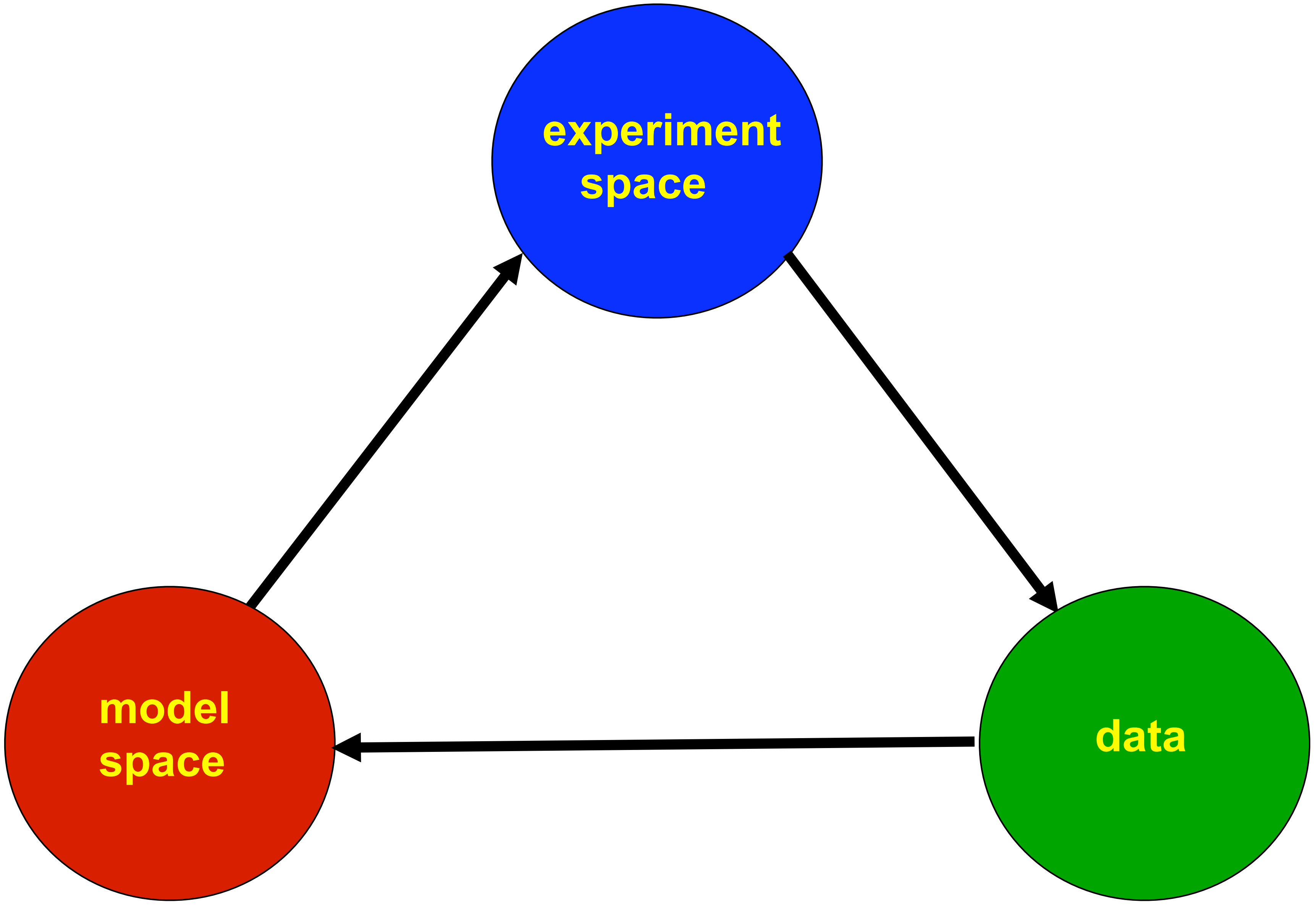
Adaptive Sensing and Sparse Interactions



Rob Nowak
www.ece.wisc.edu/~nowak

BIRS Workshop on Sparse Statistics
Optimization and Machine Learning
January 16-21 2011

Joint work with R. Castro, J. Haupt,
M. Malloy and B. Nazer

Motivation: Inferring Biological Pathways

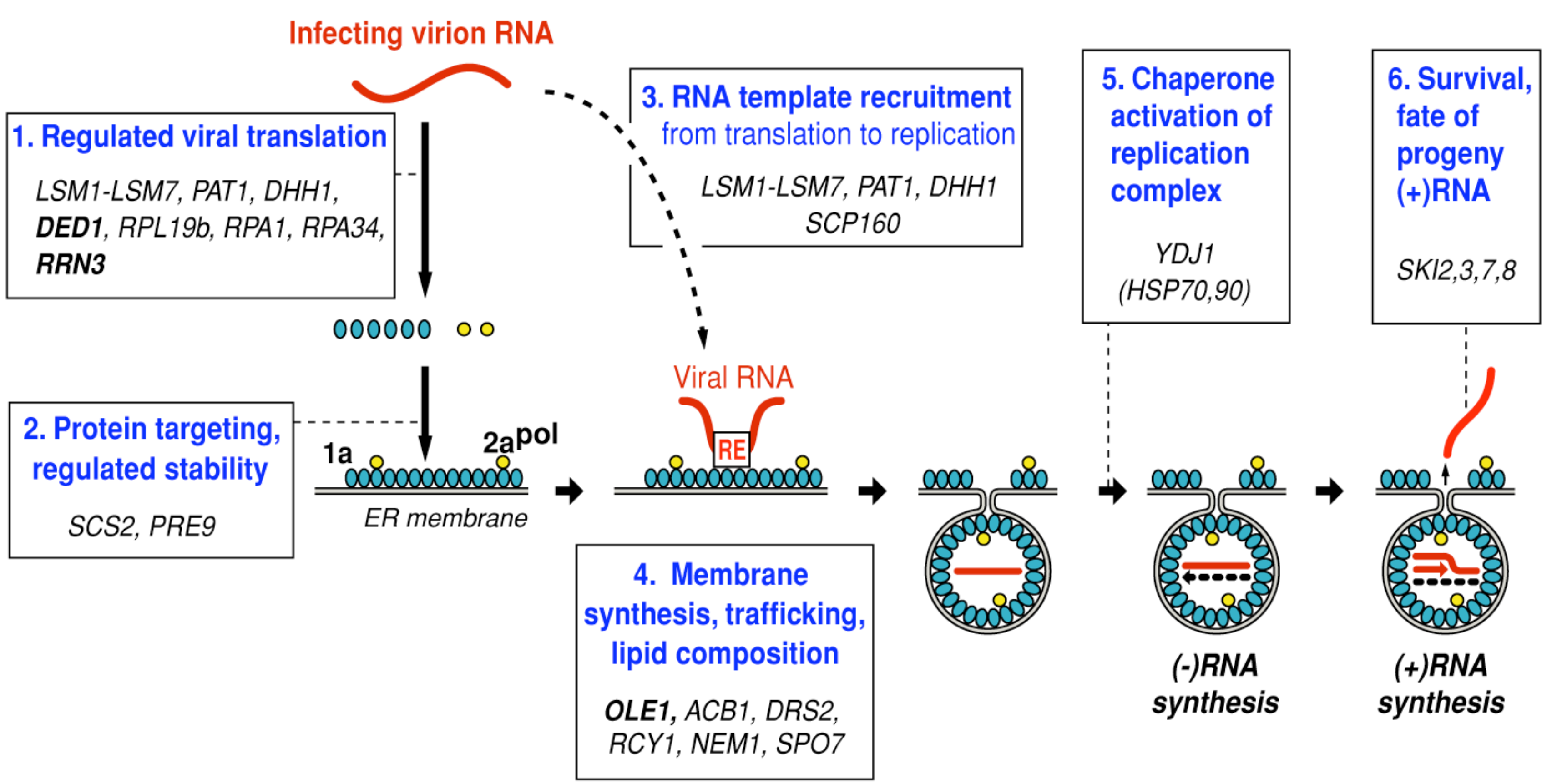
microwell array

virus

fruit fly

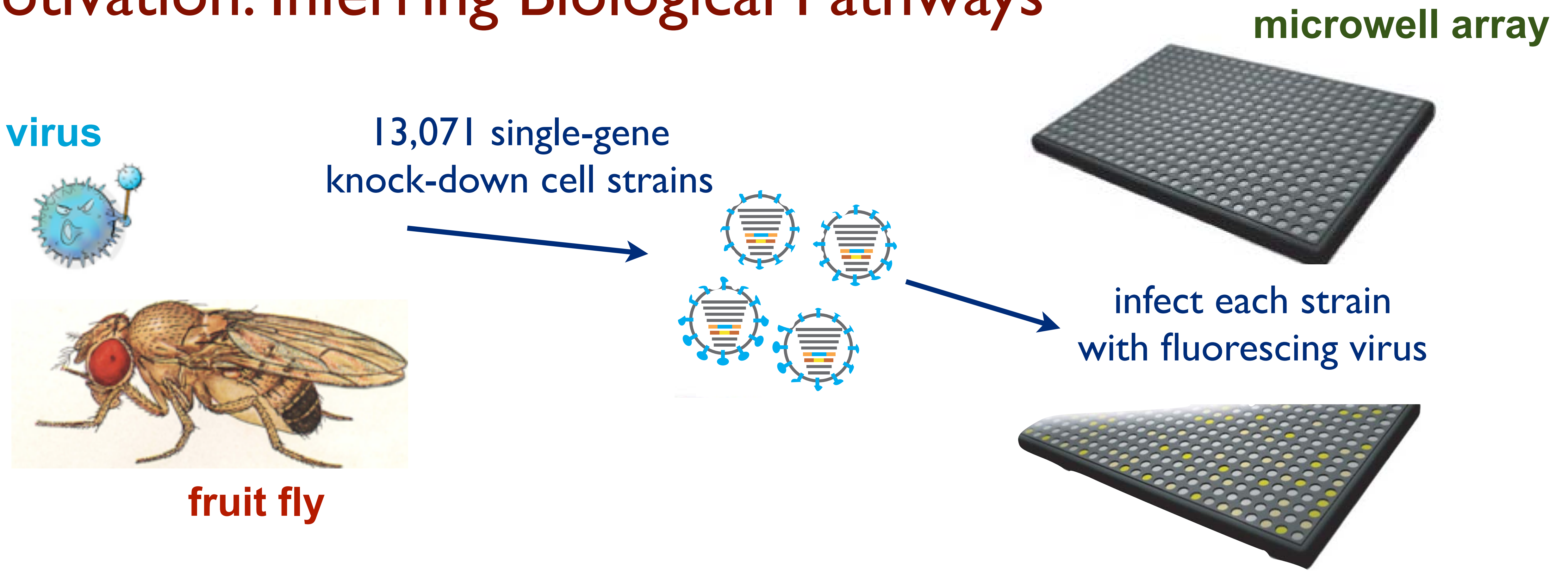
13,071 single-gene
knock-down cell strains

infect each strain
with fluorescing virus



approximately 100 significant genes/protiens discovered by Ahlquist and Kawaoka Labs at UW-Madison

Motivation: Inferring Biological Pathways



Challenges:

1. very low SNR data and huge number of experiments and tests
2. non-linear interactions

Challenge 1: High-Dimensionality and Low SNR

nature Vol 454 | 14 August 2008 | doi:10.1038/nature07151

Drosophila RNAi screen identifies host genes important for influenza virus replication

Linhui Hao[1,2]*, Akira Sakurai[3]*†, Tokiko Watanabe[3], Ericka Sorensen[1], Chairul A. Nidom[5,6], Michael A. Newton[4], Paul Ahlquist[1,2] & Yoshihiro Kawaoka[3,7,8,9]

How do they confidently determine the ~100 out of 13K genes hijacked for virus replication from extremely noisy data?

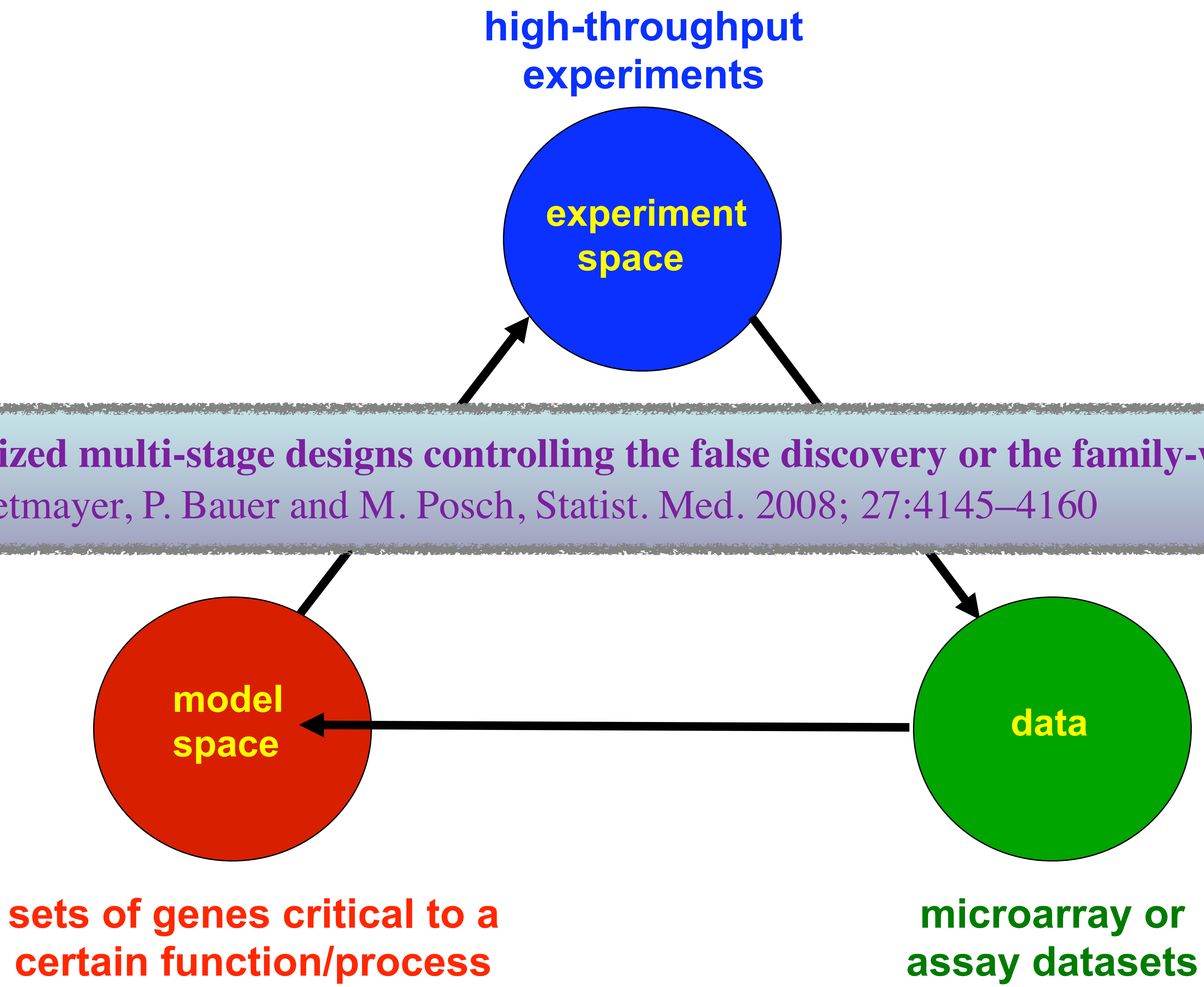
Sequential Experimental Design:

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

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

vastly more efficient that replicating all 13K experiments many times

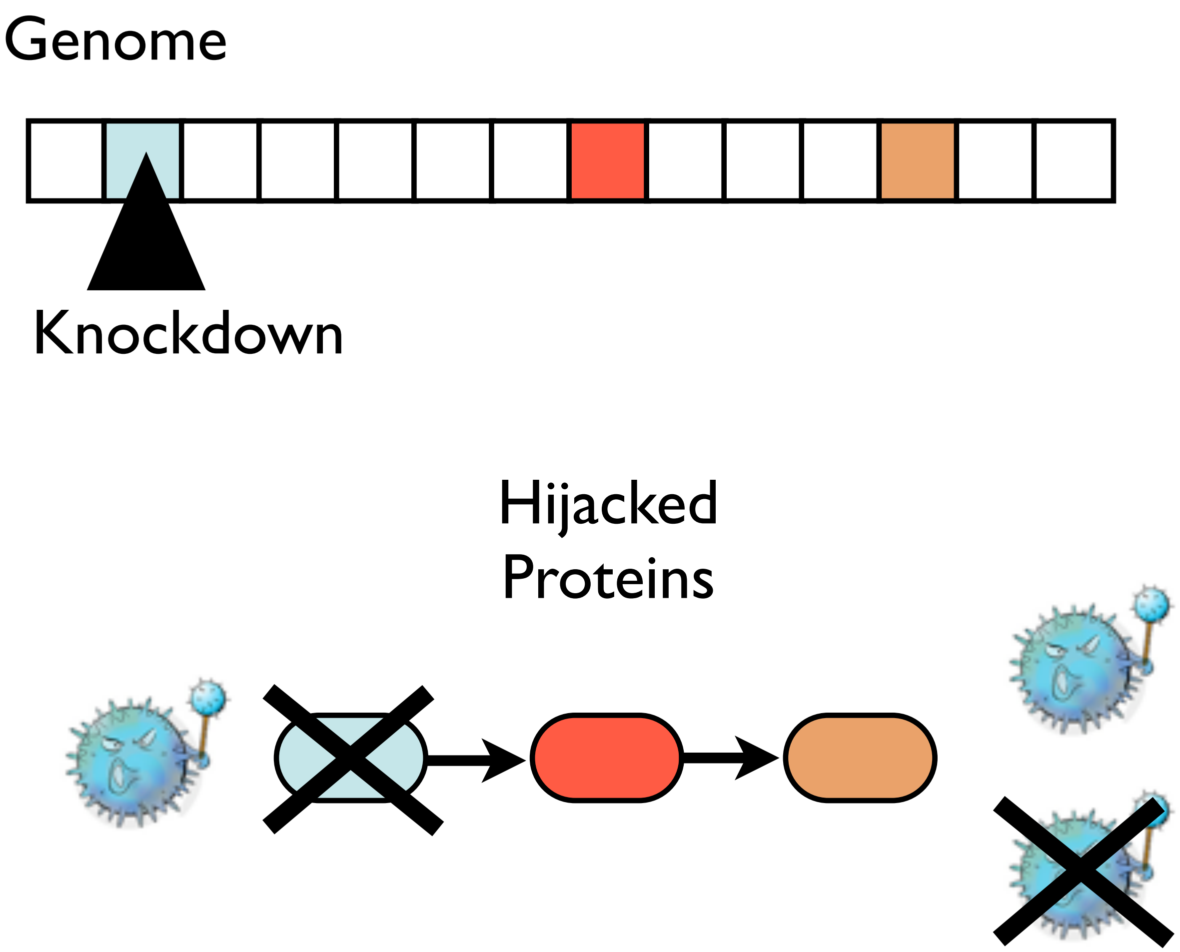
Feedback from Data Analysis to Data Collection



Optimized multi-stage designs controlling the false discovery or the family-wise error rate
S. Zehetmayer, P. Bauer and M. Posch, Statist. Med. 2008; 27:4145–4160

Challenge 2: Sparsity & Nonlinear Effects

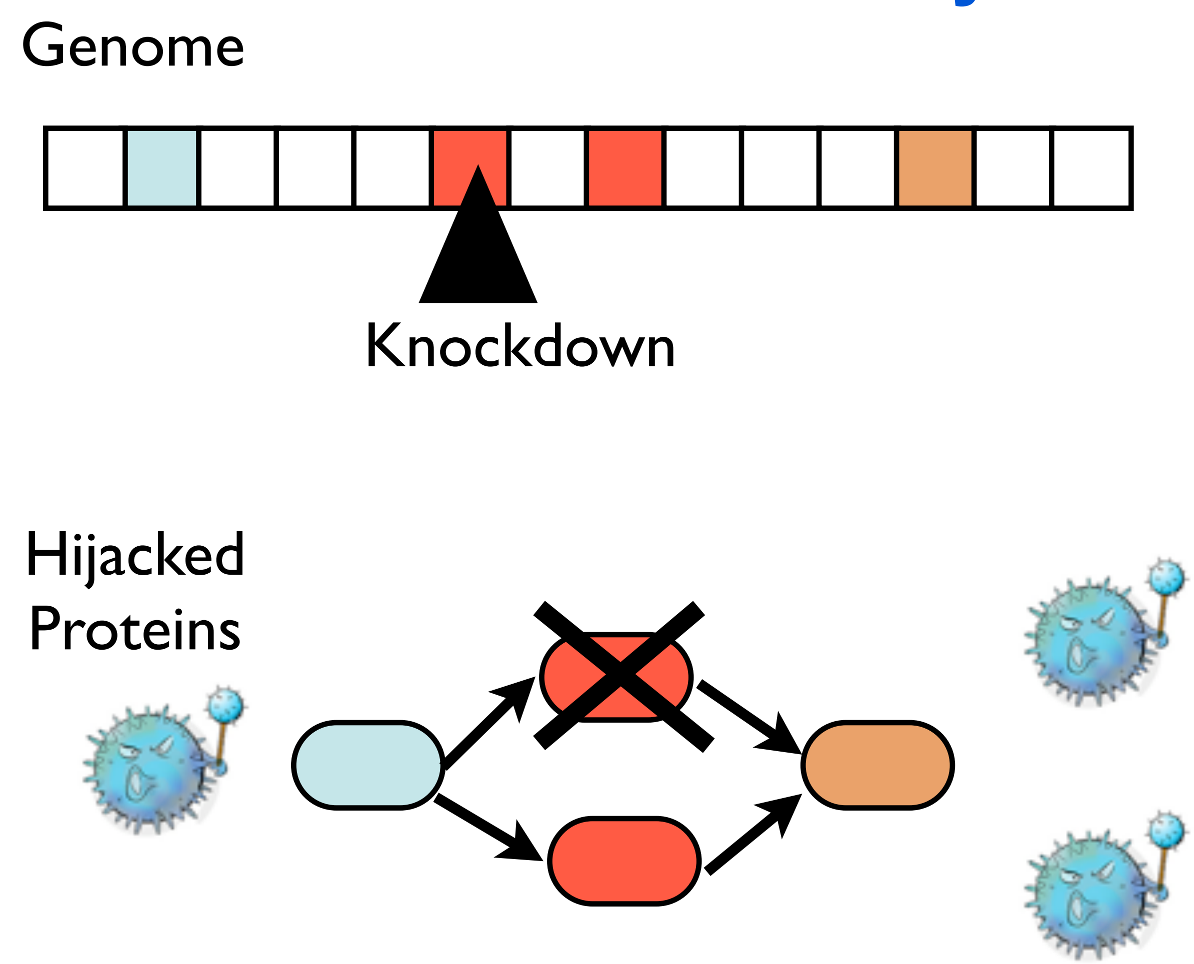
Linear Pathway



detectable effect

Linear Sparsity

Redundant Pathways



little or no observable effect !

Multilinear Sparsity

Outline of Talk

1. Sequential Experimental Designs for High-Dimensional Testing

thresholds for recovery in high-dimensional limit:

$$\begin{aligned} \textbf{non-adaptive designs} \qquad \text{SNR} &\sim \log n \\ \textbf{sequential designs} \qquad \text{SNR} &\sim \text{arbitrarily slowly growing function of } n \end{aligned}$$

2. Compressed Sensing of Sparse Multilinear Functions

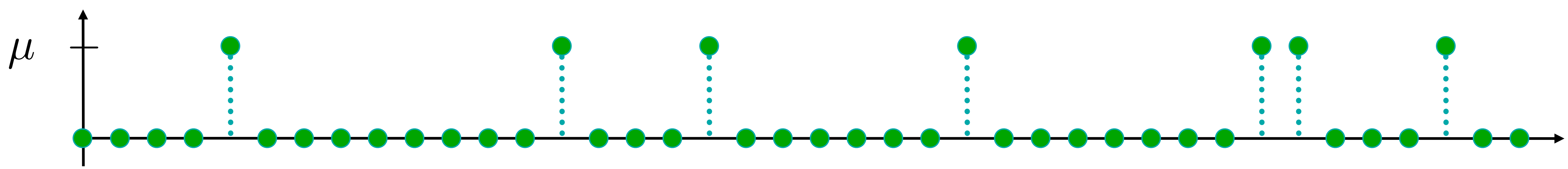
number of compressed sensing measurements for sparse recovery:

$$\begin{aligned} \textbf{linear sparsity} \qquad K &\sim S \log n \\ \textbf{multilinear sparsity} \qquad K &\sim \min\{S^2 \log n, S \log^3(S) \log n, S^\alpha \log^\alpha n\} \end{aligned}$$

where $\alpha \geq 1$ *depends* on pattern of sparsity

Sparse Signal Model

Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ be an unknown sparse vector; most (or all) of its components x_i are equal to zero.



$$x_i = \begin{cases} \mu > 0\,, & i \in \mathcal{S} \\ 0\,, & i \notin \mathcal{S} \end{cases}\,, \text{ where } |\mathcal{S}| \ll n$$

signal **support set**

deterministic
but unknown

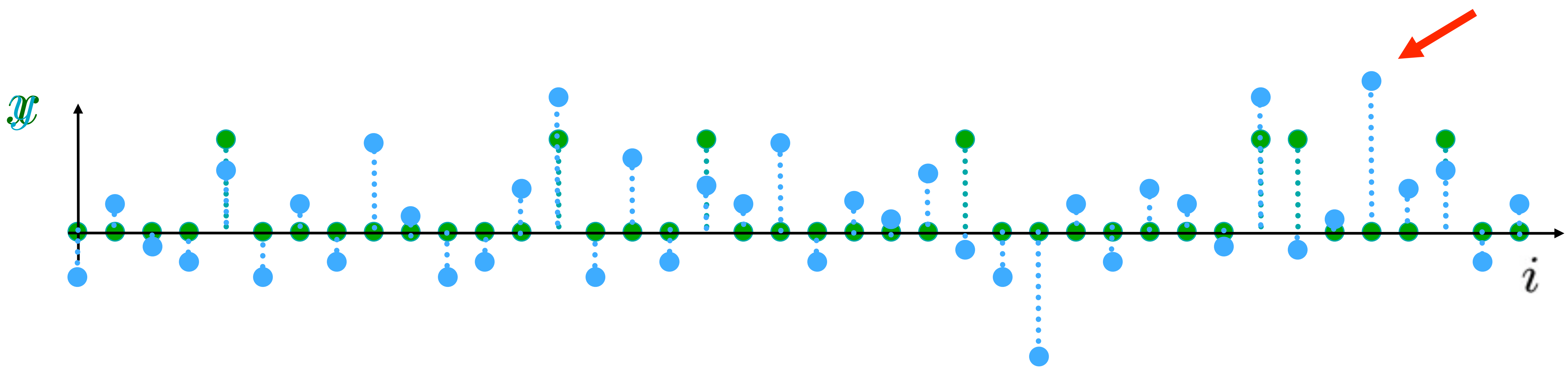
Assume sublinear sparsity level: $|\mathcal{S}| \ll n$

number of signal
components

Noisy Observation Model

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

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



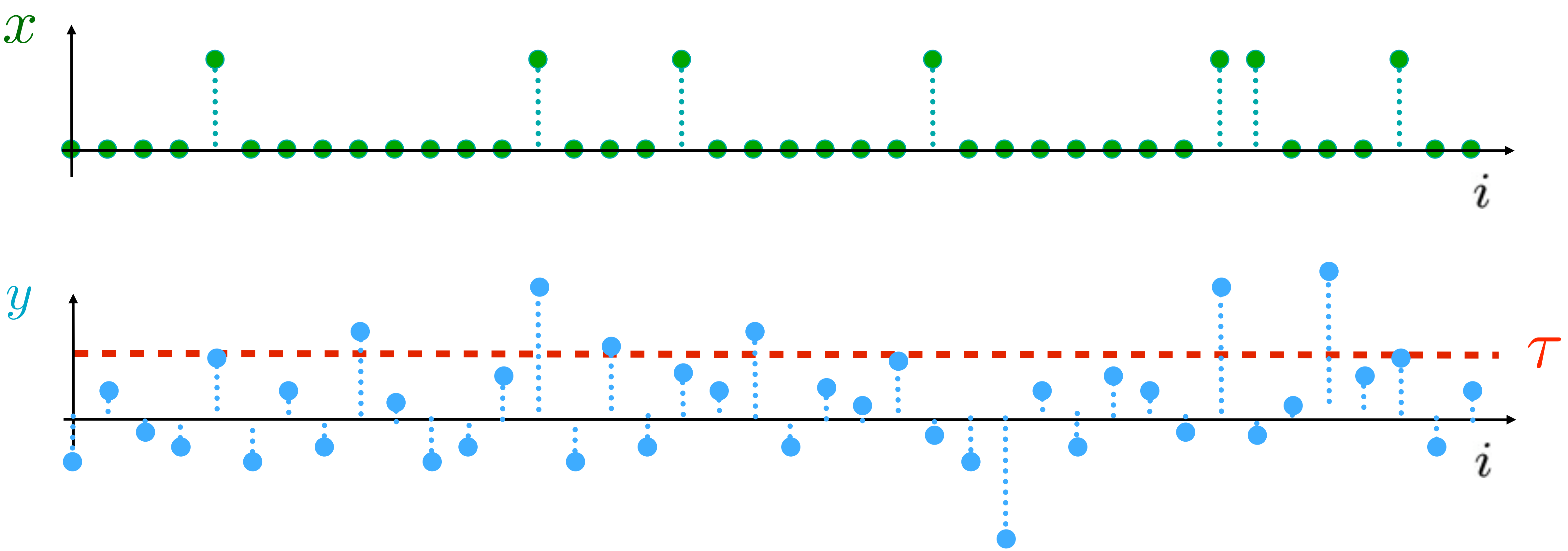
Suppose we want to locate just **one** signal component: $\widehat{i} \;=\; \arg\max_i y_i$

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

It is *impossible* to reliably detect signal components weaker than $O(\sqrt{\log n})$

Threshold Tests

Our goal is to estimate the set of non-zero components: $\mathcal{S} := \{i \,:\, x_i \neq 0\}$



Definition 1 *A* threshold test *is an estimator of the form:*

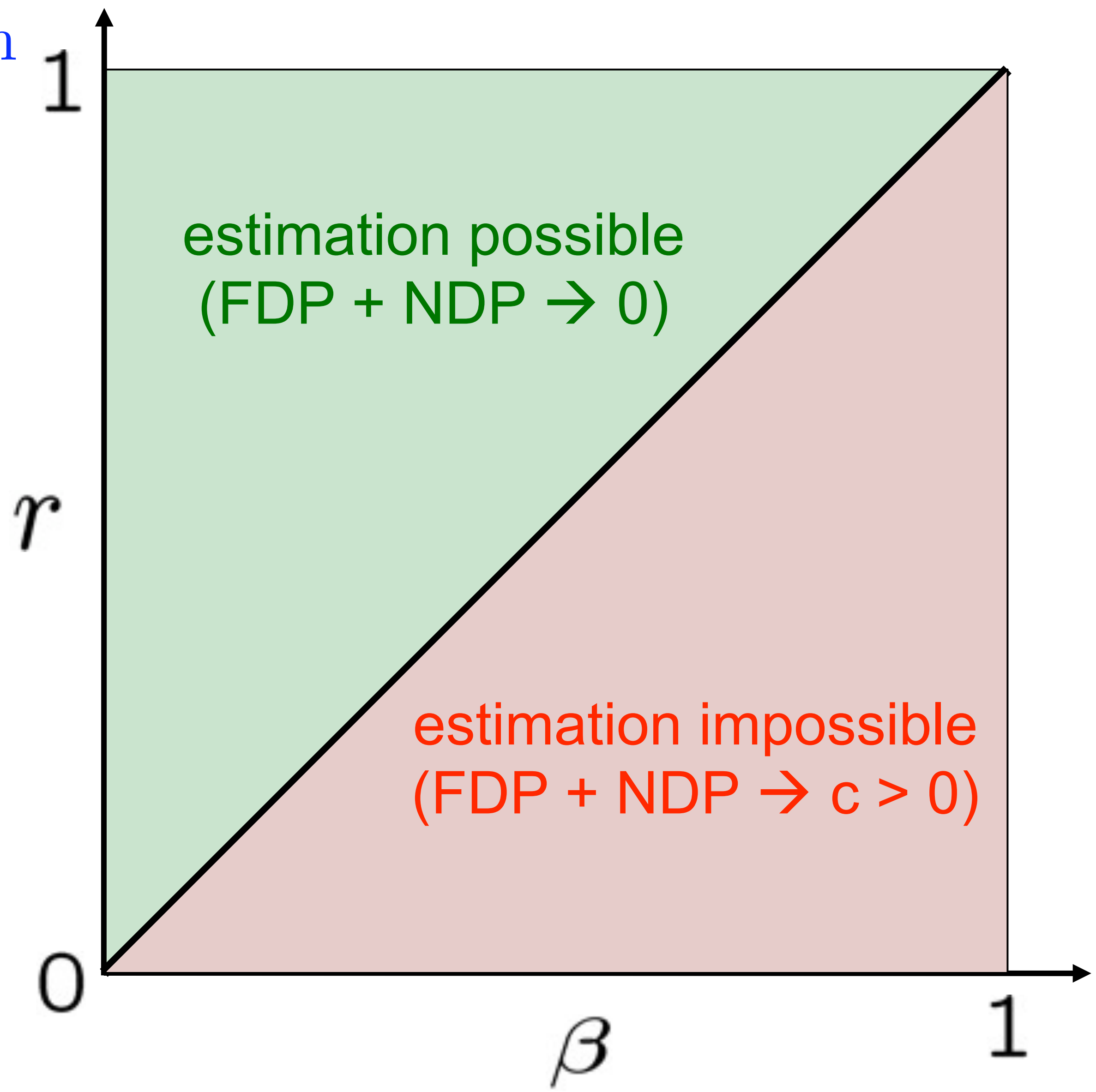
$$\widehat{\mathcal{S}}_\tau(y) \;:=\; \{i \in \{1, \dots, n\} : y_i \geq \tau > 0\}$$

Bonferroni Correction: To keep the error level small (e.g., less than 5%) the threshold must be on the order of $\sqrt{\log n}$.

False Discovery Rate Control (Ingster '97, Jin & Donoho '03)

Assume sublinear sparsity level: $|\mathcal{S}| = n^{1-\beta}$, $\beta \in (0,1)$

signal strength
$\mu = \sqrt{2r \log n}$



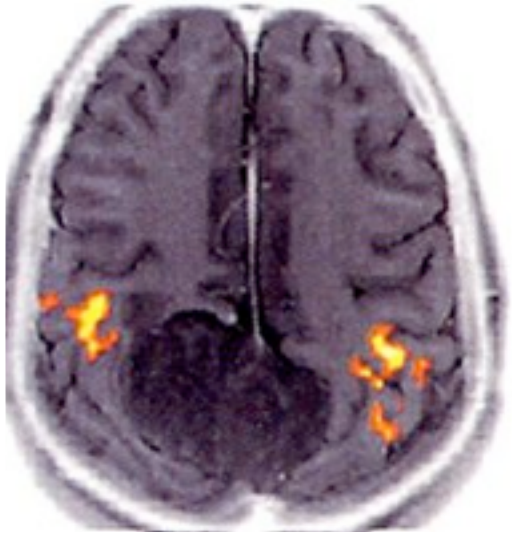
sparsity

$$\mathbf{FDP}(\widehat{\mathcal{S}}) := \frac{|\widehat{\mathcal{S}} \backslash \mathcal{S}|}{|\widehat{\mathcal{S}}|} = \frac{\#\ \text{false discoveries}}{\text{total}\ \#\ \text{discoveries}}$$

$$\mathbf{NDP}(\widehat{\mathcal{S}}) := \frac{|S \backslash \widehat{\mathcal{S}}|}{|S|} = \frac{\#\ \text{missed non-zeros}}{\#\ \text{true non-zeros}}$$

reliable detection iff $\mu \sim \sqrt{\log n}$ **!**

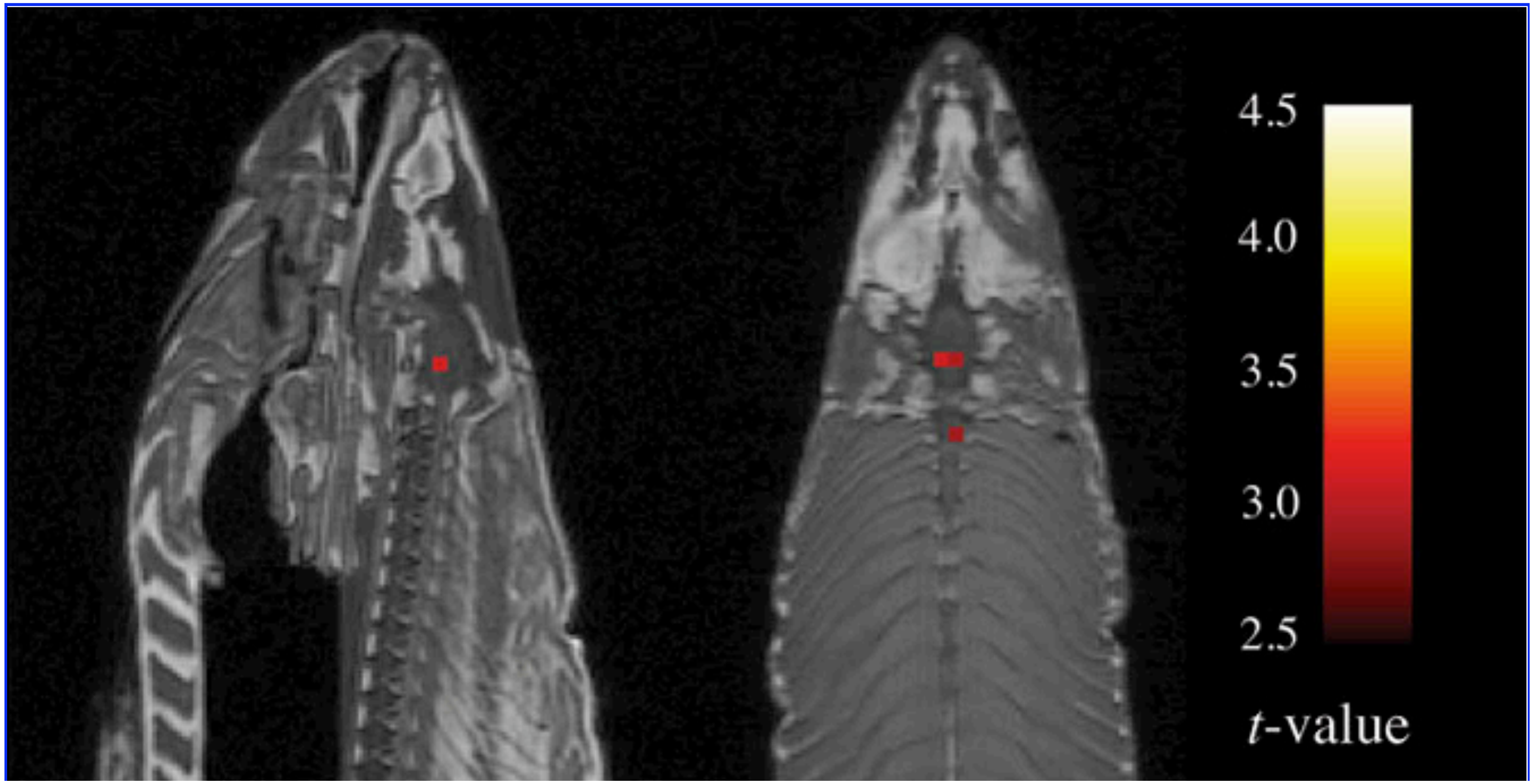
Is there really a problem ?



Wired Science
News for Your Neurons
Previous post
Next post

Scanning Dead Salmon in fMRI Machine Highlights Risk of Red Herrings

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



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

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

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 $\gamma_{i,1} = 1$, $i = 1, \dots, n$. 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

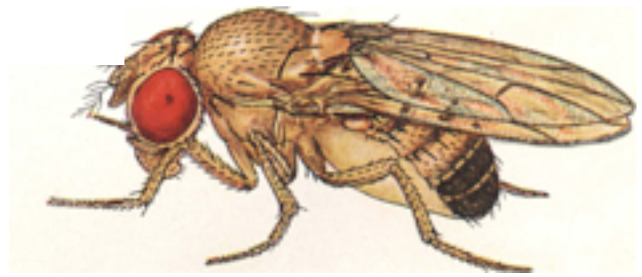
— more/fewer repeated samples or

— longer/shorter observation times

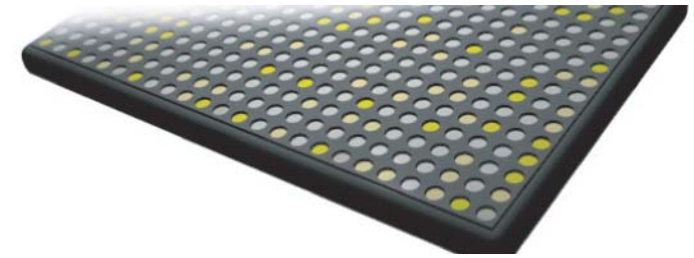
Fruit Fly Example

virus

fruit fly



assay



How to find genes involved in virus replication ?

Sequential Design Idea

Budget: **k assays**, n tests/assay

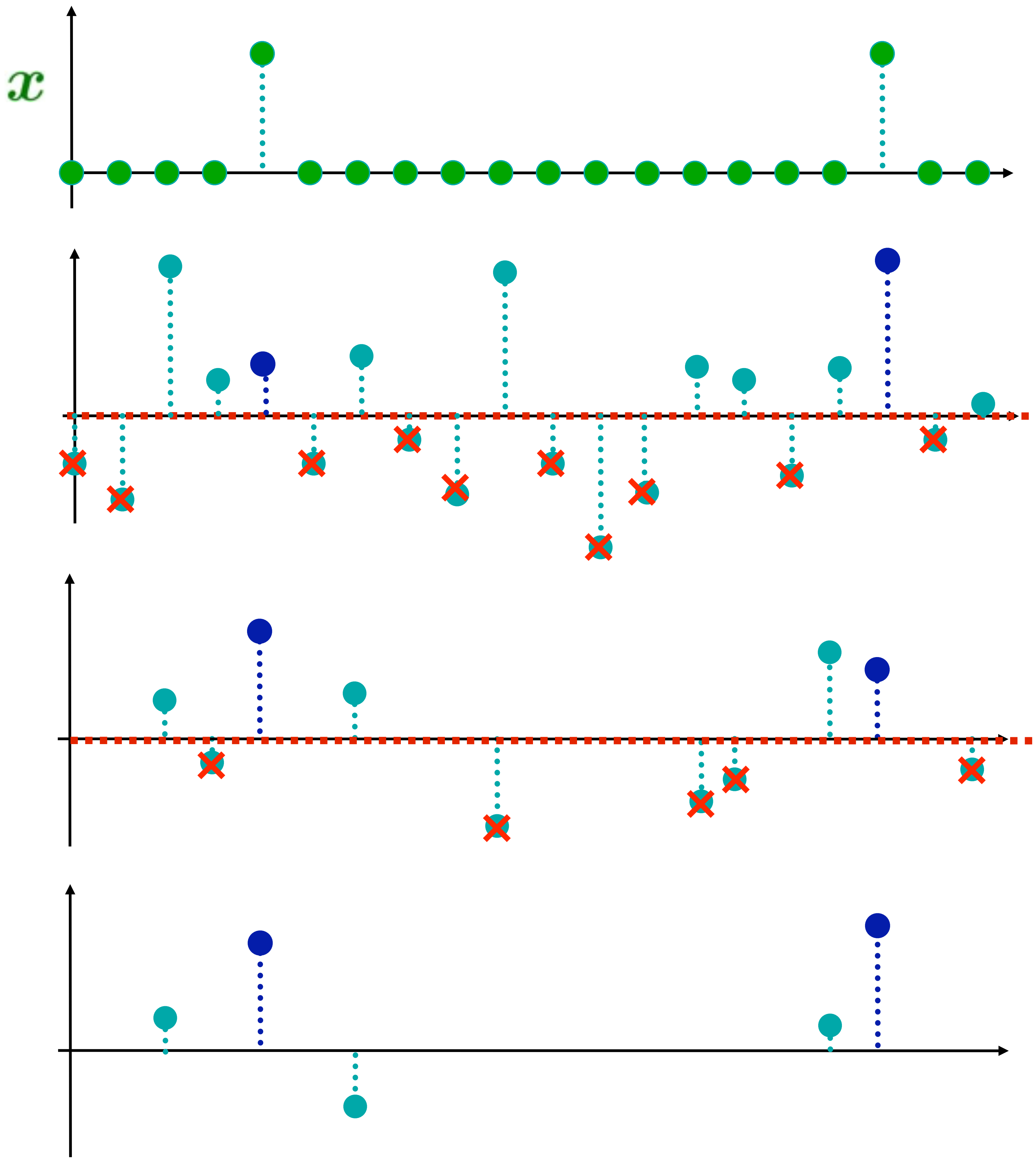
Assay 1: measure fluorescence of all n genes; discard n/2 genes with weakest fluorescence.

Assay 2: measure fluorescence for remaining n/2 genes, each tested twice (double SNR); discard n/4 genes with weakest fluorescence.

Assay 3: measure fluorescence for remaining n/4 genes, each tested four times (quadruple SNR); discard n/8 genes with weakest fluorescence.

continue distilling....

Idealized Example



Take $k = 3$ steps and split precision budget uniformly ($n/3$ per step)

$$y_{i,1} = x_i + \mathcal{N}(0, 3)$$

$$y_{i,2} = x_i + \mathcal{N}\left(0, \frac{3}{2}\right)$$

$$y_{i,3} = x_i + \mathcal{N}\left(0, \frac{3}{4}\right)$$

Distilled Sensing

***Simple* Distilled Sensing**

initialize: $\mathcal{S}_0 = \{1, \ldots, n\}$, $\gamma_{i,j}^{-1} = 2 + \epsilon$, $\epsilon > 0$

for $j = 1, \ldots, k$

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

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

end

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

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

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

(for n large)

probability of error:

$$
\begin{aligned}
\mathbb{P}(\mathcal{S}_k \neq \mathcal{S}) &= \mathbb{P}\left(\{\mathcal{S}^c \cap \mathcal{S}_k \neq \emptyset\} \cup \{\mathcal{S} \cap \mathcal{S}_k^c \neq \emptyset\}\right) \\
&\leq \mathbb{P}\left(\mathcal{S}^c \cap \mathcal{S}_k \neq \emptyset\right) + \mathbb{P}\left(\mathcal{S} \cap \mathcal{S}_k^c \neq \emptyset\right)
\end{aligned}
$$

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

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

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

Probability of Error Bound

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

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

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

Second term tends to zero if

$$
\mu \;\geq\; \sqrt{2(2+\epsilon)\log(|\mathcal{S}|(1+\epsilon)\log_2 n)}
$$

Gains of Sequential Design

non-adaptive threshold:

$$\mu \;\geq\; \sqrt{2 \log n}$$

DS threshold:

$$\mu \;\geq\; \sqrt{2(2+\epsilon)\log(|\mathcal{S}|(1+\epsilon)\log_2 n)}$$

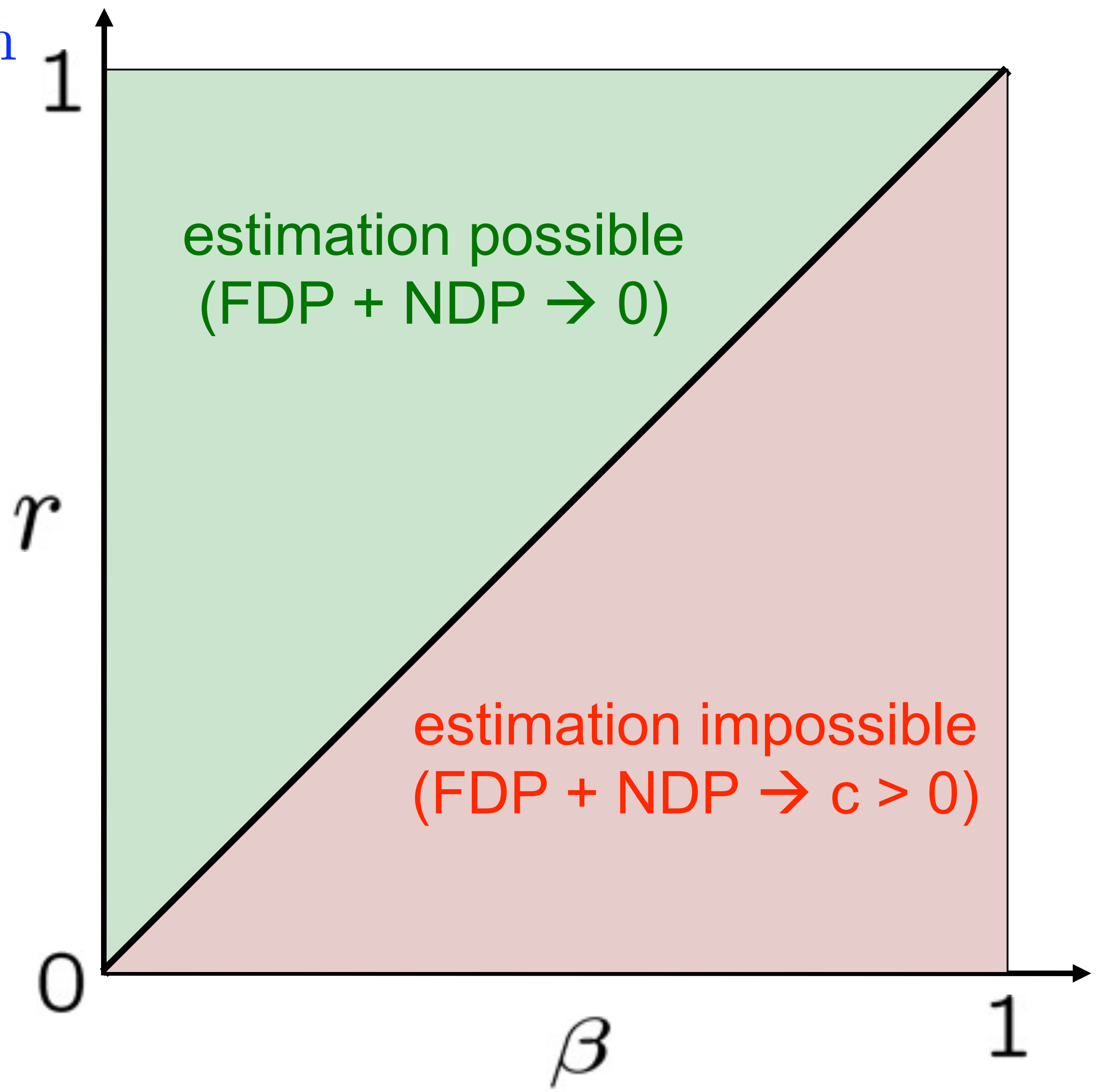
We get a gain whenever $|\mathcal{S}| \preceq n^{1/2}$

Punchline: In ultra-sparse setting, say $|\mathcal{S}| = C \log n$, DS drives error to zero if $\mu \geq \sqrt{(8+\epsilon)\,\log\log n}$, compared to the non-adaptive requirement $\mu \geq \sqrt{2\log n}$.

False Discovery Rate Control (Ingster '97, Jin & Donoho '03)

Assume sublinear sparsity level: $|\mathcal{S}| = n^{1-\beta}$, $\beta \in (0,1)$

signal strength
$\mu = \sqrt{2r \log n}$



sparsity

$$\mathbf{FDP}(\widehat{\mathcal{S}}) := \frac{|\widehat{\mathcal{S}} \backslash \mathcal{S}|}{|\widehat{\mathcal{S}}|} = \frac{\#\text{ false discoveries}}{\text{total }\#\text{ discoveries}}$$

$$\mathbf{NDP}(\widehat{\mathcal{S}}) := \frac{|S \backslash \widehat{\mathcal{S}}|}{|S|} = \frac{\#\text{ missed non-zeros}}{\#\text{ true non-zeros}}$$

non-sequential methods require $\mu \sim \sqrt{\log n}$

FDR-type Control using DS

***FDR* Distilled Sensing**

initialize: $\mathcal{S}_0 = \{1, \dots, n\}$, $k = \lceil \log \log n \rceil$

$\gamma_{i,j} = (\frac{3}{4})^j \frac{n}{8} / |\mathcal{S}_{j-1}|$, $j = 1, \dots, k-1$

$\gamma_{i,k} = \frac{n}{2|\mathcal{S}_{k-1}|}$

for $j = 1, \dots, k$

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

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

end

output: $\mathcal{S}_k = \{i : y_{i,k} \geq 4\}$

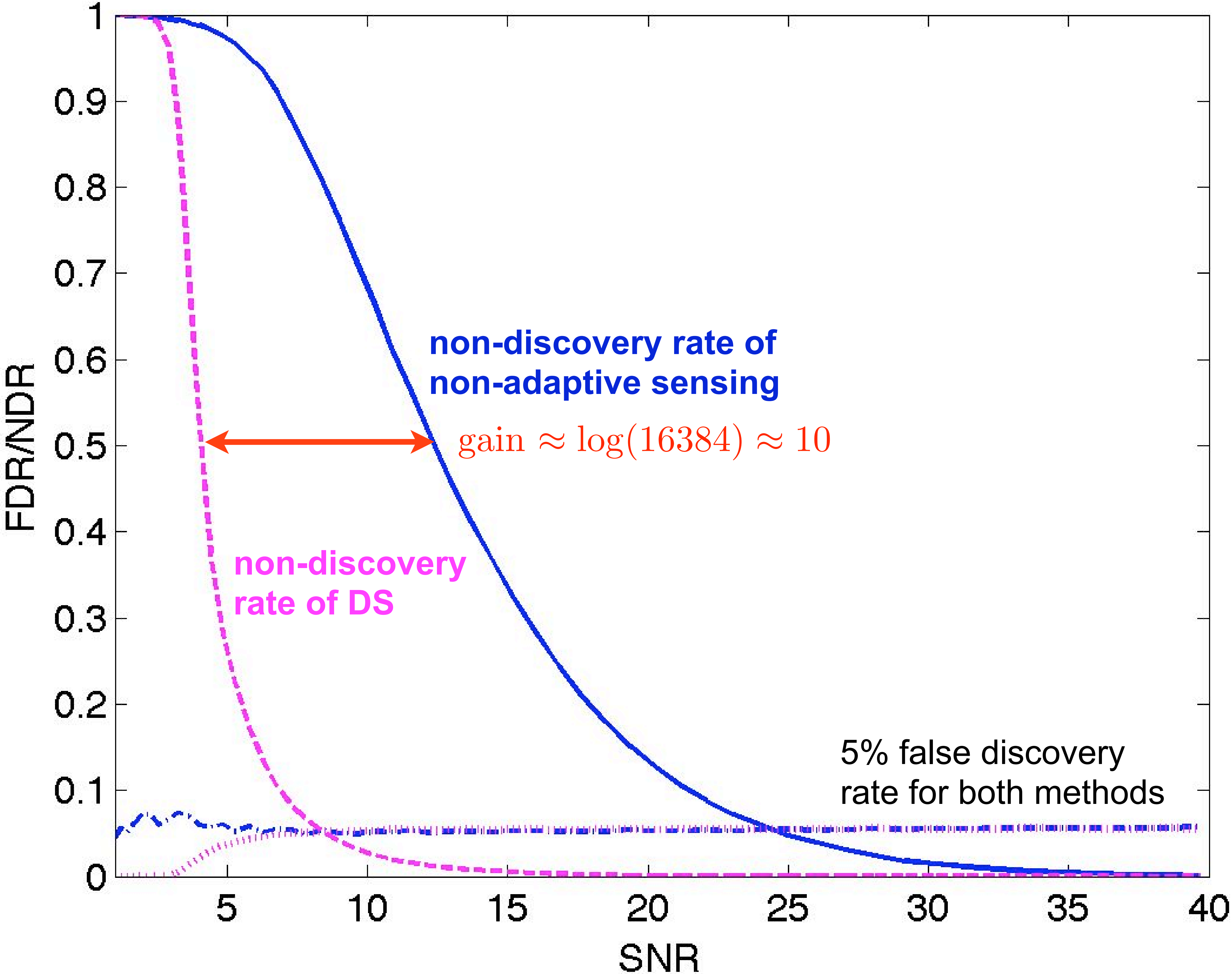
sublinear sparsity:
$|\mathcal{S}| = n^{1-\beta}$, $\beta \in (0,1)$

To guarantee that the proportions of FDP and NDP to zero as $n \to \infty$

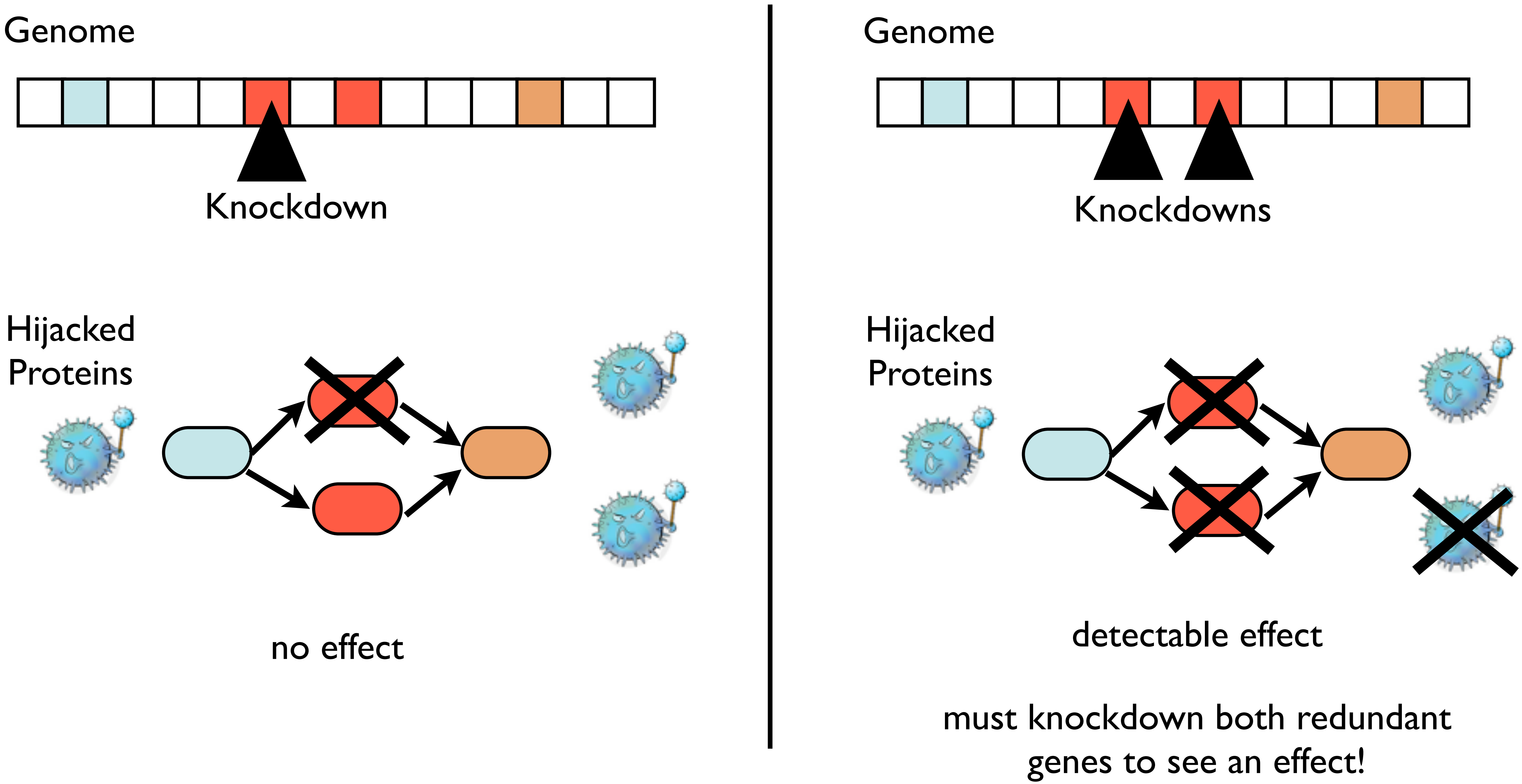
Distilled Sensing	$\mu \sim$ arbitrarily slowly growing function of n
non-adaptive	$\mu \sim \sqrt{\log n}$

Adaptivity effectively eliminates the fundamental statistical challenge in high-dimensional multiple testing.

Example $n = 2^{14}$, $\|x\|_0 = \sqrt{n} = 128$

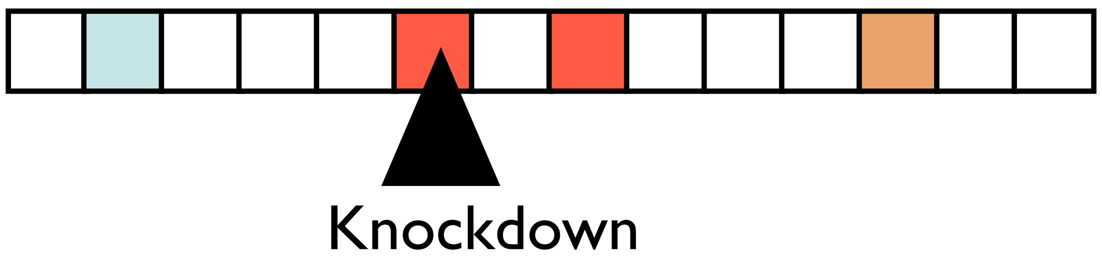


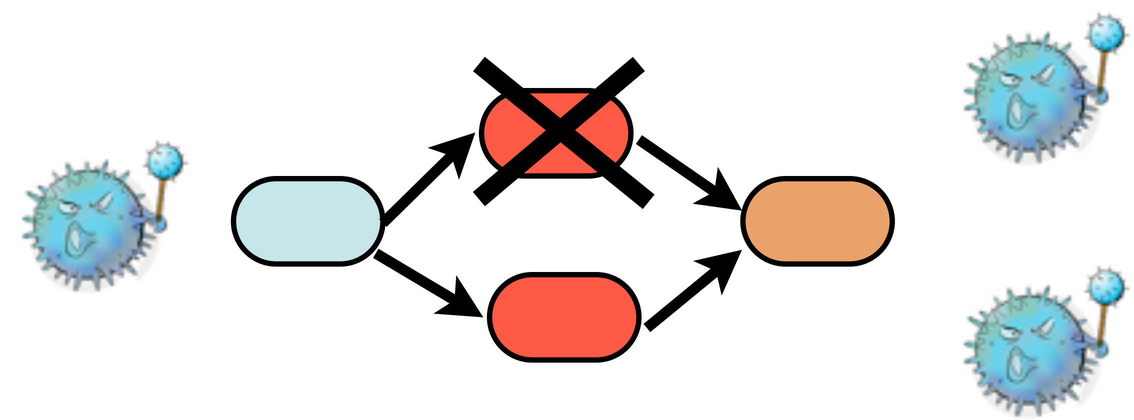
Challenge 2: Nonlinearities



$\binom{13000}{2} \approx 85,000,000$ possible two-fold gene deletion strains !

Sparse Interaction Models





$$y = \sum_i a_i x_i^{(1)} + \sum_{i<j} a_i a_j x_{ij}^{(2)}$$

Approximate output (virus reproduction) with a sparse bilinear system.

$x_i^{(1)}$ non-zero iff gene is critical to pathway

$x_{ij}^{(2)}$ non-zero iff gene pair is critical to pathway

a_i 1 if gene is knocked down; 0 otherwise

sparsity = most are 0

Sensing Sparse Interactions

Linear model: $y(x) \;=\; \sum_i x_i\, a_i$

Bilinear model: $y(x) \;=\; \sum_i x_i^{(1)}\, a_i \;+\; \sum_{i<j} x_{i,j}^{(2)}\, a_i a_j$

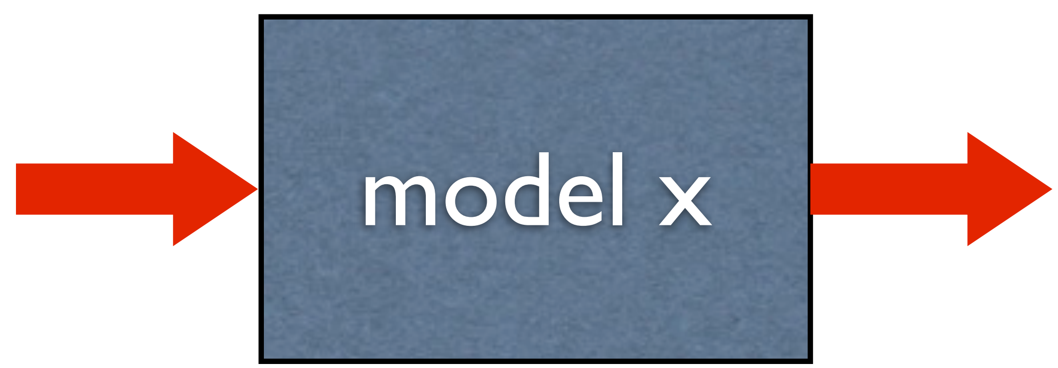
$\binom{13000}{2} \approx 85{,}000{,}000$ possible pairwise interactions !

Since most coefficients, $\{x_i\}$ or $\{x_i^{(1)}, x_{ij}^{(2)}\}$, are zero our goal is to identify critical components and interactions from using very few measurements

$a = \{+, -, +, +, \cdots, -, -, +\}$

random bit sequence

e.g., random selection of multiple simultaneous gene knock-downs



$y(x)$

collect $K \ll$ size(x) measurements $y_1, y_2, \ldots, y_K$ using K random inputs

(Linear) Compressed Sensing

This is the conventional compressed sensing problem for the linear model.

$$y_k \;=\; \sum_i a_{ki}\, x_i \;, \quad k = 1, \ldots, K$$

$\mathbf{x} \in \mathbb{R}^n, \;\; \mathbf{a} \in \{-1, +1\}^n$

sparse $\mathbf{x}$: $\;\|\mathbf{x}\|_0 \;=\; S \;\ll n$

find sparse solution to $\mathbf{y} \;=\; \mathbf{A}\,\mathbf{x}$

If the measurement matrix satisfies the restricted isometry property (RIP) with $\delta_{2S} < \sqrt{2} - 1$ for all S-sparse vectors:

$$(1 - \delta_S)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_S)\|\mathbf{x}\|_2^2$$

then x can be recovered from y by convex optimization:

$$\min \|\mathbf{z}\|_1 \text{ subject to } \mathbf{A}\mathbf{z} = \mathbf{y}$$

RIP holds with high probability if $K \;\geq\; c\,S\,\log(n/S)$

Multilinear Compressed Sensing

Multilinear model:

$$y \;=\; \sum_{i_1 < i_2 < \cdots < i_D} a_{i_1} a_{i_2} \cdots a_{i_D} x_{i_1 i_2 \cdots i_D} \qquad \mathbf{x} \in \mathbb{R}^N, \;\; \mathbf{a} \in \{-1, 1\}^n$$

$$\|x\|_0 \;=\; S \;\ll N \qquad N = \binom{n}{D}$$

y is called a ***Rademacher chaos*** of order D

Compressed sensing problem for multilinear model

K measurements of this form: $\mathbf{y} = [y_1 \cdots y_K]^T$

find sparse solution to $\mathbf{y} \;=\; \mathbf{A}\,\mathbf{x}$

matrix $\mathbf{A}$ now composed of monomials in $a_{i,j}$

Does it satisfy the RIP property?

On Average, Things Look Good

RIP: $(1-\delta_S)\|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1+\delta_S)\|\mathbf{x}\|_2^2$

isotropic measurements: $\mathbb{E}\left[\|\mathbf{A}\mathbf{x}\ \|^2\right] \ = \ \|\mathbf{x}\|_2^2$

symmetric binary random inputs: $\mathbb{P}(a_{i,j}=+1) \ = \ \mathbb{P}(a_{i,j}=-1) \ = \ 1/2$

Linear CS: $n=3$ inputs, $K=2$ measurements and $D=1$,

$$\mathbf{A} = \frac{1}{\sqrt{2}}\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix} \ \Rightarrow \ \mathbb{E}[\mathbf{A}^T\,\mathbf{A}] \ = \text{Identity}$$

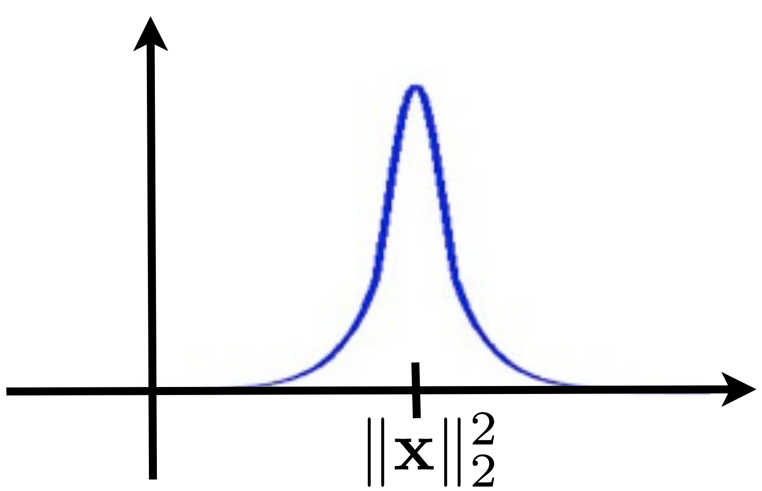
Bilinear CS: $n=3$ inputs, $K=2$ measurements and $D=2$,

$$\mathbf{A} = \frac{1}{\sqrt{2}}\begin{bmatrix} a_{1,1}a_{1,2} & a_{1,1}a_{1,3} & a_{1,2}a_{1,3} \\ a_{2,1}a_{2,2} & a_{2,1}a_{2,3} & a_{2,2}a_{2,3} \end{bmatrix} \ \Rightarrow \ \mathbb{E}[\mathbf{A}^T\,\mathbf{A}] \ = \text{Identity}$$

What about the distributions?

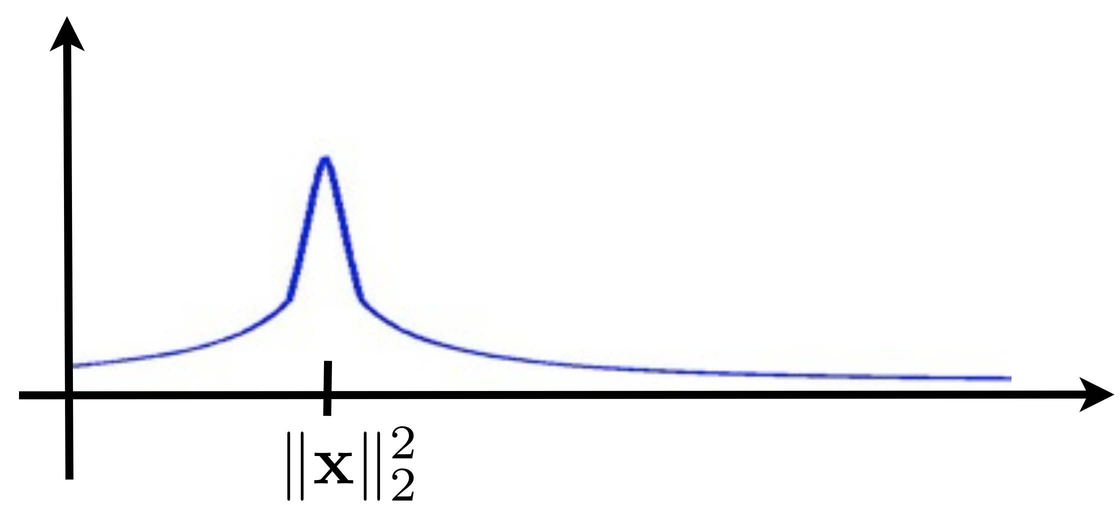
$$\mathbf{y} \;=\; \mathbf{A}\,\mathbf{x}\;,\;\; \mathbb{E}\|\mathbf{y}\|_2^2 = \|\mathbf{x}\|_2^2$$

$$\mathbb{P}(|\|\mathbf{y}\|^2 - \|\mathbf{x}\|_2^2| > t) \;\sim\; \exp(-\text{poly}(t))$$



Gaussian tails
(as in linear CS)

or



heavy tails?

Tail Behavior

Best case: decoupled chaos

$$\begin{aligned} y &= a_1\,a_2\,x_{1,2} + a_3\,a_4\,x_{3.4} + \cdots + a_{2k-1}\,a_{2k}\,x_{2k-1,2k} \\ &\equiv \widetilde{a}_1\,x_{1,2} + \widetilde{a}_2\,x_{3,4} + \cdots + \widetilde{a}_k\,x_{2k-1,2k} \end{aligned}$$

equivalent to iid binary symmetric sensing

$\Rightarrow$ subgaussian tails independent of D: $\mathbb{P}(y^2 - \|x\|_2^2 > t) \;\leq\; \exp(-c\,t)$

Worst case: strongly coupled chaos

$$y = \sum_{1\leq i<j\leq \ell} a_i\,a_j\,x_{i,j}\;, \qquad k := \binom{\ell}{2}$$

significant probability of large deviations from mean:

$$\text{if } x_{i,j} = 1/\sqrt{k}, \text{ then } \mathbb{P}(y^2 \geq k) = 2^{-\ell} = 2^{-c\,k^{1/2}}$$

$\Rightarrow$ heavy tails depending on D: $\mathbb{P}(y^2 - \|x\|_2^2 > t) \;\geq\; \exp(-c\,t^{1/D})$

Combinatorial Dimension of Rademacher Chaos

The combinatorial dimension $1 \leq \alpha \leq D$ measures the level of dependence introduced by a particular pattern of sparsity.

$$y \;=\; \sum_{i_1 < i_2 < \cdots < i_D} a_{i_1} a_{i_2} \cdots a_{i_D} x_{i_1 i_2 \cdots i_D}$$

Blei-Janson '04: A Rademacher chaos with combinatorial dimensional α satisfies

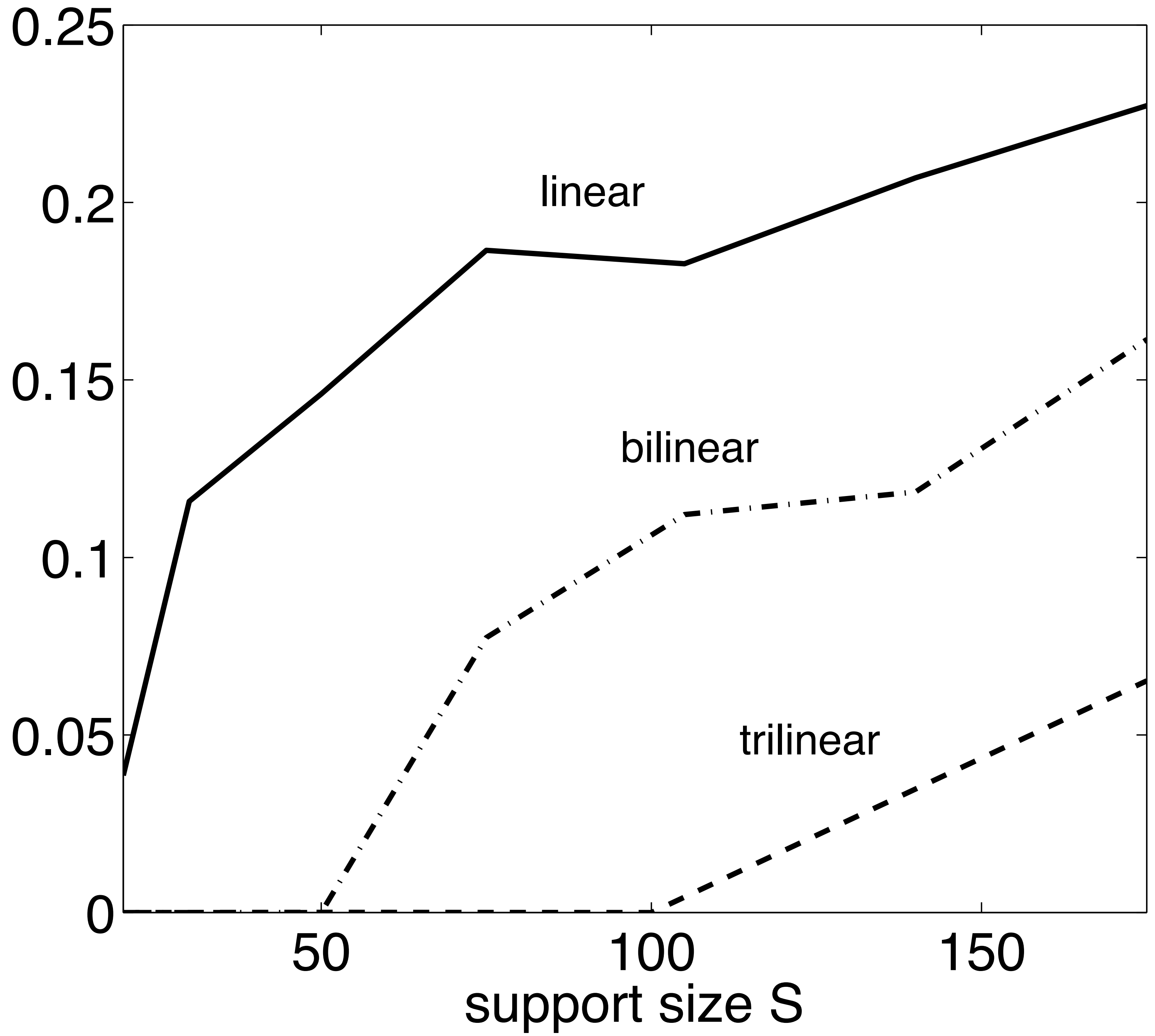
$$\exp\left(-c_1 t^{1/\alpha}\right) \leq \mathbb{P}\left(|y|^2 > t\right) \leq \exp\left(-c_2 t^{1/\alpha}\right)$$

tails are light to heavy, depending on $1 \leq \alpha \leq D$

Dependencies Matter (in practice)

$$\text{RIP:} \quad (1-\delta_S)\|\mathbf{x}\|_2^2 \le \|\mathbf{A}\mathbf{x}\|_2^2 \le (1+\delta_S)\|\mathbf{x}\|_2^2$$

$1-\delta_S$



$$K \;=\; 5\,S$$

Dependencies Matter (in theory)

$K =$ number of measurements needed to recovery S-sparse multilinear forms

proof technique	bound on K	ingredients
Gershgorin	$S^2 \log N$	empirical 2nd moment bounds, union bound
Rudelson-Vershynin	$S(\log^3 S)(\log N)$	empirical 2nd moment bounds, no union bound
Rademacher chaos	$S^\alpha \log^\alpha (N/S)$ $1 \leq \alpha \leq D$	tail bounds union bound

compare with linear CS bound: $K \geq S \log(N/S)$

Gershgorin Bound

i) Control each element of (partial) Gram matrix $\mathbf{G}_\mathcal{T} = \mathbf{A}_\mathcal{T}^T \mathbf{A}_\mathcal{T}$ using Hoeffding's inequality and bound probability that $\mathbf{G}_\mathcal{T}$ is approximately diagonal.

$$\mathbf{G}_\mathcal{T} = \begin{bmatrix} 1 & \frac{\delta_S}{S} & \cdots & \frac{\delta_S}{S} \\ \frac{\delta_S}{S} & 1 & \cdots & \frac{\delta_S}{S} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\delta_S}{S} & \frac{\delta_S}{S} & \cdots & 1 \end{bmatrix}$$

ii) Gershgorin's Disc Theorem guarantees that eigenvalues lie in the range

$$g_{ii} - \sum_{j \neq i} |g_{ij}| \;\leq\; \lambda_i(\mathbf{G}_\mathcal{T}) \;\leq\; g_{ii} + \sum_{j \neq i} |g_{ij}|$$

iii) union bound over all $\binom{N}{S}$ sparsity patterns. RIP holds if

$$K \geq c\, S^2 \log N$$

Heavy-Tailed Restricted Isometries

Theorem 1 (Vershynin) Let $\widetilde{\mathbf{A}}$ be a $K \times N$ measurement matrix whose rows $\mathbf{a}_i^T$ are independent isotropic random vectors in $\mathbb{R}^N$. Let B be a number such that all entries $|a_{ij}| \leq B$ almost surely. Then the normalized matrix $\mathbf{A} = \frac{1}{\sqrt{K}}\widetilde{\mathbf{A}}$ satisfies the following for $K \leq N$, for every sparsity level $S \leq N$ and $0 < \epsilon < 1$:

if the number of measurements satisfies

$$K \geq C\ \epsilon^{-2} S \log N \log^3(S)$$

then the RIP constant δ_S of $\mathbf{A}$ satisfies $\mathbb{E}[\delta_S] \leq \epsilon$.

Check conditions:

- isotropy: $\mathbf{G} := \mathbf{A}^T\mathbf{A}$, $\mathbb{E}[\mathbf{G}] = \text{Identity}$
- elements of $\widetilde{\mathbf{A}}$ bounded by 1.

Chaos Tail Bound

Lemma 1 Assume that y_k, $k = 1, \dots, K$, are i.i.d. Rademacher variables of order D with combinatorial dimension $1 \leq \alpha \leq D$ and $\mathbb{E} y_k^2 = 1$. There exist constants $c, C > 0$ such that

$$\mathbb{P}\left(\left| \frac{1}{K} \sum_{k=1}^{K} y_k^2 - 1 \right| > t \right) \leq C \exp(-c \min(K t^2, K^{1/\alpha} t^{1/\alpha}))$$

proof technique:

- Blei-Jansen chaos tail bounds
- moment bound for sums of symmetric i.i.d. variables due to R. Latala
- apply lemma and union bound over ϵ-net for sparse vectors (technique from Baraniuk-Devore-Davenport-Wakin ’08)

RIP holds if $K \geq C\ S^\alpha \log^\alpha(N/S)$

Conclusions

1. Sequential Experimental Designs for High-Dimensional Testing

thresholds for recovery in high-dimensional limit:

non-adaptive designs $\quad \text{SNR} \sim \log n$

sequential designs $\quad \text{SNR} \sim$ arbitrarily slowly growing function of n

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

2. Compressed Sensing of Sparse Multilinear Functions

number of compressed sensing measurements for sparse recovery:

linear sparsity $\quad K \sim S \log n$

multilinear sparsity $\quad K \sim \min\{S^2 \log n, S \log^3(S) \log n, S^\alpha \log^\alpha n\}$

where $\alpha \geq 1$ *depends* on pattern of sparsity

Sparse Interactions: Identifying High-Dimensional. Multilinear
Systems via Compressed Sensing, B. Nazer and RN, Allerton 2010