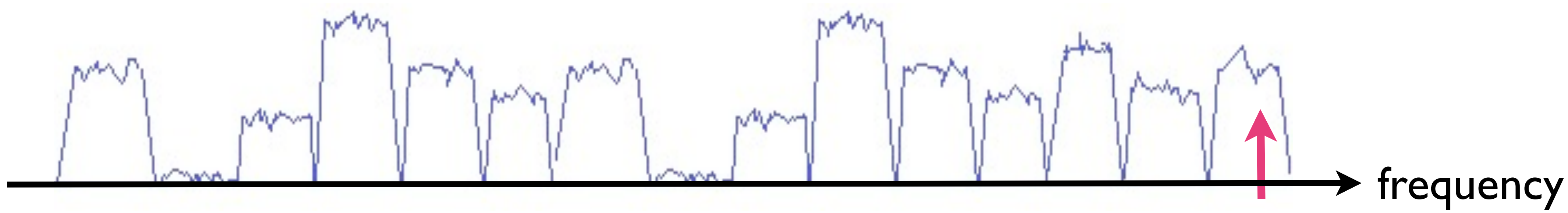
goal: scan to find open channel(s) as quickly as possible

channel samples: $y_{i,j} \overset{\text{iid}}{\sim} \mathcal{CN}(0, \theta)\ , \quad \theta_0 > \theta_1 = 1$

test statistic: $t_i(m_i) \ = \ \sum_{j=1}^{m_i} |y_{i,j}|^2 \ \sim \begin{cases} \Gamma(m_i, \theta_0)\,, & i \notin \mathcal{S} \\ \Gamma(m_i, 1)\,, & i \in \mathcal{S} \end{cases}$



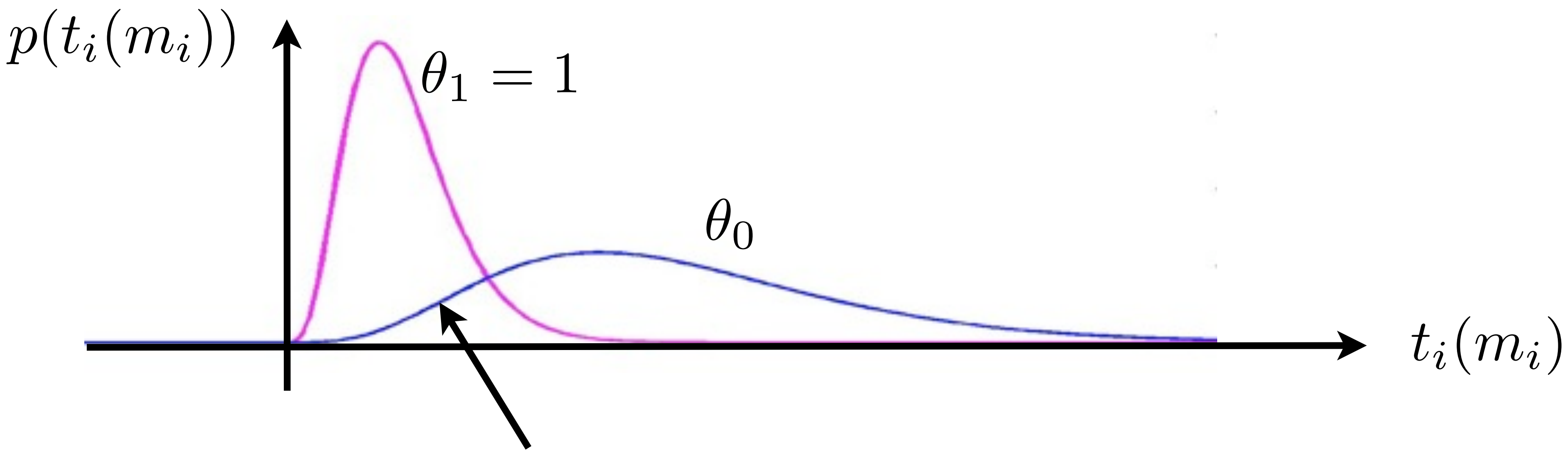
Spectrum Sensing



goal: scan to find open channel(s) as quickly as possible

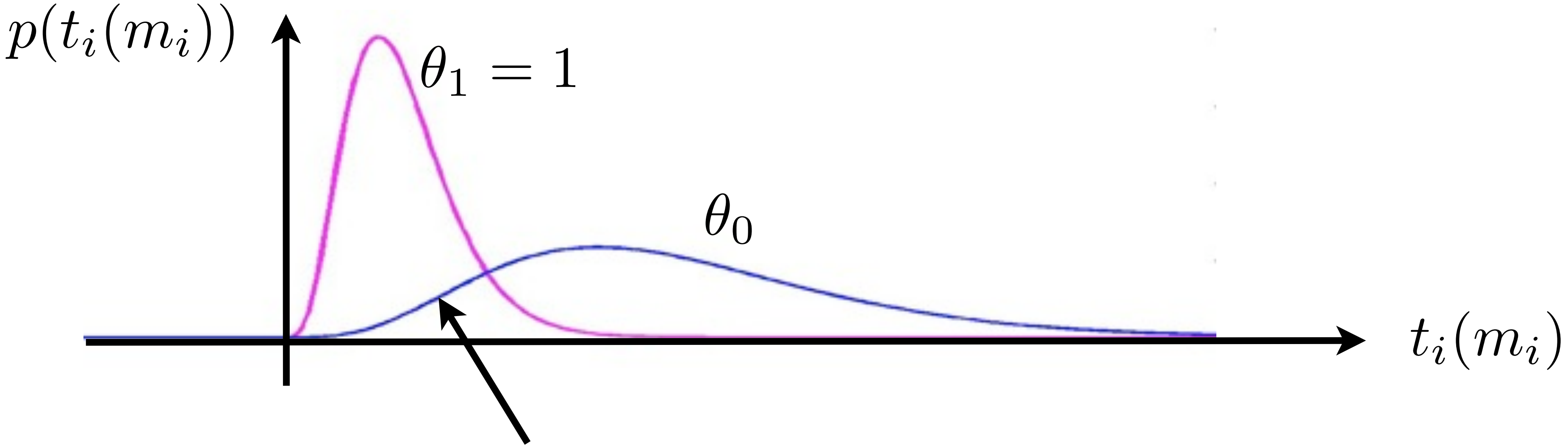
channel samples: $y_{i,j} \overset{\text{iid}}{\sim} \mathcal{CN}(0,\theta) \;, \quad \theta_0 > \theta_1 = 1$

test statistic: $t_i(m_i) \;=\; \sum_{j=1}^{m_i} |y_{i,j}|^2 \;\sim\; \begin{cases} \Gamma(m_i,\theta_0)\,, & i \notin \mathcal{S} \\ \Gamma(m_i,1)\,, & i \in \mathcal{S} \end{cases}$



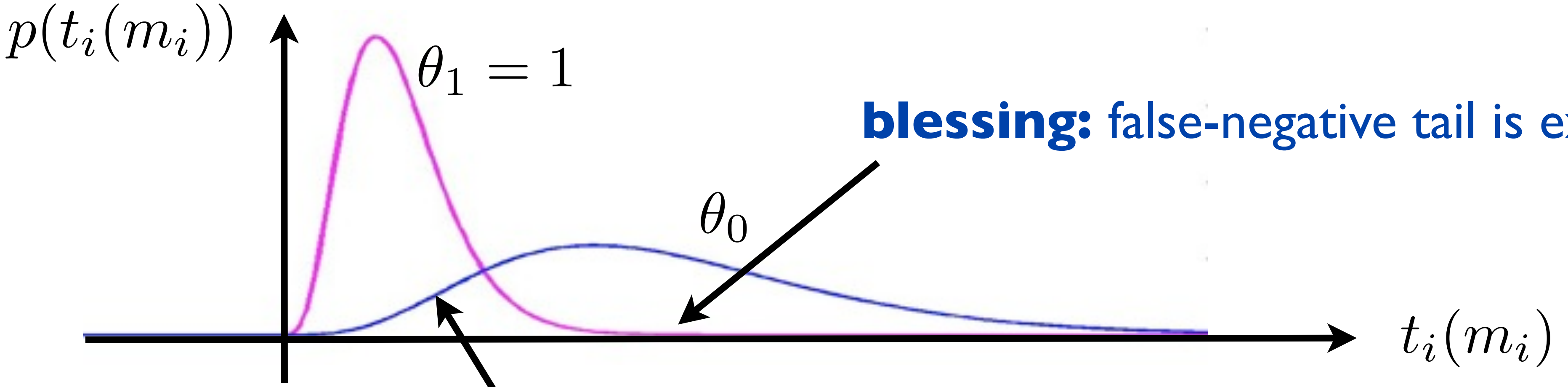
curse: false-positive error tail is polynomial

Spectrum Sensing Application



curse: false-positive error tail is polynomial

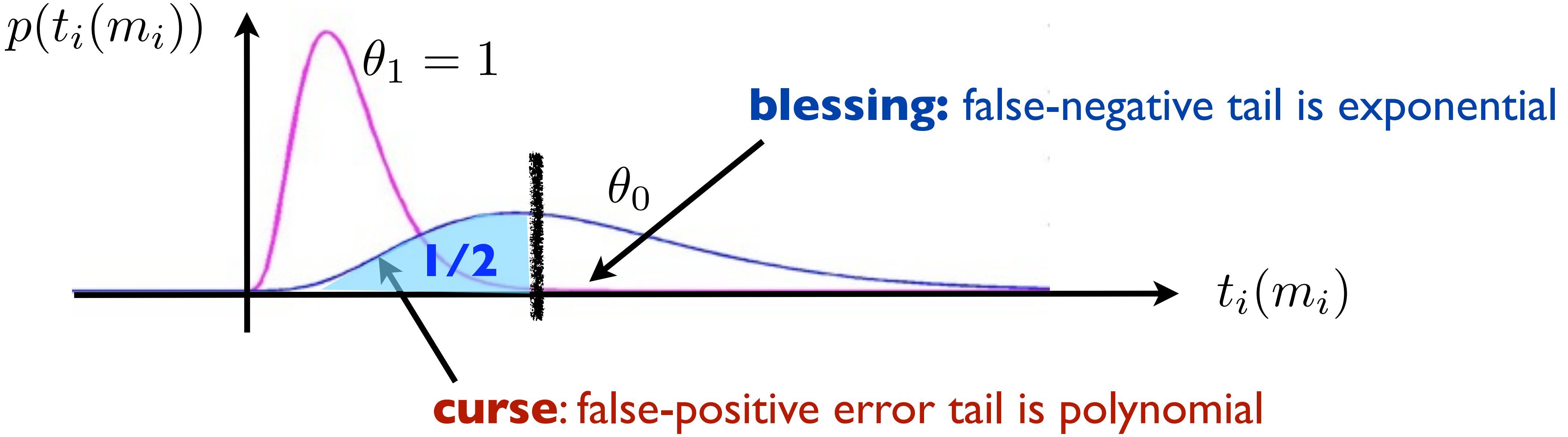
Spectrum Sensing Application



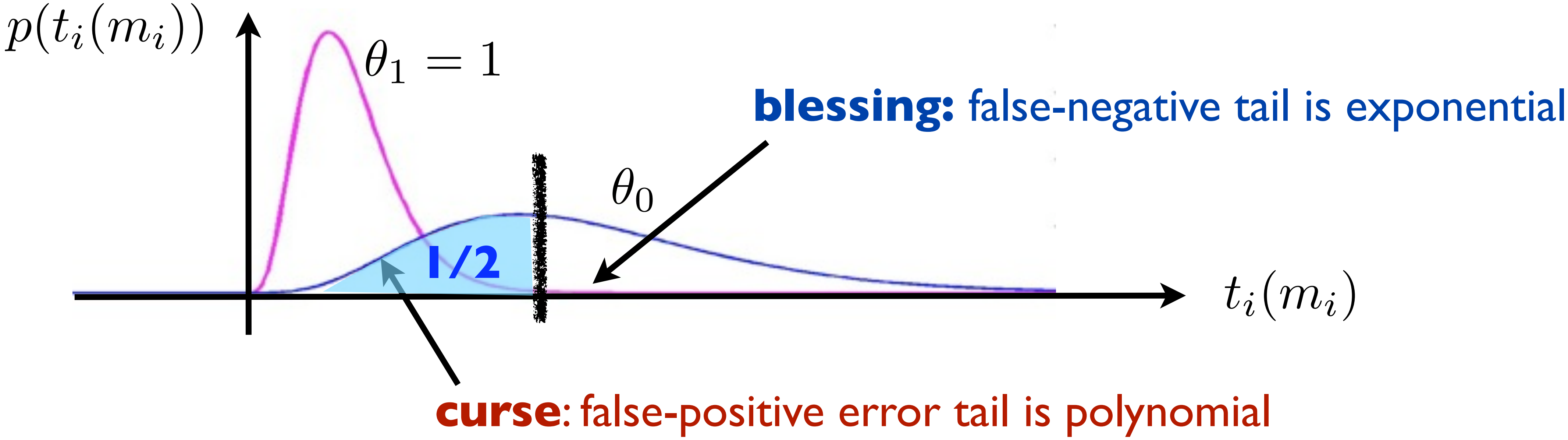
blessing: false-negative tail is exponential

curse: false-positive error tail is polynomial

Spectrum Sensing Application

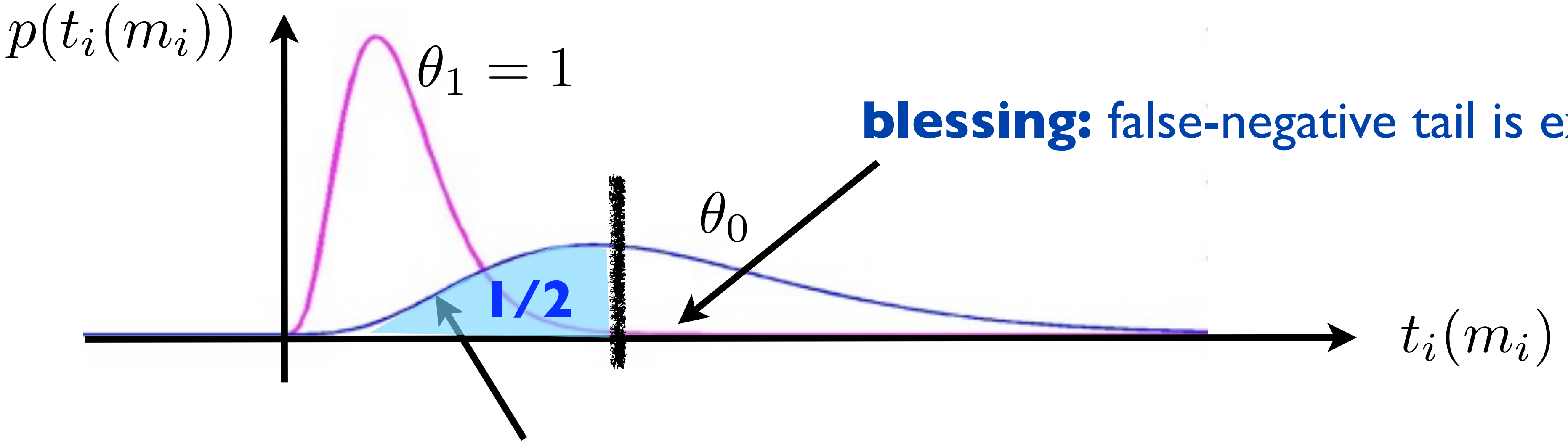


Spectrum Sensing Application



scanning budget: $m =$ average number of samples per channel

Spectrum Sensing Application



scanning budget: $m =$ average number of samples per channel

non-sequential: $\theta_0 > 2(m-1)(n-s)^{1/2m} \sim n^{1/2m}$ (necessary)

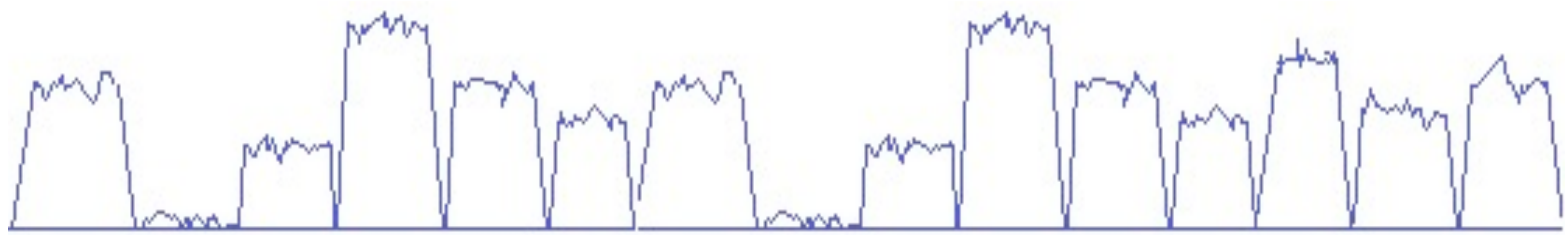
SPRT: $\theta_0 > \frac{1}{m} \log s$ minimum requirement for any testing scheme with average sample budget m

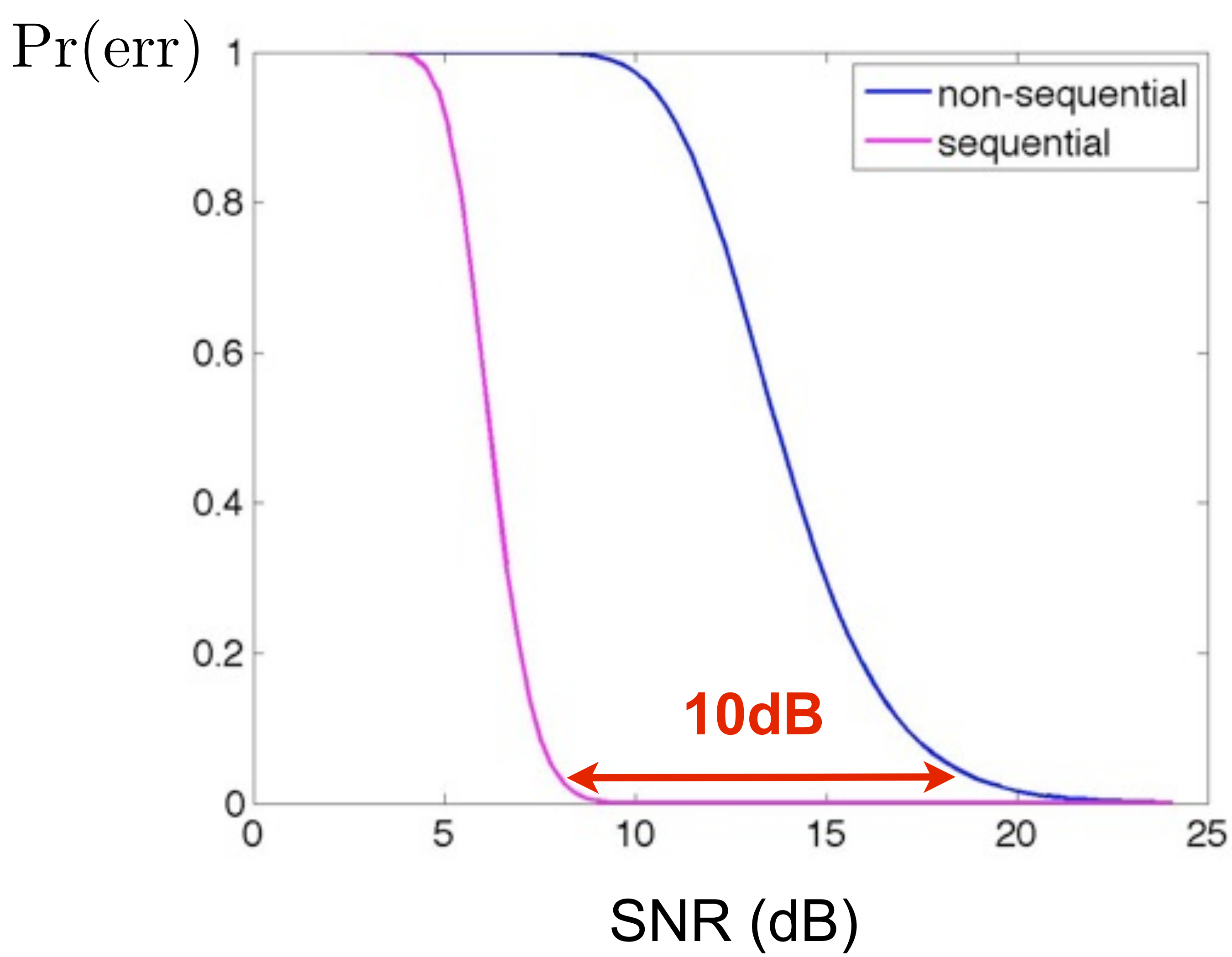
sequential thresholding: $\theta_0 > \frac{1}{m}(\log s + \log \log_2 n)$ (sufficient)

... and automatically adaptive to sparsity level

Performance

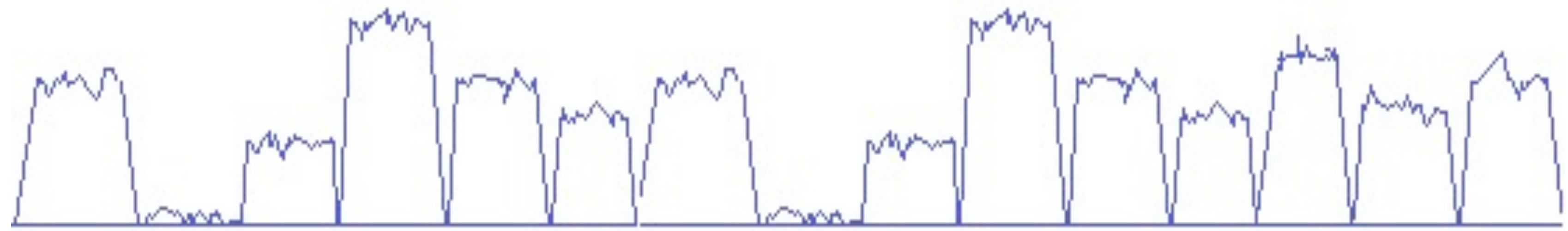
n=1000 channels, s=6 unoccupied

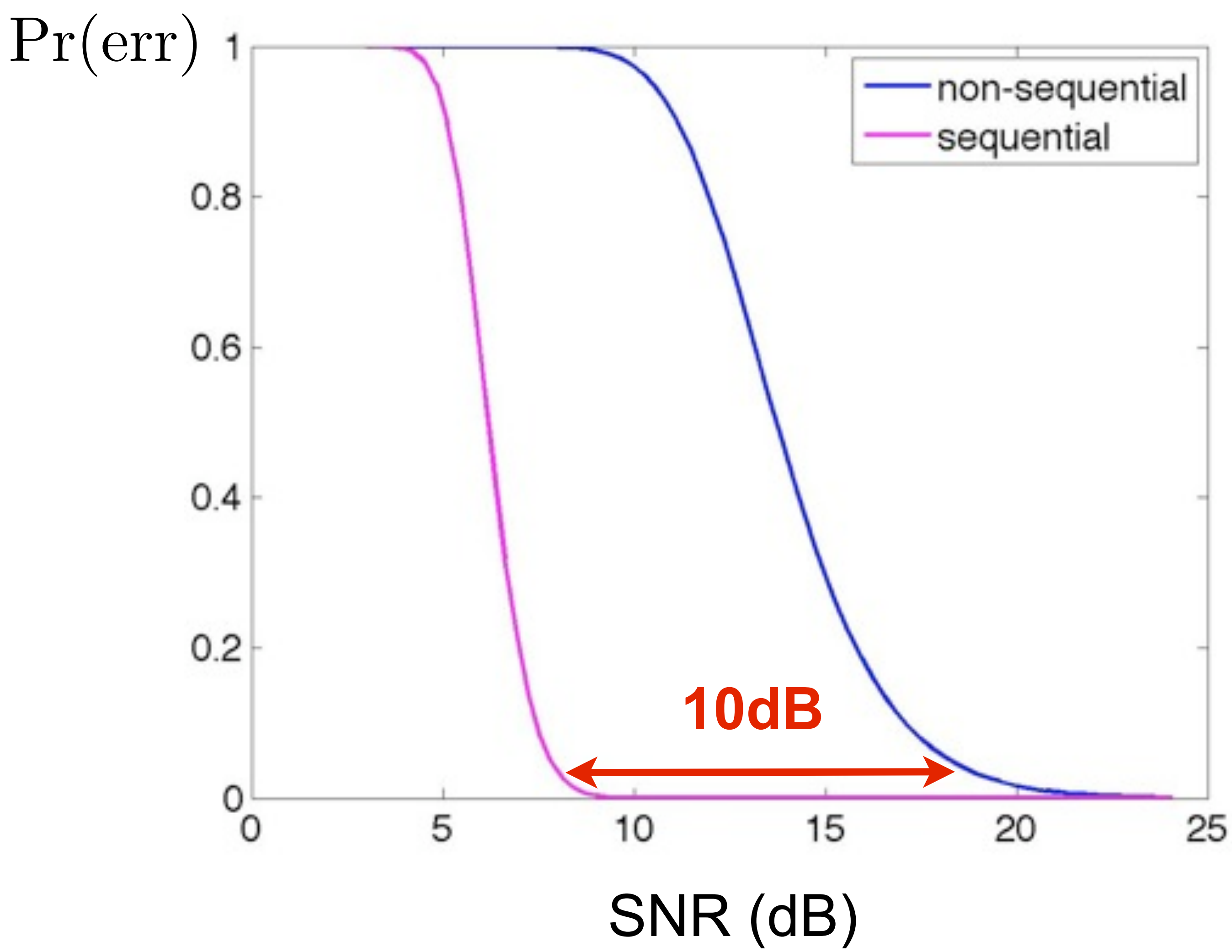


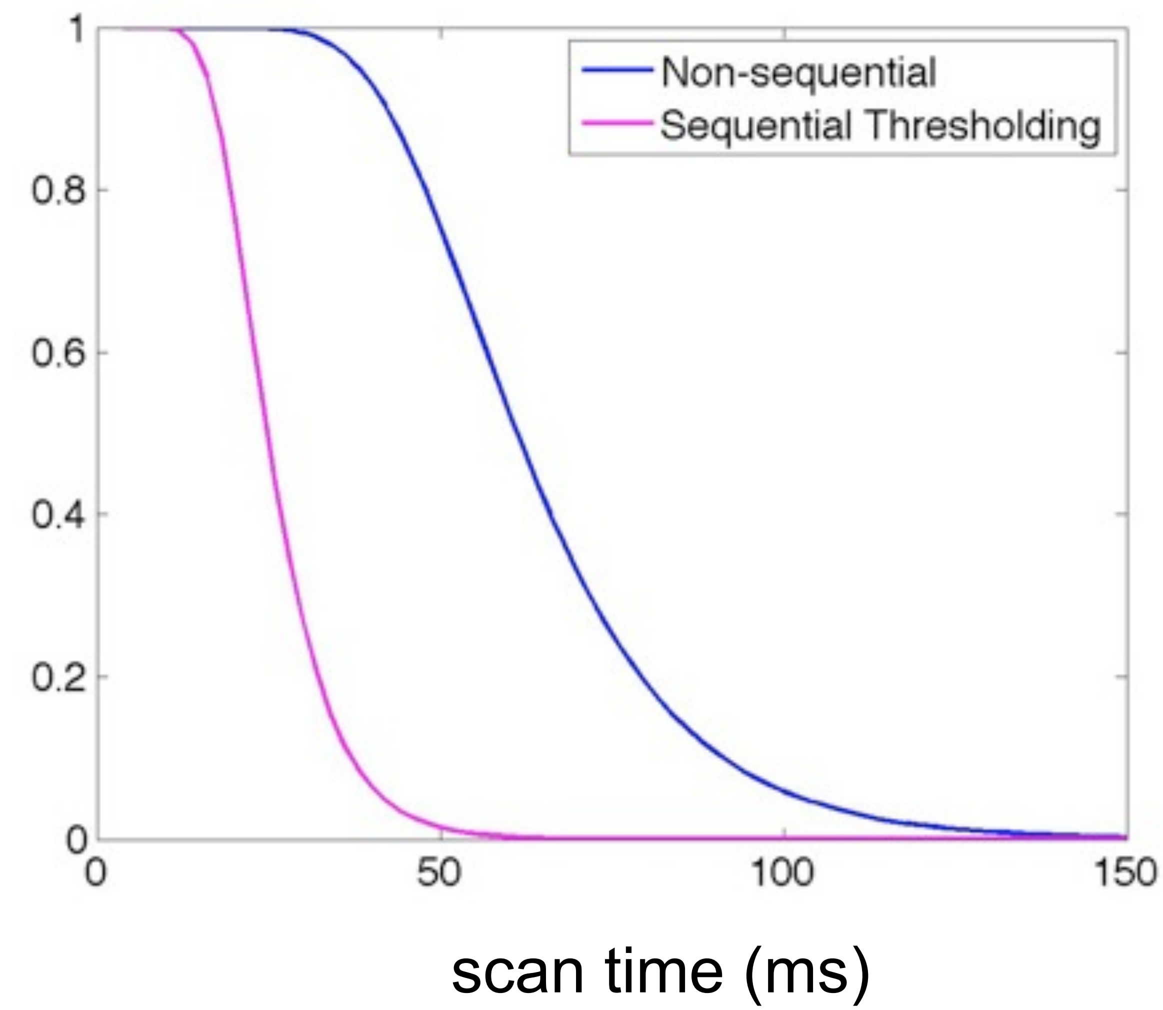


Performance

n=1000 channels, s=6 unoccupied

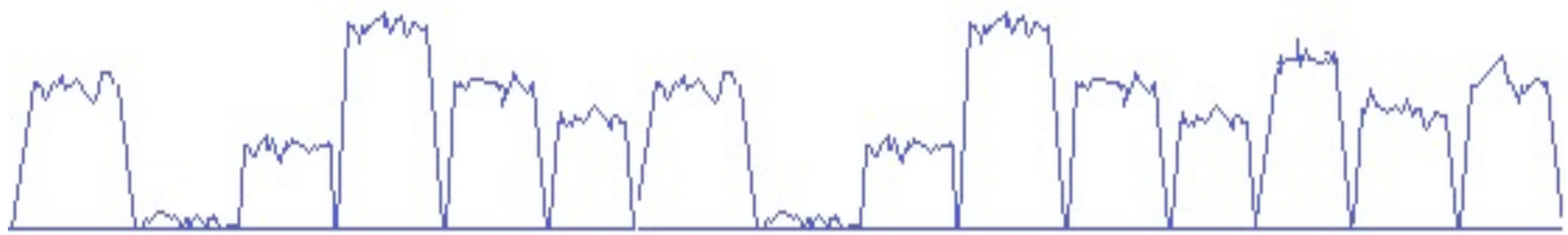


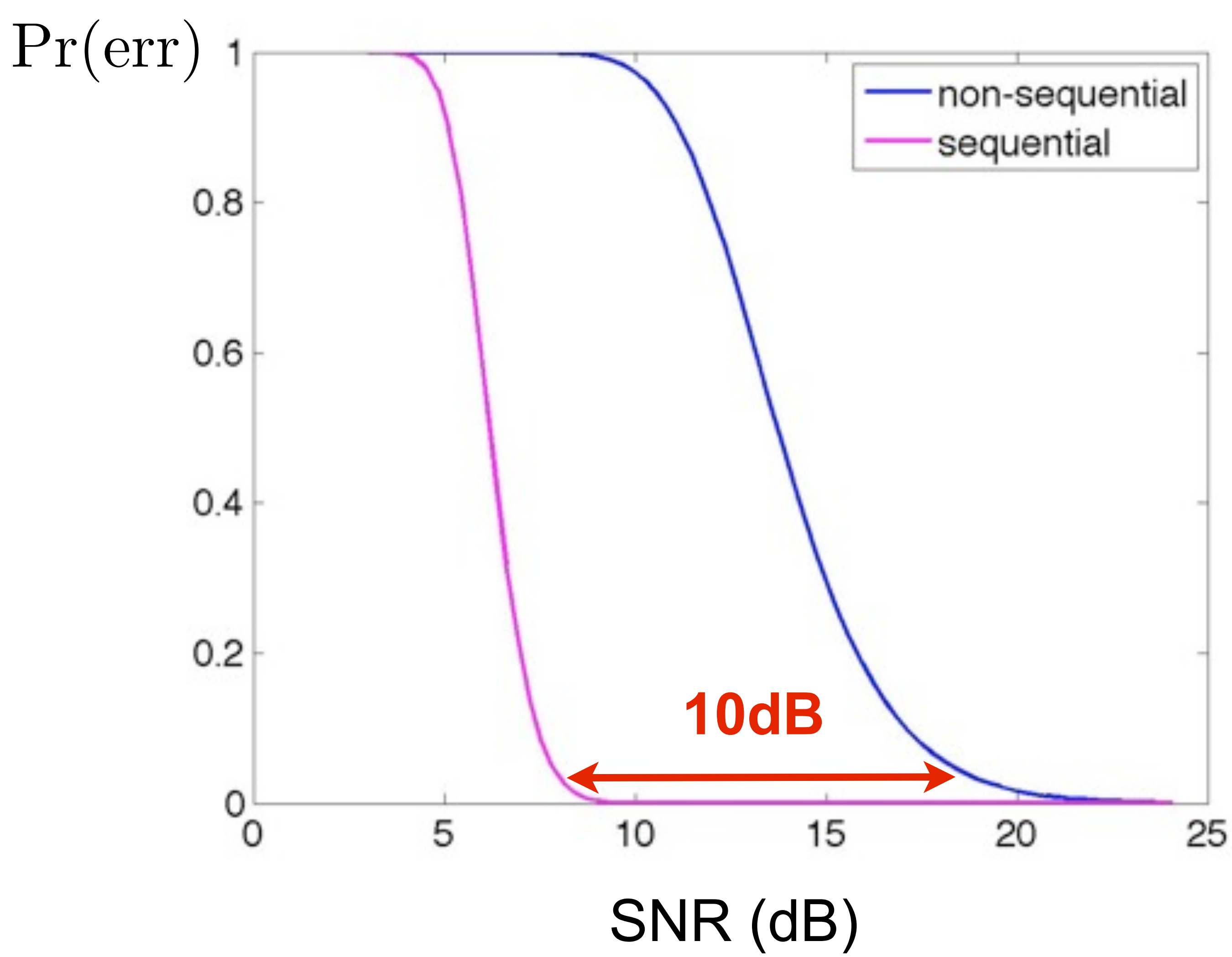


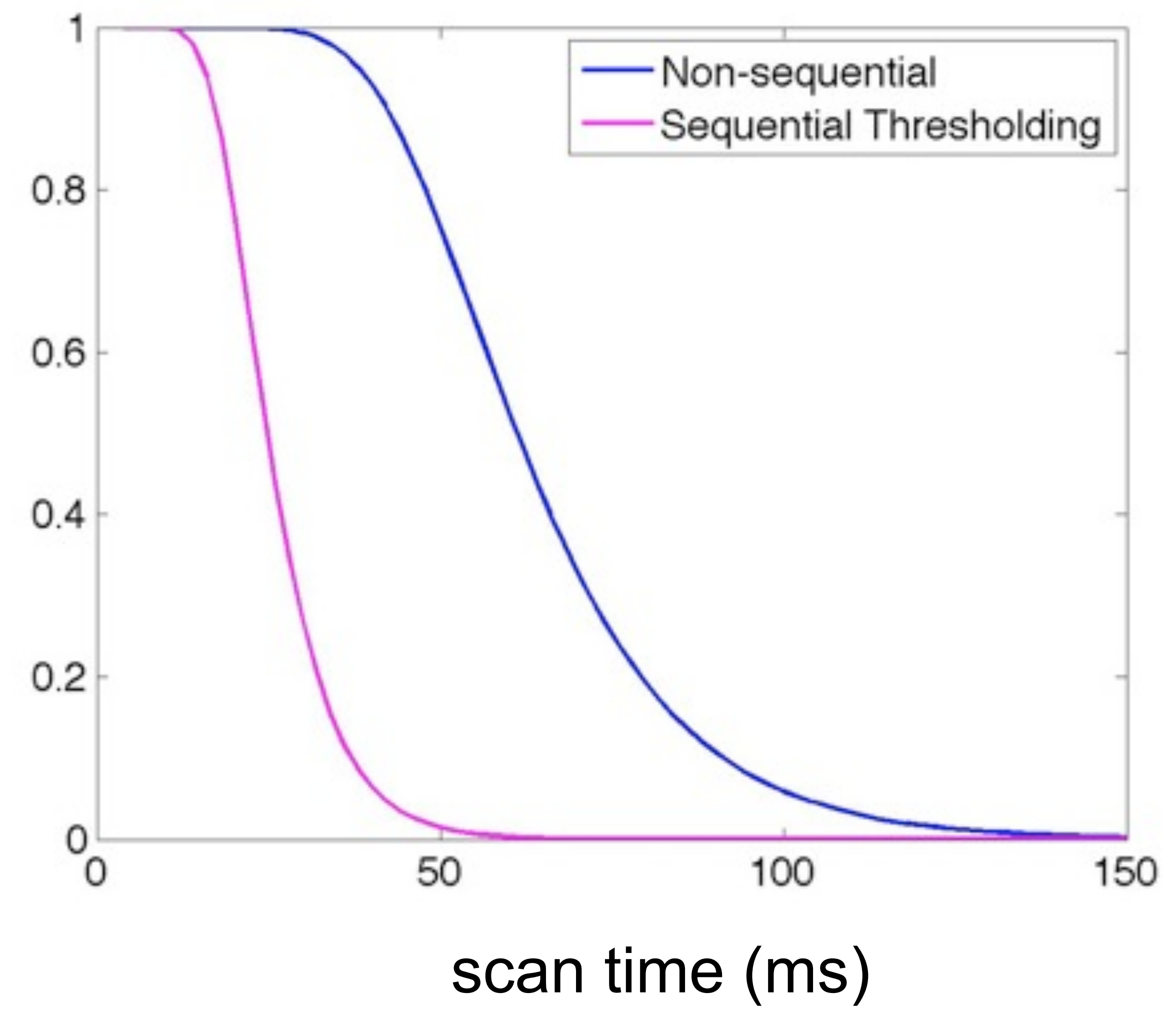


Performance

n=1000 channels, s=6 unoccupied

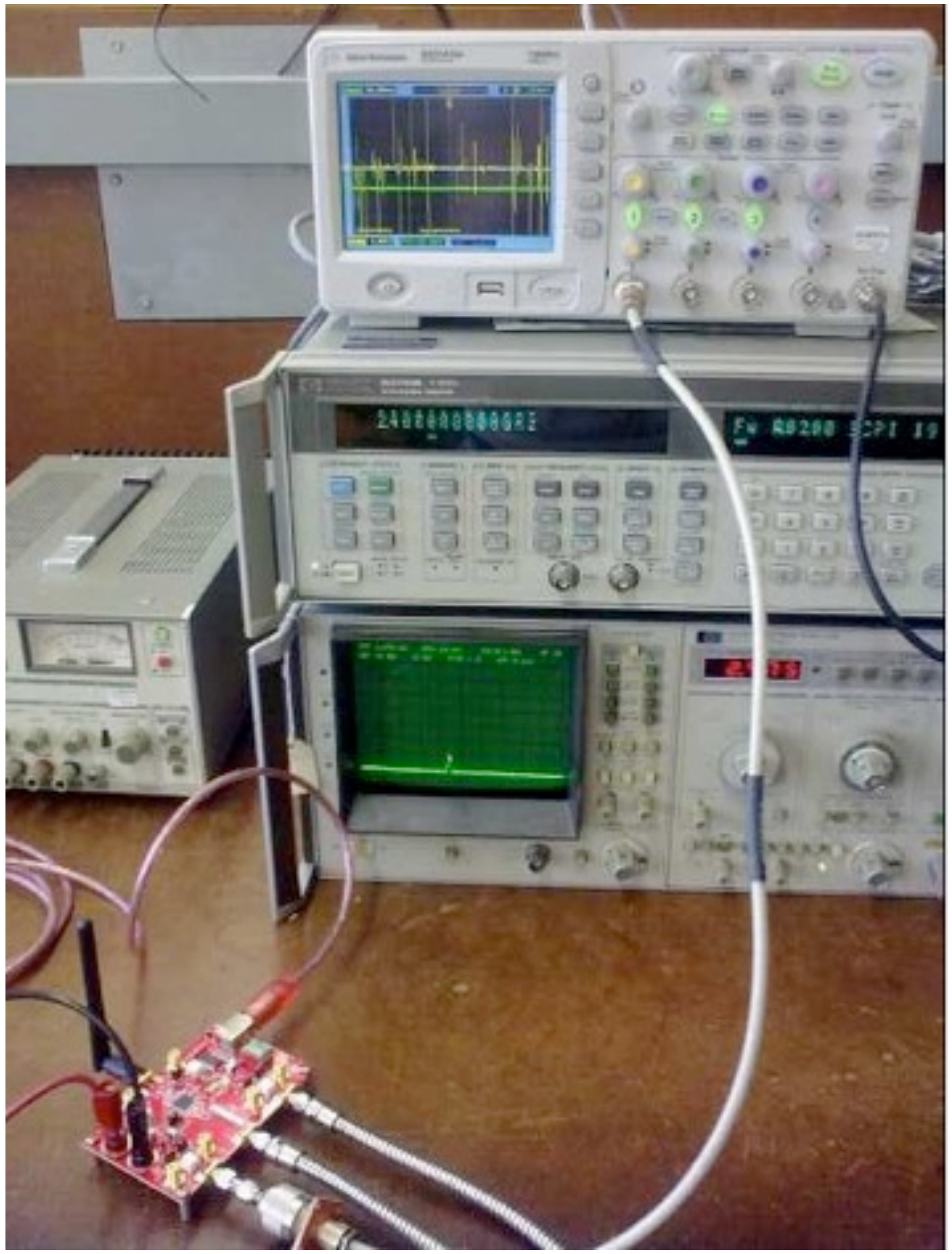






sequential thresholding is about 10 times more sensitive
(for equal scan time) or scans 3 times faster (with same reliability)

Spectrum Sensing in the Lab



Matt Malloy in the lab

Faster, Better, Stronger

requirements for reliable sequential testing in high-dimensional sparse problems:

1. SNR ~ $\max(\log(s), \log\log\log(n))$
2. total number of samples ~ $2n$

Faster, Better, Stronger

requirements for reliable sequential testing in high-dimensional sparse problems:

1. SNR ~ max(log(s) , loglog log(n))
2. total number of samples ~ 2n

Can we break the log(s) barrier ?

Faster, Better, Stronger

requirements for reliable sequential testing in high-dimensional sparse problems:

1. SNR ~ max(log(s) , loglog log(n))
2. total number of samples ~ 2n

Can we break the log(s) barrier ?

Yes, if we relax the probability of error requirement and only require that most of the signal components are detected.
SNR can be any slowly diverging function of n.

Faster, Better, Stronger

requirements for reliable sequential testing in high-dimensional sparse problems:

1. SNR ~ max(log(s) , loglog log(n))
2. total number of samples ~ 2n

Can we break the log(s) barrier ?

Yes, if we relax the probability of error requirement and only require that most of the signal components are detected. SNR can be any slowly diverging function of n.

Is reliable detection possible using fewer than 2n samples ?

Faster, Better, Stronger

requirements for reliable sequential testing in high-dimensional sparse problems:

1. SNR ~ max(log(s) , loglog log(n))
2. total number of samples ~ 2n

Can we break the log(s) barrier ?

Yes, if we relax the probability of error requirement and only require that most of the signal components are detected. SNR can be any slowly diverging function of n.

Is reliable detection possible using fewer than 2n samples ?

Yes. Sequential versions of compressed sensing require only s log(n) samples.

Faster, Better, Stronger

requirements for reliable sequential testing in high-dimensional sparse problems:

1. SNR ~ max(log(s) , logloglog(n))
2. total number of samples ~ 2n

Can we break the log(s) barrier ?

Yes, if we relax the probability of error requirement and only require that most of the signal components are detected.
SNR can be any slowly diverging function of n.

Is reliable detection possible using fewer than 2n samples ?

Yes. Sequential versions of compressed sensing require only s log(n) samples.

see Jarvis Haupt's poster

Conclusions

Sequential Experimental Designs for High-Dimensional Models

thresholds for recovery in high-dimensional limit:

non-sequential designs	SNR $\sim$ log(dimension)	(or worse)
sequential designs	SNR $\sim$ log(sparsity level)	(or better)

Sequential Analysis High-Dimensional Multiple Testing and Sparse Recovery
M. Malloy and RN, **arXiv:1103.5991**

Distilled Sensing: Adaptive Sampling for Sparse Detection and Estimation
J. Haupt, R. Castro, and RN, **arXiv:1001.5311v2**