Written by: Edo Liberty
Presented by: Ryan Rogers
With Some Slides from: Edo Liberty

# SIMPLE AND DETERMINISTIC MATRIX SKETCHING

# Set up

- $A$ is an $n$ x $m$ matrix
- We want to compute the $m$ x $m$ matrix: $A^T A$
- Problem: $n$ > machine memory.
- Goal: Find ‘good’ approximate $d$ x $m$ matrix $B$ for any $||x|| = 1$

$$||A^T A - B^T B|| \leq \text{Small}$$

# Set up

- $A$ is an $n \times m$ matrix
- We want to compute the $m \times m$ matrix: $A^T A$
- Problem: $n$ > machine memory.
- Goal: Find 'good' approximate $d \times m$ matrix $B$ for any $||x|| = 1$

$$||A^T A - B^T B|| \le \varepsilon ||A||_f^2$$

# Sketches

# Sketches

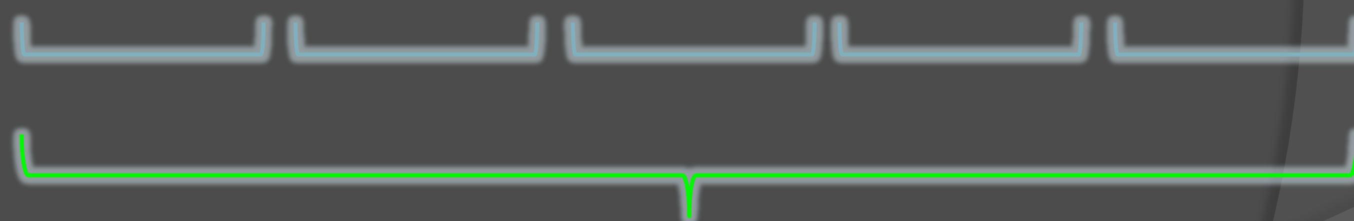
- A **sketch** of a matrix *A* is another matrix *B*, that is significantly smaller than *A* but still approximates *A* well.
- We need this if:
  - Rows of matrix can be processed only once
  - Storage is limited

# Frequent Items

- Universe $U = \{a_1, \ldots, a_m\}$ and a stream $A_1, A_2, \ldots, A_n$
- Frequency $f_i$ of item $a_i$ in the stream
- Use only $O(d)$ space to produce approximate counts $g_i$, such that

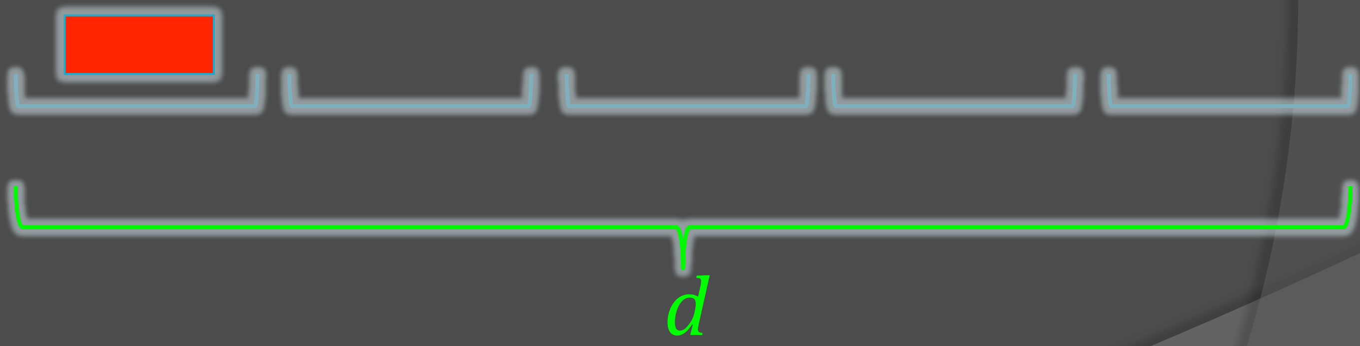
$$|f_i - g_i| < n/d$$

# Frequent Items

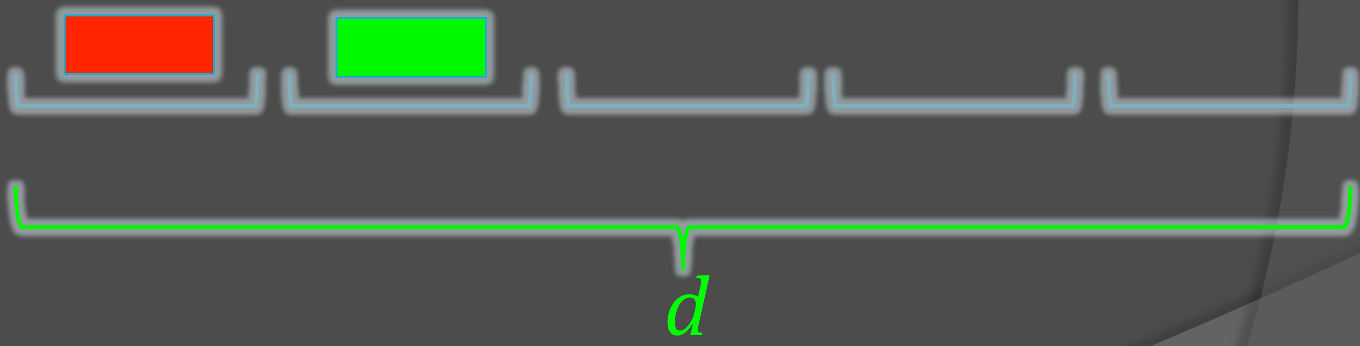


$d$

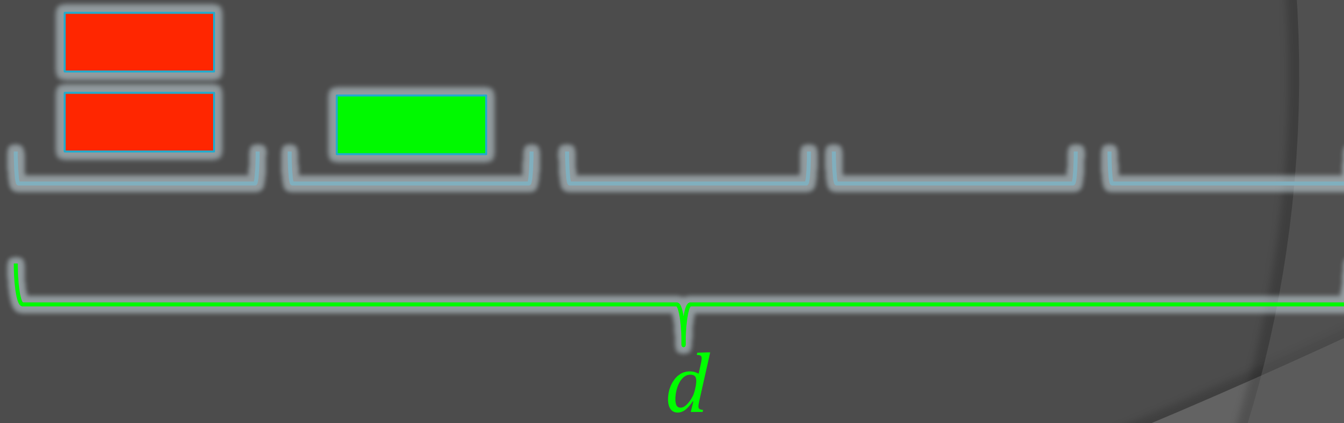
# Frequent Items



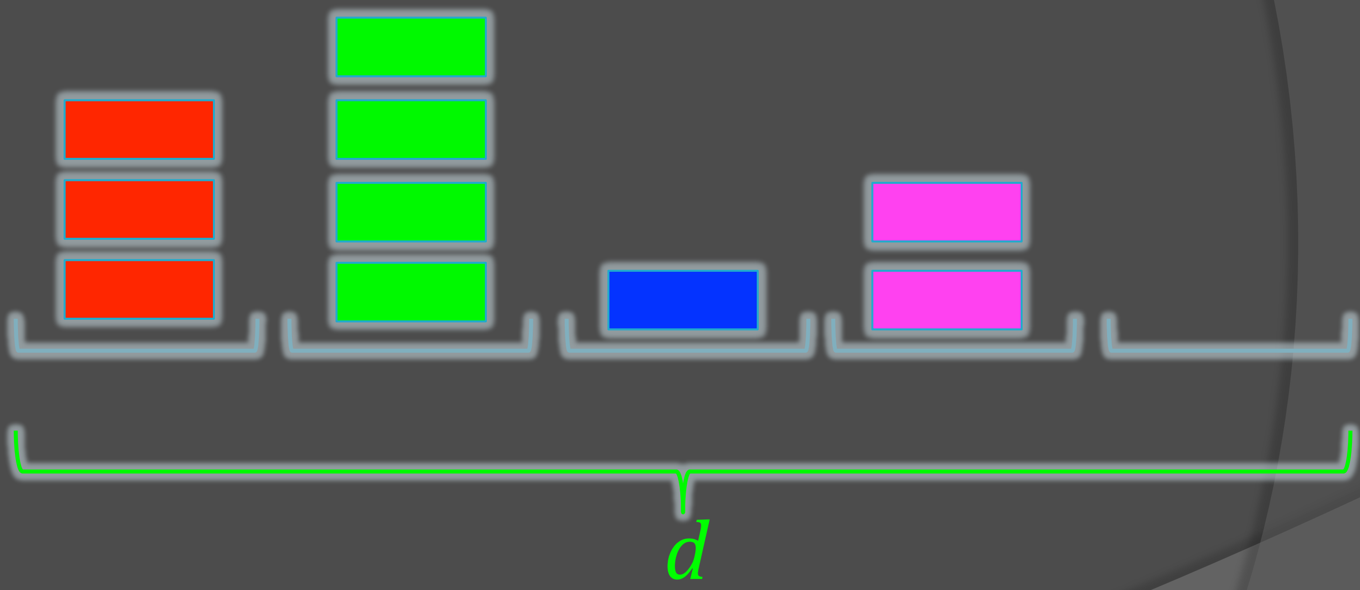
# Frequent Items



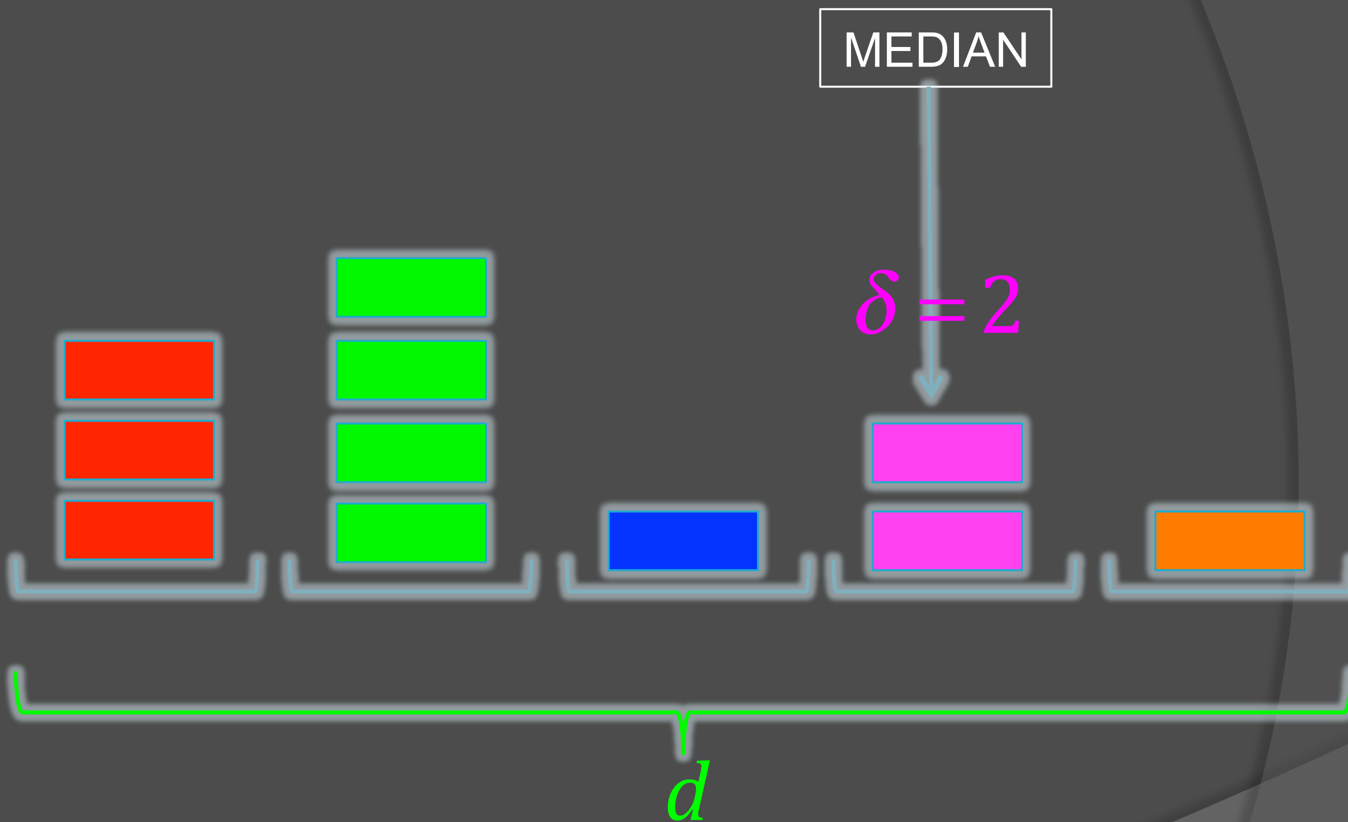
# Frequent Items



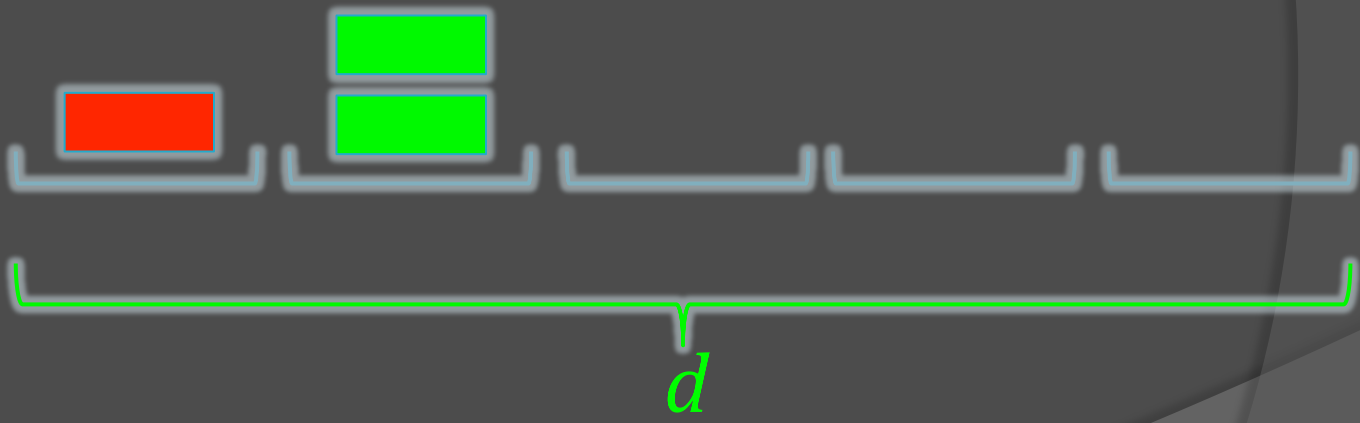
# Frequent Items



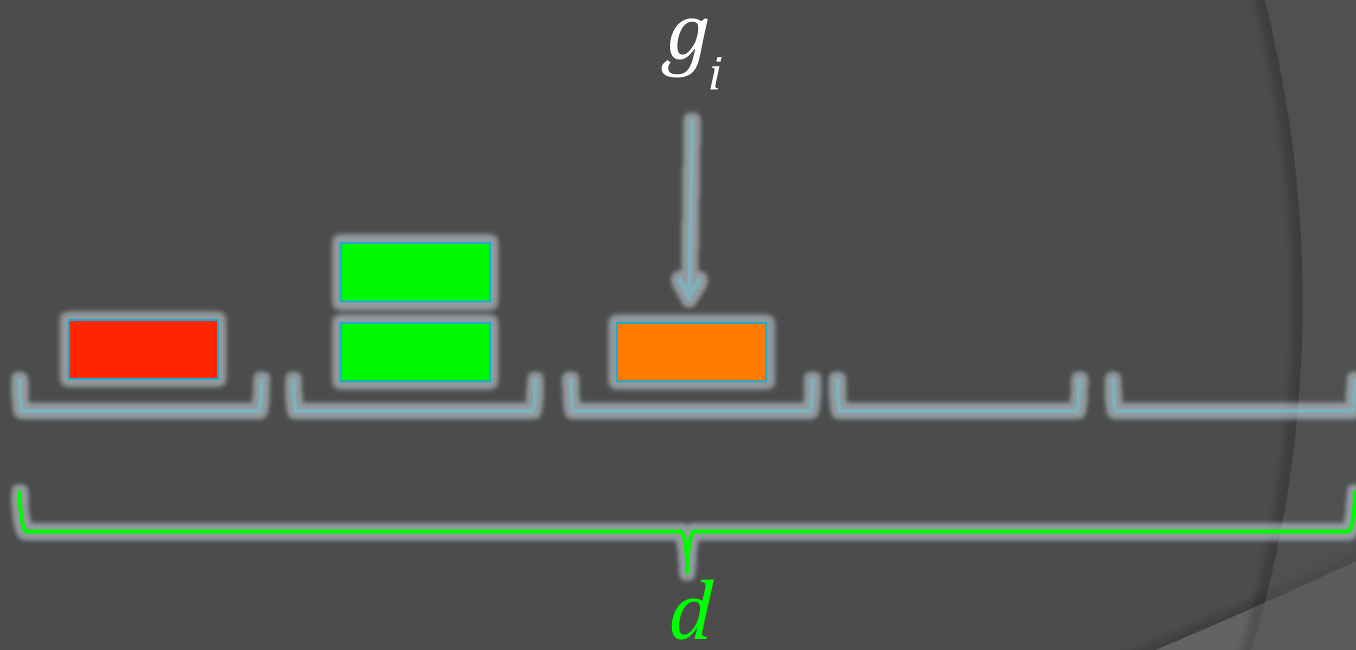
# Frequent Items



# Frequent Items



# Frequent Items



# Frequent Items – Observations

- We always get an undercount $g_i \le f_i$
- If we let $\delta_t$ be the amount we decrease counter at time *t* then
$$g_i \ge f_i - \sum_t \delta_t$$
- Sum up the undercounts
$$0 \le \sum_{i=1}^{d} g_i \le \sum_{t=1}^{n} \left( 1 - \frac{d}{2} \delta_t \right)$$

# Frequent Items – Observations

- Thus, we get $\sum_t \delta_t \leq 2n/d$
- Set $d = 2/\varepsilon$:

$$|f_i - g_i| \leq \varepsilon n$$

# Frequent Directions

$$\left[ \qquad\qquad\qquad\qquad \right]$$

d x m

# Frequent Directions

We now need to zero out some rows to make room for more!


d x m

# Frequent Directions

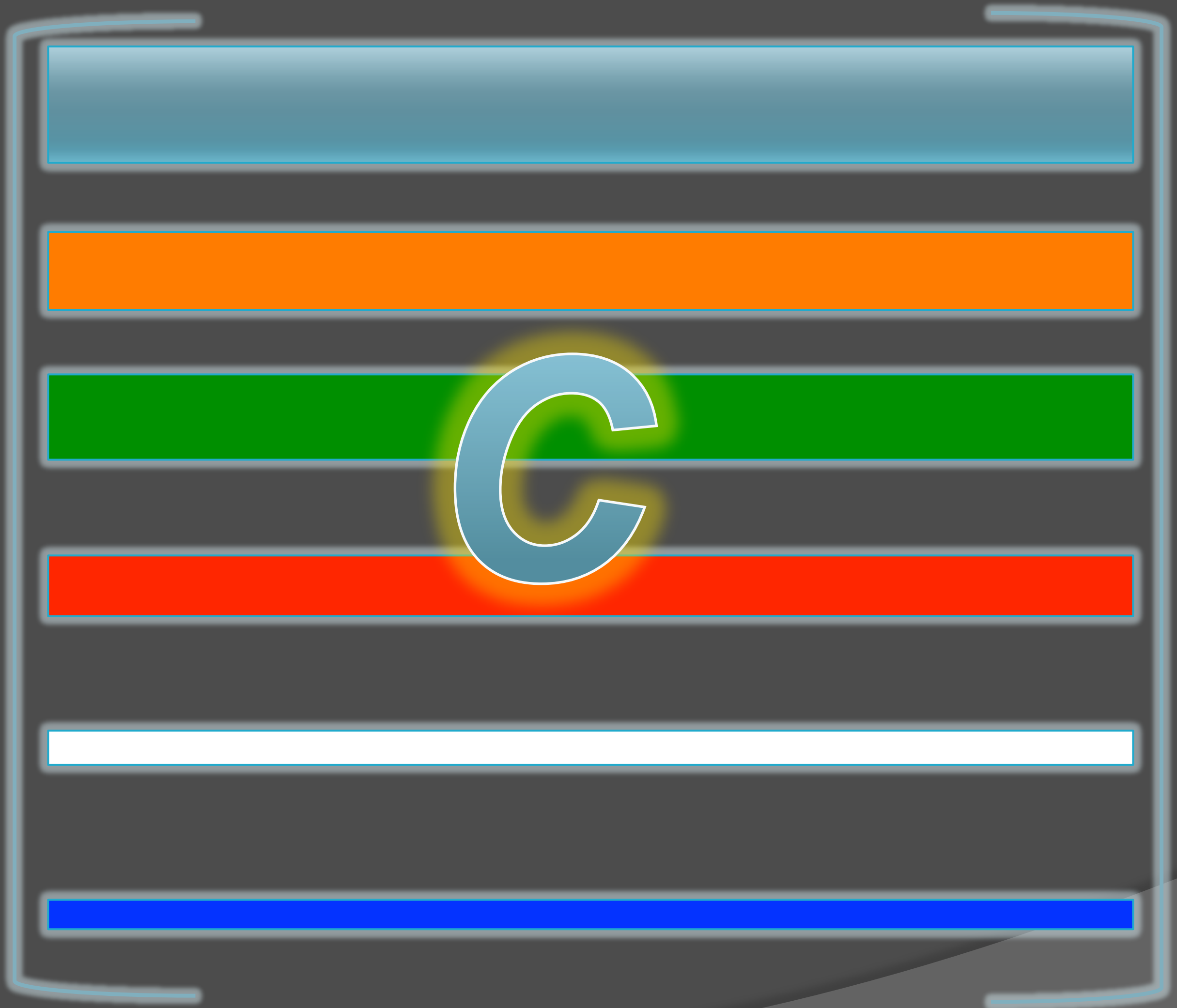
- Find SVD of *B*:

$$\underbrace{U_{d\times d}\,\overbrace{\Sigma_{d\times m} V^T_{m\times m}}^{=C}}= B$$

$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \sigma_3 & & & & \\ & & & \sigma_4 & & & \\ & & & & \sigma_5 & & \\ & & & & & \sigma_6 & \\ & & & & & & \sigma_7 \end{bmatrix}$$

Decreasing

# Frequent Directions




d x m

# Frequent Directions

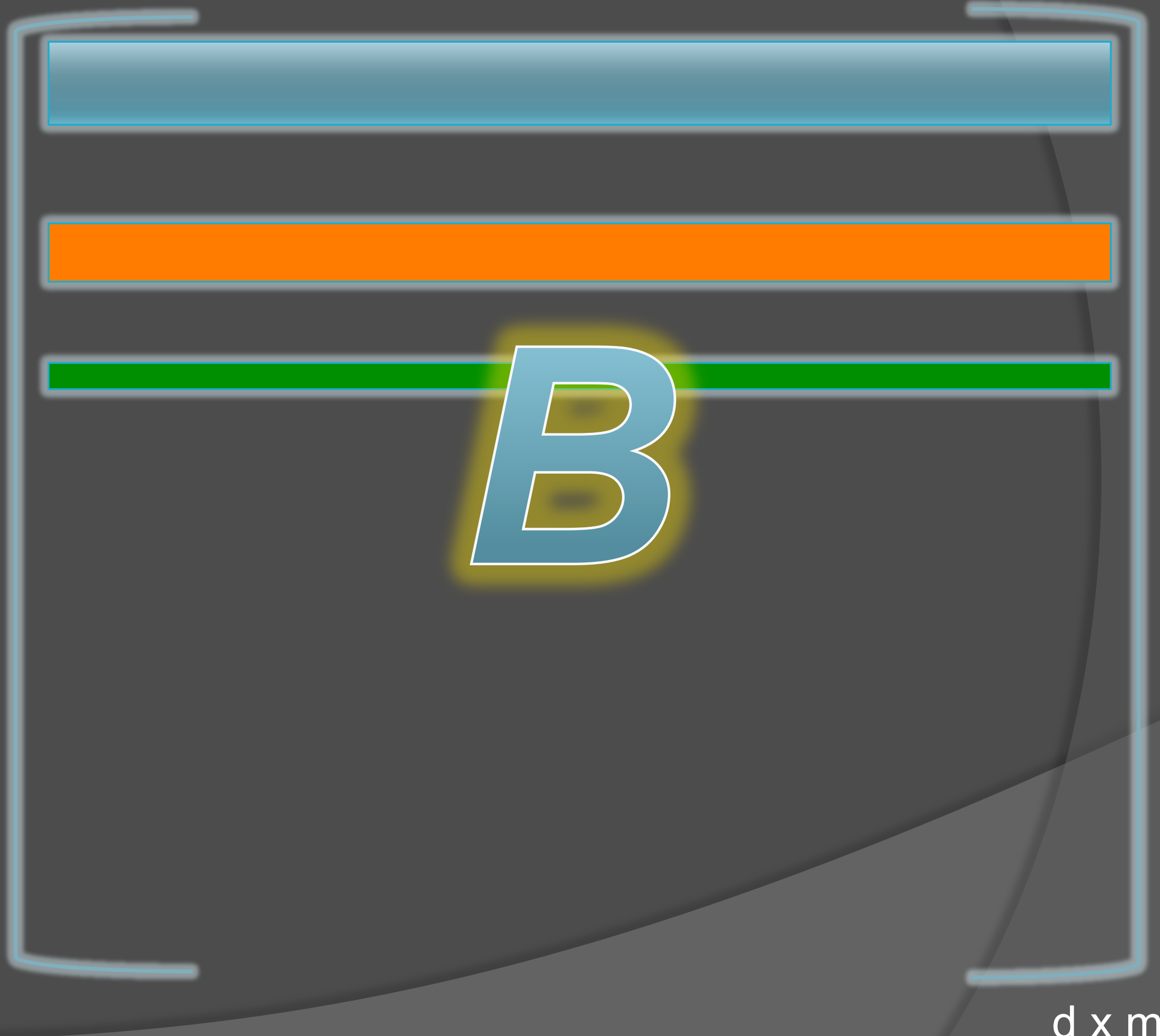
$$\Sigma = \begin{bmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \\ & & \sigma_3 & & & & \\ & & & \sigma_4 & & & \\ & & & & \sigma_5 & & \\ & & & & & \sigma_6 & \\ & & & & & & \sigma_7 \end{bmatrix}$$

$$\sigma_{d/2} = \sqrt{\delta}$$

$$\hat{\Sigma} = \sqrt{\max\left(\Sigma^2 - \delta I, 0\right)}$$

# Frequent Directions

$$B = \hat{\Sigma} V^T$$



# Frequent Directions

**Algorithm 1** *Frequent-directions*

**Input:** $\ell$, $A \in \mathbb{R}^{n \times m}$
$B \leftarrow$ all zeros matrix $\in \mathbb{R}^{\ell \times m}$
**for** $i \in [n]$ **do**
  Insert $A_i$ into a zero valued row of $B$
  **if** $B$ has no zero valued rows **then**
    $[U, \Sigma, V] \leftarrow \mathrm{SVD}(B)$
    $C \leftarrow \Sigma V^T$ # Only needed for proof notation
    $\delta \leftarrow \sigma_{\ell/2}^2$
    $\check{\Sigma} \leftarrow \sqrt{\max(\Sigma^2 - I_\ell \delta, 0)}$
    $B \leftarrow \check{\Sigma} V^T$ # At least half the rows of $B$ are all zero
  **end if**
**end for**
**Return:** $B$

# Analysis – Claim 1

- $B^TB$, $A^TA$, $A^TA\text{-}B^TB$ are all P.S.D.
- Proof: Check

$$||Ax||_2^2 - ||Bx||_2^2 \geq 0$$

# Analysis – Claim 2

- With sketch *B* of size *d* from Frequent Directions we have

$$||A^T A - B^T B|| \le 2||A||_f^2 / d$$

- Proof: First prove that for any unit vector *x*

$$||Ax||^2 - ||Bx||^2 \le 2/d\left(||A||_f^2 - ||B||_f^2\right)$$

# Analysis – Proof Continued

- Now we must show that for the largest e-vector $x$ that

$$||A^T A - B^T B|| = ||Ax||^2 - ||Bx||^2$$

# Run Time

- SVD of an *d x m* matrix of rank *r* takes

$$O(dmr) = O(d^2m)$$

- SVD is done once every *d/2* rows
- When SVD is not done, it takes time

$$O(m)$$

- Total run time:

$$O(dnm)$$

# Parallelization

If we have $A = \begin{bmatrix} A_1 \\ \hline A_2 \end{bmatrix}$ and $B_i = FD(A_i)$

then

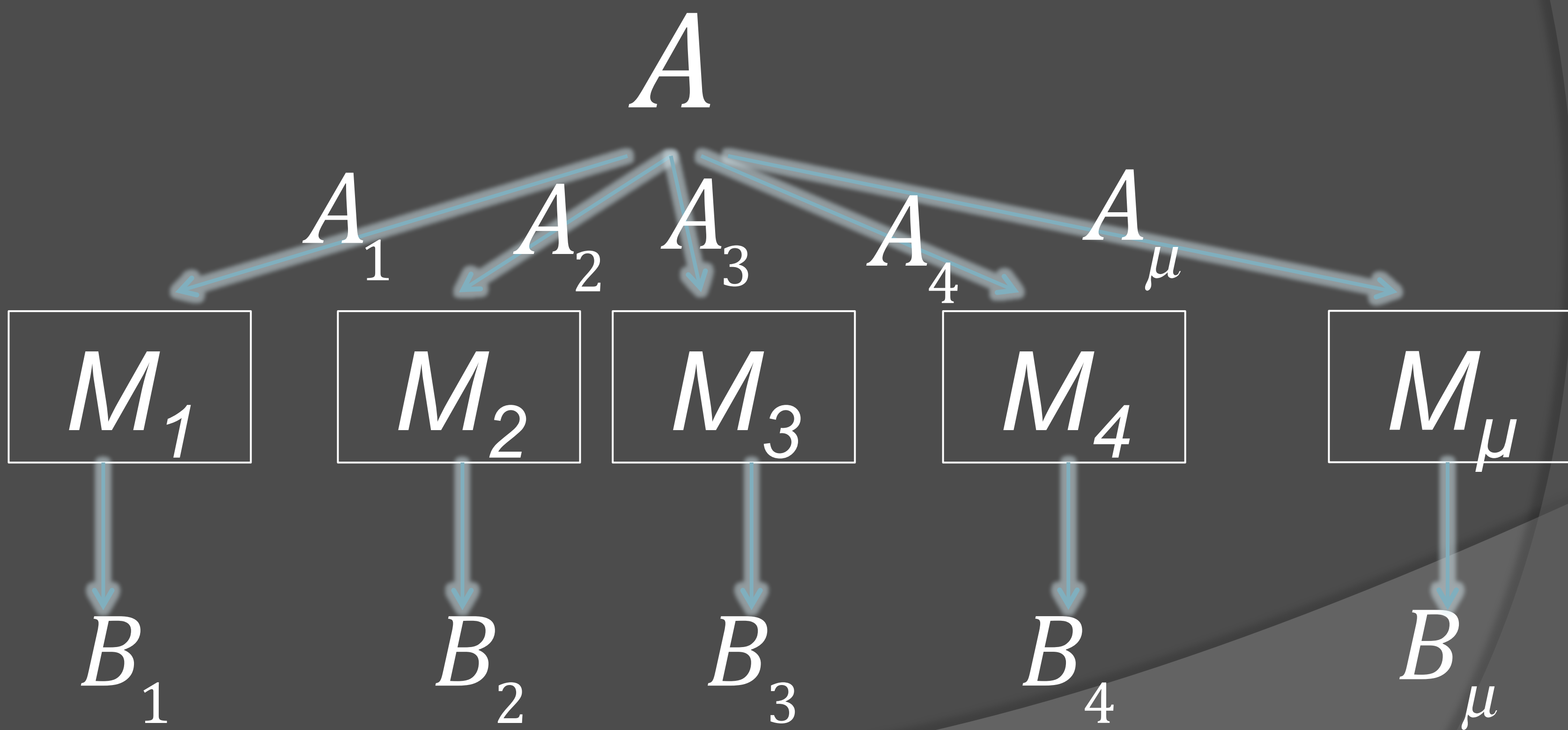
$$\|A^T A - D^T D\| \leq 2\|A\|_f^2 / d$$

where

$$D = FD\begin{pmatrix} B_1 \\ \hline B_2 \end{pmatrix}$$

# Parallelization

- Let there be $\mu$ machines and each takes $n/\mu$ many rows



# Parallelization

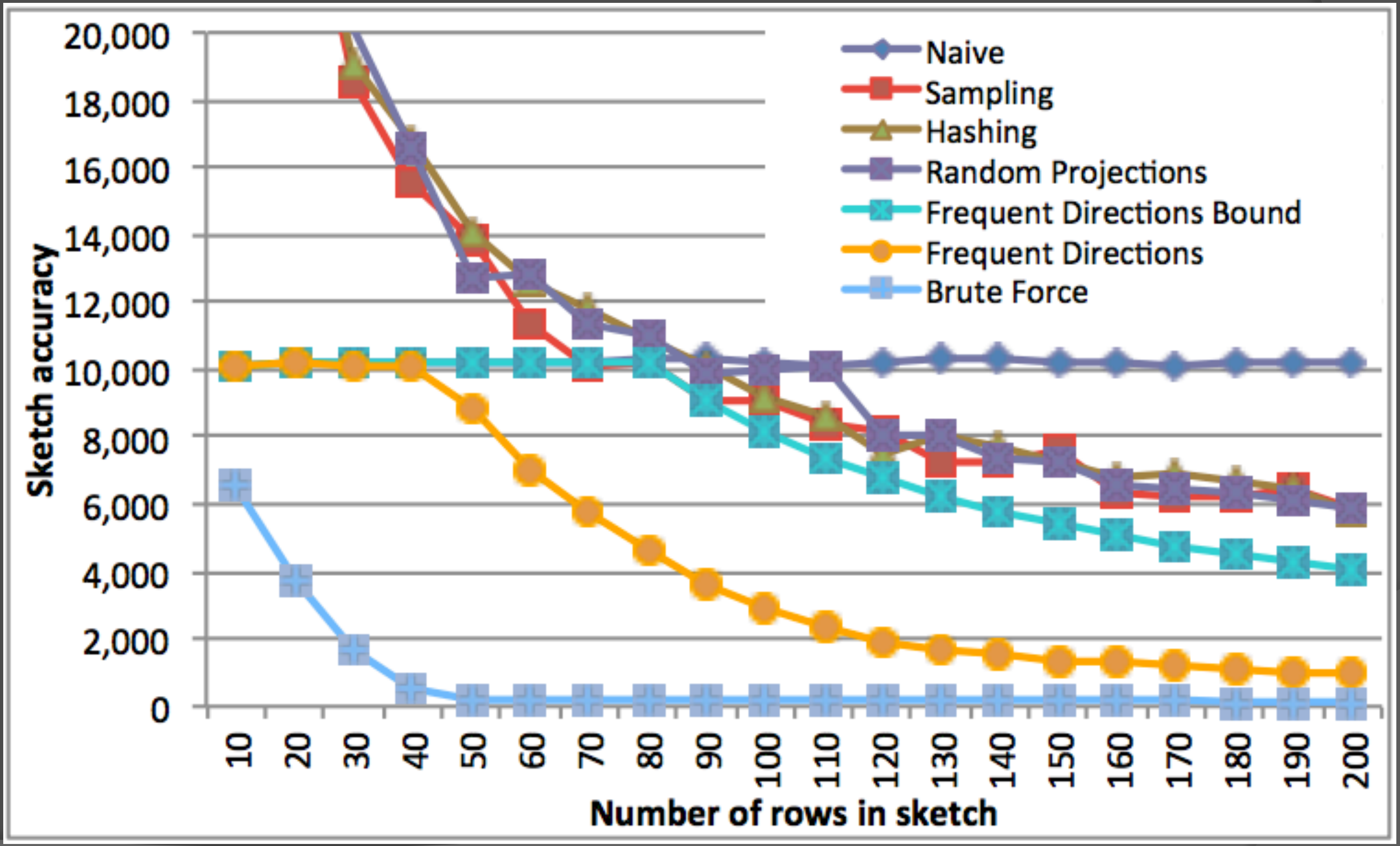
- Each $B_i$ has dimension $d \times m$.

- Each $M_i$ took time $O(dmn/\mu)$
- To then combine the others, can take $\mu$ more machines, and total run time

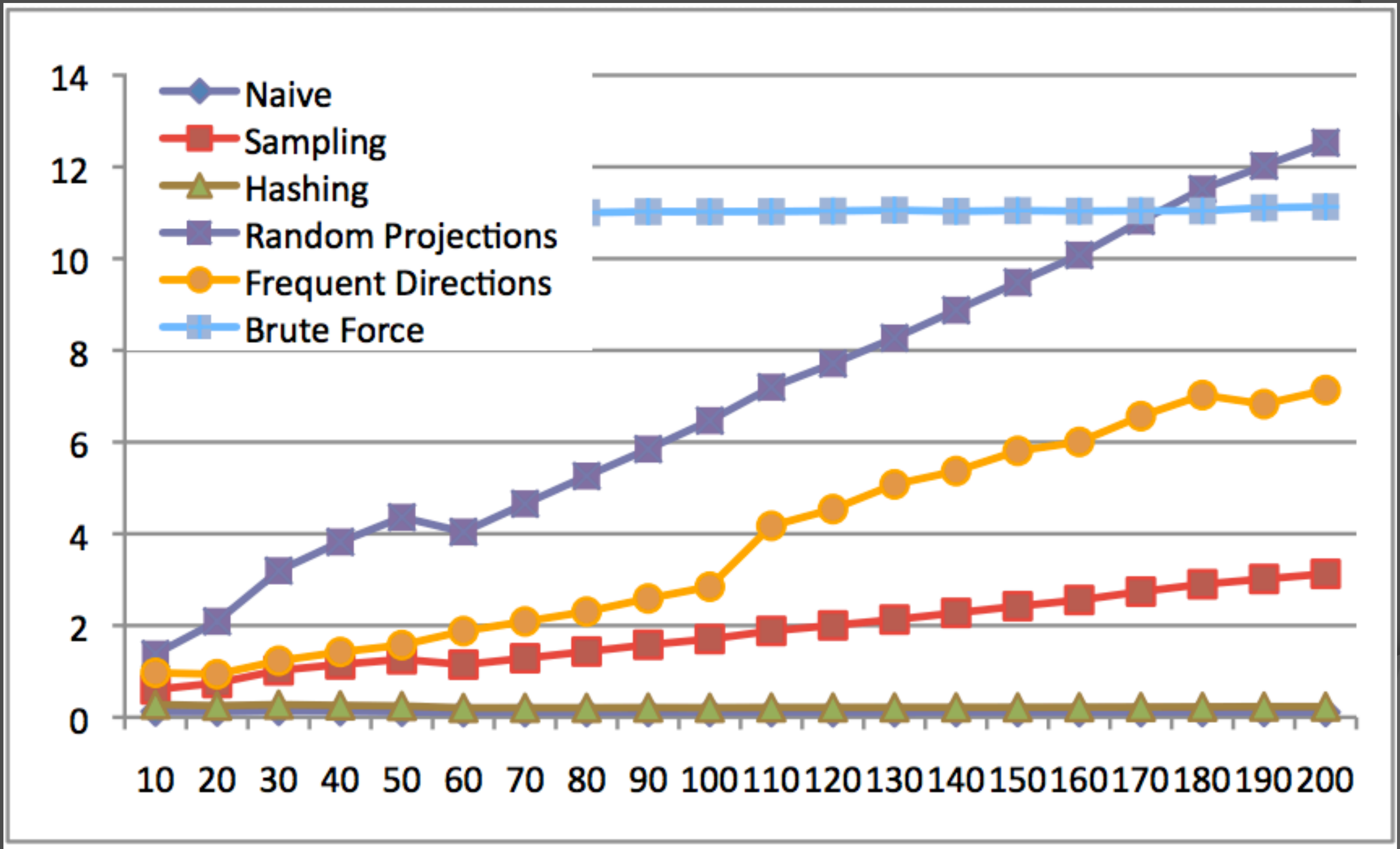
$$O\left(dmn/\mu + \log(\mu)d^2 m\right)$$

- Set $\mu = \Theta\left(\frac{n}{d}\right) = \Theta(\varepsilon n)$ => run time $O\left(\frac{m\log(n)}{\varepsilon^2}\right)$

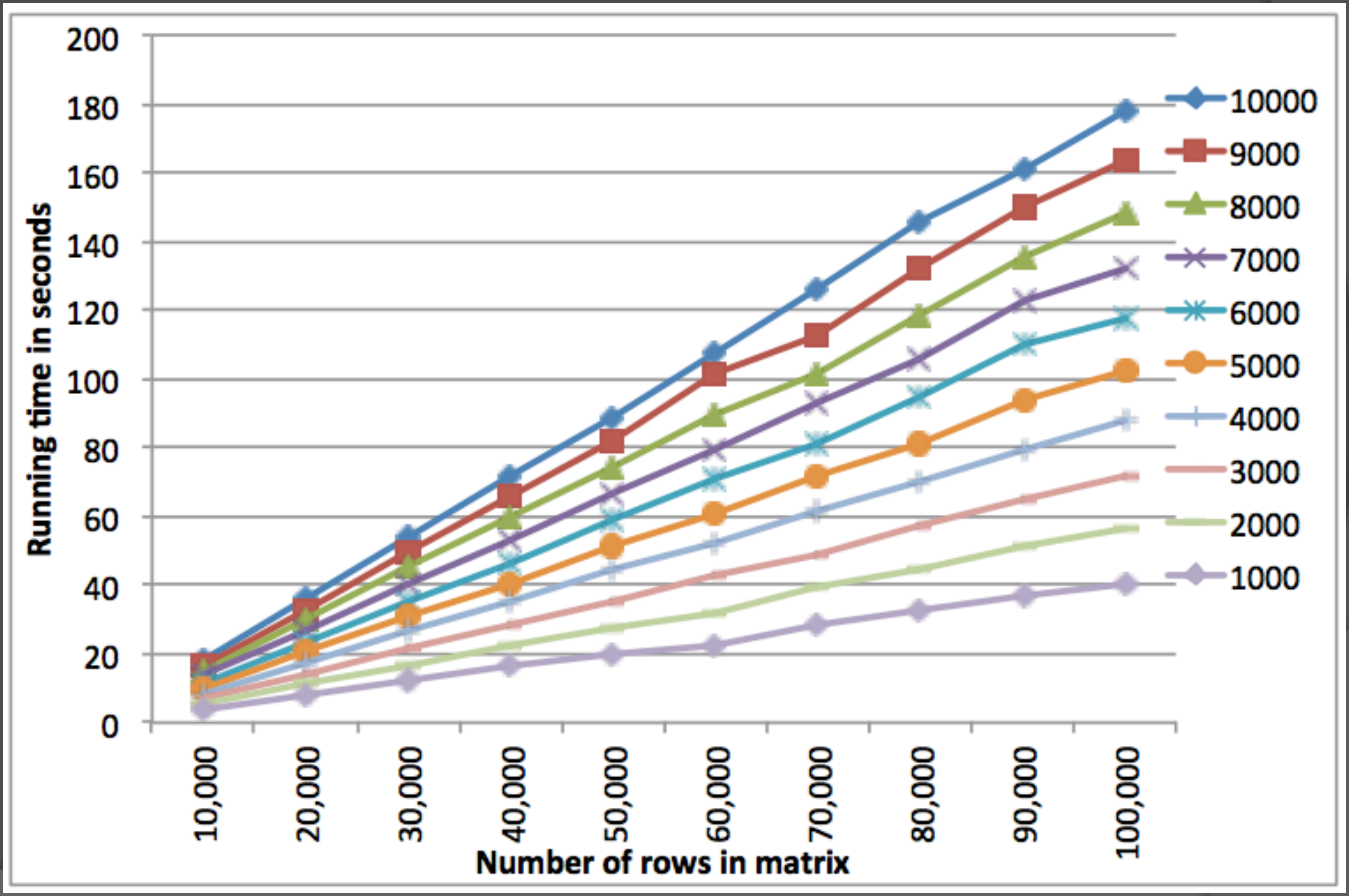
# Results – Accuracy

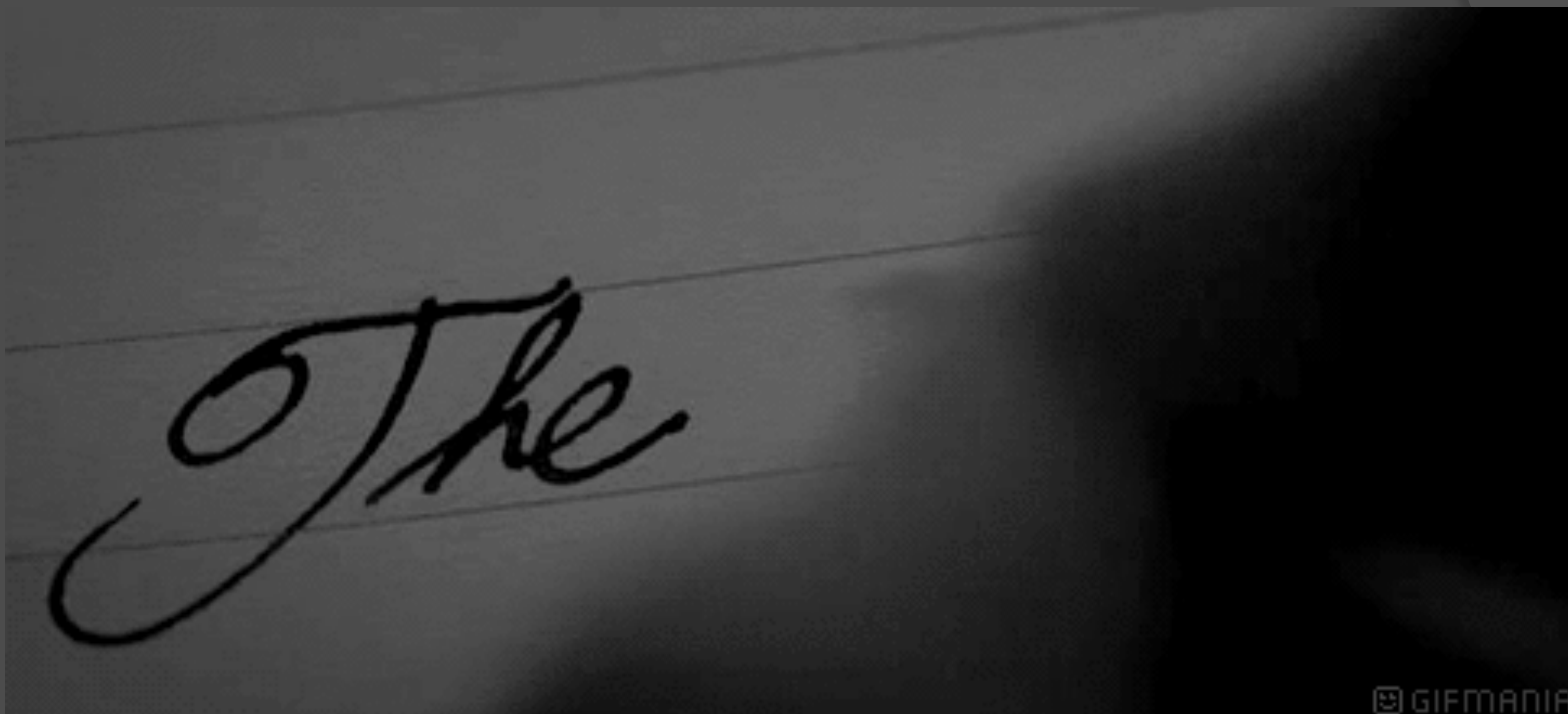


# Results – Run Time vs. Others



# Results – Run Time for FD




The
GIFMANIA