

# GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Iowa State University

September 29, 2013

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Outline

## Biological background

- DNA and RNA

- Central dogma

- Examples of gene regulation

- RNA-seq

- Hybrid vigor

## The model

## The Gibbs sampler

- Gibbs steps

- Estimated heterosis probabilities

- GPU parallelism

## The software

GPU-parallel Gibbs  
sampling of a  
hierarchical model  
for hybrid vigor in  
RNA-seq  
experiments

Will Landau

Biological  
background

- DNA and RNA

- Central dogma

- Examples of gene  
regulation

- RNA-seq

- Hybrid vigor

The model

The Gibbs sampler

- Gibbs steps

- Estimated heterosis  
probabilities

- GPU parallelism

The software

# Outline

## Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

## The model

## The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

## The software

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

## Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

## The model

## The Gibbs sampler

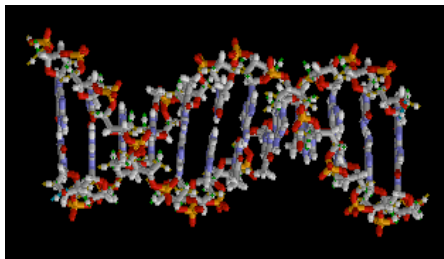
Gibbs steps

Estimated heterosis probabilities

GPU parallelism

## The software

# DNA



```

... — GTGCATCTGACTCCTGAGGAGAAG — ...
... — CACGTAGACTGAGGACTCCTCTTC — ...
  
```

GPU-parallel Gibbs  
sampling of a  
hierarchical model  
for hybrid vigor in  
RNA-seq  
experiments

Will Landau

Biological  
background

DNA and RNA

Central dogma

Examples of gene  
regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

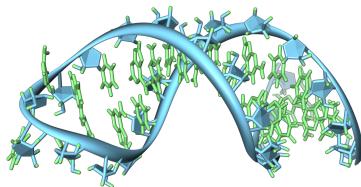
Gibbs steps

Estimated heterosis  
probabilities

GPU parallelism

The software

# RNA



... — GUGCAUCUGACUCCUGAGGAGAAG — ...

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

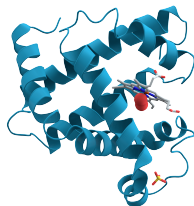
Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Proteins



GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Central dogma: how organisms make proteins

... GTGCATCTGACTCCTGAGGAGAAG ...  
... CACGTAGACTGAGGACTCCTCTTC ... DNA

GPU-parallel Gibbs  
sampling of a  
hierarchical model  
for hybrid vigor in  
RNA-seq  
experiments

Will Landau

Biological  
background  
DNA and RNA  
Central dogma  
Examples of gene  
regulation  
RNA-seq  
Hybrid vigor

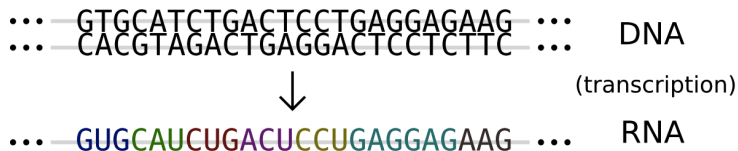
The model

The Gibbs sampler

Gibbs steps  
Estimated heterosis  
probabilities  
GPU parallelism

The software

# Central dogma: how organisms make proteins



GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background  
DNA and RNA  
Central dogma  
Examples of gene regulation  
RNA-seq  
Hybrid vigor

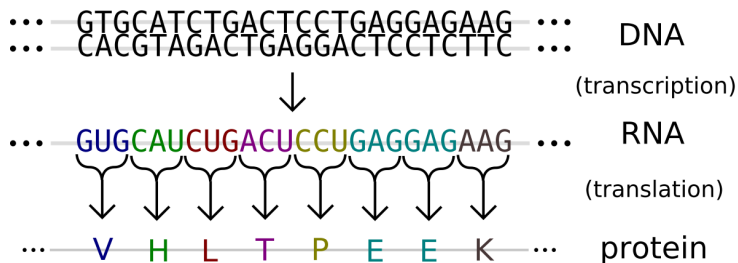
The model

The Gibbs sampler  
Gibbs steps  
Estimated heterosis probabilities  
GPU parallelism

The software



# Central dogma of genetics



GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background  
DNA and RNA  
Central dogma  
Examples of gene regulation  
RNA-seq  
Hybrid vigor

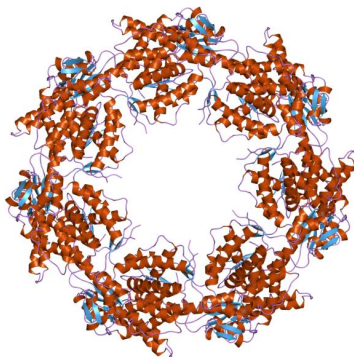
The model

The Gibbs sampler  
Gibbs steps  
Estimated heterosis probabilities  
GPU parallelism

The software

# HSP60

- ▶ HSP = heat shock protein.
- ▶ Prevent heat damage to other proteins.



GPU-parallel Gibbs  
sampling of a  
hierarchical model  
for hybrid vigor in  
RNA-seq  
experiments

Will Landau

Biological  
background

DNA and RNA

Central dogma

Examples of gene  
regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

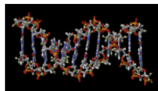
Estimated heterosis  
probabilities

GPU parallelism

The software

# Temperature spike triggers HSP60 production.

HSP60 Gene



GPU-parallel Gibbs  
sampling of a  
hierarchical model  
for hybrid vigor in  
RNA-seq  
experiments

Will Landau

Biological  
background  
DNA and RNA  
Central dogma  
Examples of gene  
regulation  
RNA-seq  
Hybrid vigor

The model

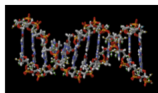
The Gibbs sampler

Gibbs steps  
Estimated heterosis  
probabilities  
GPU parallelism

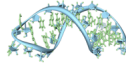
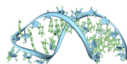
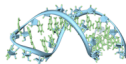
The software

# Temperature spike causes HSP60 expression.

HSP60 Gene



HSP60 RNA



...

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

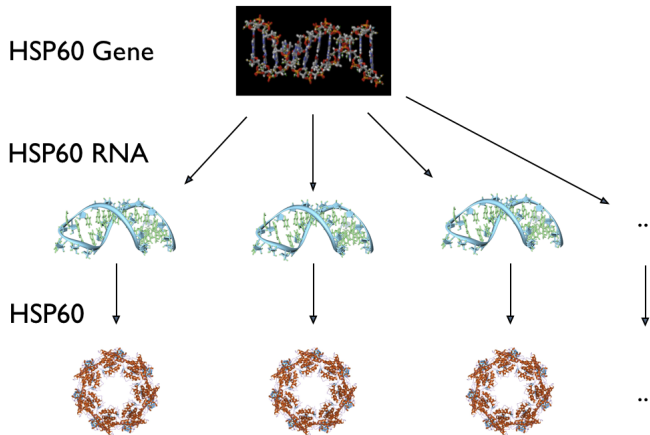
Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Temperature spike causes HSP60 expression.



GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

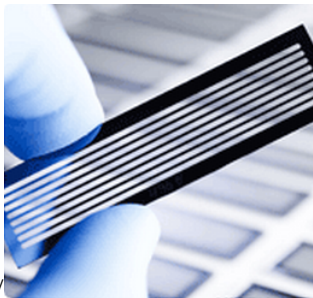
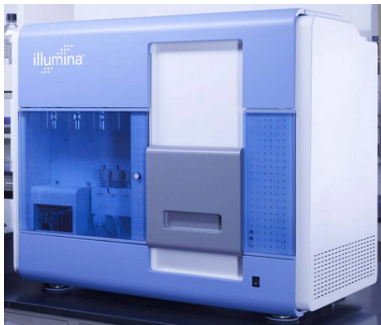
Estimated heterosis probabilities

GPU parallelism

The software

# RNA-seq

- ▶ RNA sequencing: measure gene expression using relative abundance of RNA.
- ▶ Illumina Genome Analyzer:



GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# RNA-seq data: counts of amplified RNA fragments

	Treatment 1			Treatment 2		Treatment 3	
Gene 1	100	225	0	70	279	300	106
Gene 2	0	1	1	50	501	2	7
Gene 3	3	4	2	700	900	0	0
Gene 4	893	400	760	5	5	1000	513
...	...	...	...	...	...	...	...
Gene 34897	10	13	6	819	761	902	912

- Goal: use RNA-seq to study hybrid vigor (heterosis).

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# High-parent heterosis: child's trait surpasses both parents

Parent 1



Parent 2



Child



GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software



# Low-parent heterosis: child's trait is weaker than in each parent

Parent 1



Parent 2



Child



GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Mid-parent heterosis: child's trait is different than average of parents

Parent 1



Parent 2



Child



GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis

probabilities

GPU parallelism

The software

# High-parent heterosis in gene expression

	Parent 1			Child		Parent 2	
Gene 1	100	225	0	70	279	300	106
Gene 2	0	1	1	50	501	2	7
Gene 3	3	4	2	700	900	0	0
Gene 4	893	400	760	5	5	1000	513
...	...	...	...	...	...	...	...
Gene 34897	10	13	6	819	761	902	912

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis

probabilities

GPU parallelism

The software

# Low-parent heterosis in gene expression

	Parent 1			Child		Parent 2	
Gene 1	100	225	0	70	279	300	106
Gene 2	0	1	1	50	501	2	7
Gene 3	3	4	2	700	900	0	0
Gene 4	893	400	760	5	5	1000	513
...	...	...	...	...	...	...	...
Gene 34897	10	13	6	819	761	902	912

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Mid-parent heterosis in gene expression

	Parent 1			Child		Parent 2	
Gene 1	100	225	0	70	279	300	106
Gene 2	0	1	1	50	501	2	7
Gene 3	3	4	2	700	900	0	0
Gene 4	893	400	760	5	5	1000	513
...	...	...	...	...	...	...	...
Gene 34897	10	13	6	819	761	902	912

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis

probabilities

GPU parallelism

The software

# Outline

## Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

## The model

### The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

### The software

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# The model

$$y_{g,n} \stackrel{\text{ind}}{\sim} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))$$

$$c_n \stackrel{\text{ind}}{\sim} N(c_n \mid 0, \sigma_c^2)$$

$$\sigma_c \sim U(\sigma_c \mid 0, \sigma_{c0})$$

$$\varepsilon_{g,n} \stackrel{\text{ind}}{\sim} N(\varepsilon_{g,n} \mid 0, \eta_g^2)$$

$$\eta_g^2 \stackrel{\text{ind}}{\sim} \text{Inv-Gamma} \left( \eta_g^2 \mid \text{shape} = \frac{d}{2}, \text{rate} = \frac{d \cdot \tau^2}{2} \right)$$

$$d \sim U(d \mid 0, d_0)$$

$$\tau^2 \sim \text{Gamma}(\tau^2 \mid \text{shape} = a_\tau, \text{rate} = b_\tau)$$

GPU-parallel Gibbs  
sampling of a  
hierarchical model  
for hybrid vigor in  
RNA-seq  
experiments

Will Landau

Biological  
background  
DNA and RNA  
Central dogma  
Examples of gene  
regulation  
RNA-seq  
Hybrid vigor

The model

The Gibbs sampler  
Gibbs steps  
Estimated heterosis  
probabilities  
GPU parallelism

The software

# The model

$$\mu(n, \phi_g, \alpha_g, \delta_g) = \begin{cases} \phi_g - \alpha_g & \text{sample } n \text{ from parent 1} \\ \phi_g + \delta_g & \text{sample } n \text{ from child} \\ \phi_g + \alpha_g & \text{sample } n \text{ from parent 3} \end{cases}$$

$$y_{g,n} \stackrel{\text{ind}}{\sim} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))$$

$$\phi_g \stackrel{\text{ind}}{\sim} N(\phi_g \mid \theta_\phi, \sigma_\phi^2)$$

$$\theta_\phi \sim N(\theta_\phi \mid 0, \gamma_\phi^2)$$

$$\sigma_\phi \sim U(\sigma_\phi \mid 0, \sigma_{\phi 0})$$

$$\alpha_g \stackrel{\text{ind}}{\sim} \pi_\alpha^{1-I(\alpha_g)} [(1 - \pi_\alpha)N(\alpha_g \mid \theta_\alpha, \sigma_\alpha^2)]^{I(\alpha_g)}$$

$$\theta_\alpha \sim N(\theta_\alpha \mid 0, \gamma_\alpha^2)$$

$$\sigma_\alpha \sim U(\sigma_\alpha \mid 0, \sigma_{\alpha 0})$$

$$\pi_\alpha \sim \text{Beta}(\pi_\alpha \mid a_\alpha, b_\alpha)$$

$$\delta_g \stackrel{\text{ind}}{\sim} \pi_\delta^{1-I(\delta_g)} [(1 - \pi_\delta)N(\delta_g \mid \theta_\delta, \sigma_\delta^2)]^{I(\delta_g)}$$

$$\theta_\delta \sim N(\theta_\delta \mid 0, \gamma_\delta^2)$$

$$\sigma_\delta \sim U(\sigma_\delta \mid 0, \sigma_{\delta 0})$$

$$\pi_\delta \sim \text{Beta}(\pi_\delta \mid a_\delta, b_\delta)$$

GPU-parallel Gibbs  
sampling of a  
hierarchical model  
for hybrid vigor in  
RNA-seq  
experiments

Will Landau

Biological  
background

DNA and RNA  
Central dogma  
Examples of gene  
regulation  
RNA-seq  
Hybrid vigor

The model

The Gibbs sampler

Gibbs steps  
Estimated heterosis  
probabilities  
GPU parallelism

The software



# Outline

## Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

## The model

## The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

## The software

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

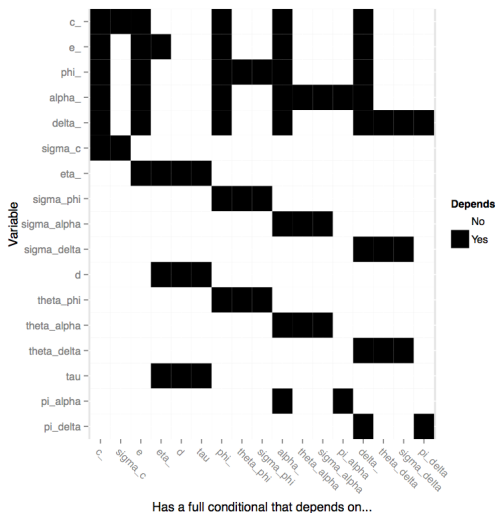
Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Partition parameters by conditional independence.



GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Use these partitions as Gibbs steps.

$$y_{g,n} \stackrel{\text{ind}}{\sim} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))$$

- From the appropriate full conditional distributions, sample the following:

1.  $c_1, \dots, c_N$
2.  $\tau, \pi_\alpha, \pi_\delta$
3.  $d, \theta_\phi, \theta_\alpha, \theta_\delta$
4.  $\sigma_c, \sigma_\phi, \sigma_\alpha, \sigma_\delta, \eta_1^2, \dots, \eta_G^2$
5.  $\varepsilon_{1,1}, \varepsilon_{1,2}, \dots, \varepsilon_{1,N}, \varepsilon_{2,N}, \dots, \varepsilon_{G,N}$
6.  $\phi_1, \dots, \phi_G$
7.  $\alpha_1, \dots, \alpha_G$
8.  $\delta_1, \dots, \delta_G$

- and then repeat.

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Estimated heterosis probabilities

$$\mu(n, \phi_g, \alpha_g, \delta_g) = \begin{cases} \phi_g - \alpha_g & \text{sample } n \text{ from parent 1} \\ \phi_g + \delta_g & \text{sample } n \text{ from child} \\ \phi_g + \alpha_g & \text{sample } n \text{ from parent 3} \end{cases}$$

$$y_{g,n} \stackrel{\text{ind}}{\sim} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))$$

Consider one chain with  $M$  iterations.

$$P(\text{high-parent heterosis in gene } g) \approx \frac{1}{M} \sum_{i=1}^M I(\delta_g^{(i)} > |\alpha_g^{(i)}|)$$

$$P(\text{low-parent heterosis in gene } g) \approx \frac{1}{M} \sum_{i=1}^M I(\delta_g^{(i)} < -|\alpha_g^{(i)}|)$$

$$P(\text{mid-parent heterosis in gene } g) \approx \frac{1}{M} \sum_{i=1}^M I(\delta_g^{(i)} \neq 0)$$

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Tons of opportunity for GPU parallelism across genes!

$$y_{g,n} \stackrel{\text{ind}}{\sim} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))$$

► Sample in parallel:

- $\phi_g$ 's
- $\alpha_g$ 's
- $\delta_g$ 's
- $\varepsilon_{g,n}$ 's
- $\eta_g$ 's

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Example: $\phi_g$ 's

$$y_{g,n} \stackrel{\text{ind}}{\sim} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))$$

$$\phi_g \stackrel{\text{ind}}{\sim} \text{N}(\phi_g \mid \theta_\phi, \sigma_\phi^2)$$

$$\theta_\phi \sim \text{N}(\theta_\phi \mid 0, \gamma_\phi^2)$$

$$\sigma_\phi \sim \text{U}(\sigma_\phi \mid 0, \sigma_{\phi 0})$$

- Using parallel random walk Metropolis steps, sample the  $\phi_g$ 's from their full conditional distributions,

$$p(\phi_g \mid \dots) \propto \exp \left( \sum_{n=1}^N [y_{g,n} \cdot \mu(n, \phi_g, \alpha_g, \delta_g) - \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g))] - \frac{(\phi_g - \theta_\phi)^2}{2\sigma_\phi^2} \right)$$

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Tons of opportunity for GPU parallelism across genes!

$$y_{g,n} \stackrel{\text{ind}}{\sim} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))$$

- ▶ Use **parallel reductions** to calculate sufficient statistics for:
  - ▶  $c_n$ 's
  - ▶  $\tau, d$
  - ▶  $\theta_\phi, \theta_\alpha, \theta_\delta$
  - ▶  $\sigma_\phi, \sigma_\alpha, \sigma_\delta, \sigma_c$
  - ▶  $\pi_\alpha, \pi_\delta$

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Pairwise summation: an example reduction

- ▶ Let's take the pairwise sum of the vector,

$$(5, 2, -3, 1, 1, 8, 2, 6)$$

using 1 block of 4 threads.

GPU-parallel Gibbs  
sampling of a  
hierarchical model  
for hybrid vigor in  
RNA-seq  
experiments

Will Landau

Biological  
background

DNA and RNA

Central dogma

Examples of gene  
regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

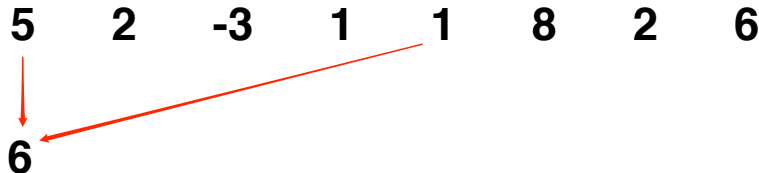
Estimated heterosis  
probabilities

GPU parallelism

The software



# Pairwise summation: an example reduction



**Thread 0**

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

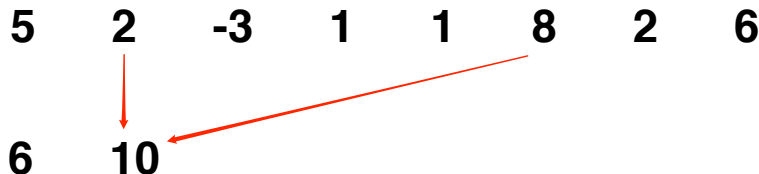
Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Pairwise summation: an example reduction



**Thread 1**

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

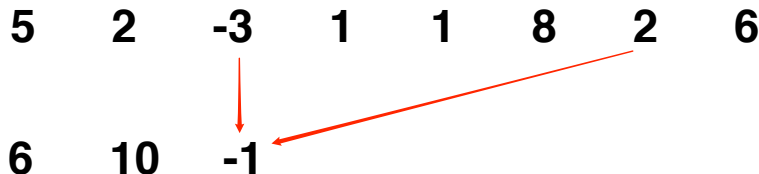
Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Pairwise summation: an example reduction



**Thread 2**

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

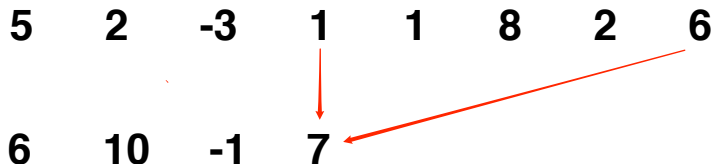
Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Pairwise summation: an example reduction



**Thread 3**

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

## Pairwise summation: an example reduction

5      2      -3      1      1      8      2      6

6      10      -1      7

---

**Synchronize threads**

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Pairwise summation: an example reduction

5      2      -3      1      1      8      2      6

6      10      -1      7

↓

5



**Thread 0**

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Pairwise summation: an example reduction

5      2      -3      1      1      8      2      6

6      10      -1      7

5      17

**Thread 1**

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Pairwise summation: an example reduction

5      2      -3      1      1      8      2      6

6      10      -1      7

5      17

**Synchronize Threads**

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software



# Pairwise summation: an example reduction

5      2      -3      1      1      8      2      6

6      10      -1      7

5      17

**Thread 0**

22

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Example: $\tau^2$

$$y_{g,n} \stackrel{\text{ind}}{\sim} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))$$

$$\varepsilon_{g,n} \stackrel{\text{ind}}{\sim} \text{N}(\varepsilon_{g,n} \mid 0, \eta_g^2)$$

$$\eta_g^2 \stackrel{\text{ind}}{\sim} \text{Inv-Gamma} \left( \eta_g^2 \mid \text{shape} = \frac{d}{2}, \text{rate} = \frac{d \cdot \tau^2}{2} \right)$$

$$d \sim \text{U}(d \mid 0, d_0)$$

$$\tau^2 \sim \text{Gamma}(\tau^2 \mid \text{shape} = a_\tau, \text{rate} = b_\tau)$$

$$p(\tau^2 \mid \dots)$$

$$= \text{Gamma} \left( \tau^2 \mid \text{shape} = a_\tau + \frac{Gd}{2}, \text{rate} = b_\tau + \frac{d}{2} \sum_{g=1}^G \frac{1}{\eta_g^2} \right)$$

- Using a parallel reduction (NVIDIA's CUDA C/C++ Thrust library), calculate the sufficient statistic:

$$\sum_{g=1}^G \frac{1}{\eta_g^2}$$

- Use an efficient rejection sampler to sample  $\tau^2$ .

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Example: $d$

$$y_{g,n} \stackrel{\text{ind}}{\sim} \text{Poisson}(y_{g,n} \mid \exp(c_n + \varepsilon_{g,n} + \mu(n, \phi_g, \alpha_g, \delta_g)))$$

$$\varepsilon_{g,n} \stackrel{\text{ind}}{\sim} \text{N}(\varepsilon_{g,n} \mid 0, \eta_g^2)$$

$$\eta_g^2 \stackrel{\text{ind}}{\sim} \text{Inv-Gamma} \left( \eta_g^2 \mid \text{shape} = \frac{d}{2}, \text{rate} = \frac{d \cdot \tau^2}{2} \right)$$

$$d \sim \text{U}(d \mid 0, d_0)$$

$$\tau^2 \sim \text{Gamma}(\tau^2 \mid \text{shape} = a_\tau, \text{rate} = b_\tau)$$

$$p(d \mid \dots) \propto \Gamma(d/2)^{-G} \left( \frac{d \cdot \tau^2}{2} \right)^{Gd/2} \left( \prod_{g=1}^G \eta_g^2 \right)^{-(d/2+1)} \exp \left( -\frac{d \cdot \tau^2}{2} \sum_{g=1}^G \frac{1}{\eta_g^2} \right) I(0 < d < d_0)$$

- Using parallel reductions (NVIDIA's CUDA C/C++ Thrust library), calculate the sufficient statistics:

$$\prod_{g=1}^G \eta_g^2 \qquad \sum_{g=1}^G \frac{1}{\eta_g^2}$$

- Use a random-walk metropolis step to sample  $d$ .

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# Outline

## Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

## The model

## The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

## The software

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background

DNA and RNA

Central dogma

Examples of gene regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis probabilities

GPU parallelism

The software

# The software

- ▶ Ordinary C and GPU-accelerated versions, along with an R package wrapper, are available for download at <https://github.com/wlandau/heterosis>.
- ▶ Time for a demo...

GPU-parallel Gibbs  
sampling of a  
hierarchical model  
for hybrid vigor in  
RNA-seq  
experiments

Will Landau

Biological  
background

DNA and RNA

Central dogma

Examples of gene  
regulation

RNA-seq

Hybrid vigor

The model

The Gibbs sampler

Gibbs steps

Estimated heterosis  
probabilities

GPU parallelism

The software

# Sources

1. A. Gelman, J. B. Carlin, H. S. Stern, and D. S. Rubin. Bayesian Data Analysis. Chapman & Hall/CRC, 2 edition, 2004.
2. Prof. Jarad Niemi's STAT 544 lecture notes.
3. J. Sanders and E. Kandrot. *CUDA by Example*. Addison-Wesley, 2010.
4. <http://www.astrochem.org/sci/Nucleobases.php>
5. <http://www.biologycorner.com/bio1/DNA.html>
6. <http://www.qualitysilks.com/images/products/artificial-corn-stalk.jpg>
7. <http://en.wikipedia.org/wiki/dna>
8. <http://en.wikipedia.org/wiki/rna>
9. <http://en.wikipedia.org/wiki/HSP60>

GPU-parallel Gibbs sampling of a hierarchical model for hybrid vigor in RNA-seq experiments

Will Landau

Biological background  
DNA and RNA  
Central dogma  
Examples of gene regulation  
RNA-seq  
Hybrid vigor

The model

The Gibbs sampler  
Gibbs steps  
Estimated heterosis probabilities  
GPU parallelism

The software