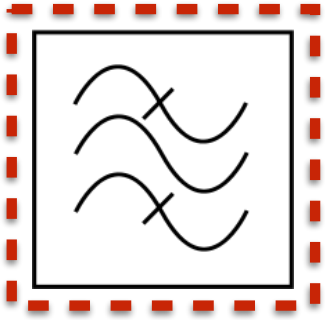
Extracting Posteriors from a Gaussian Mixture Model

Xinkun Nie and Paul Myers

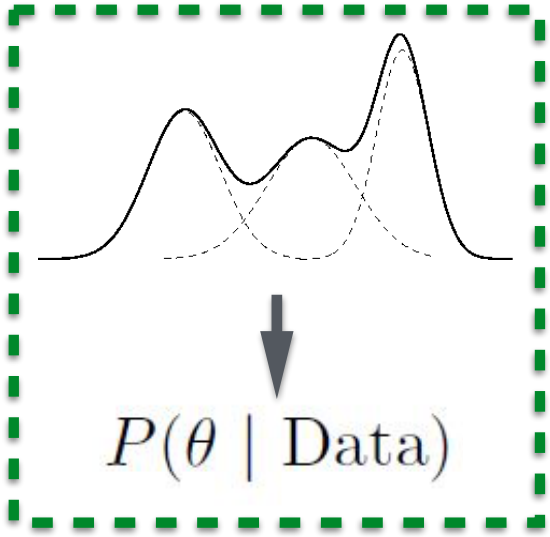
6.375 Final Project

May 11, 2016

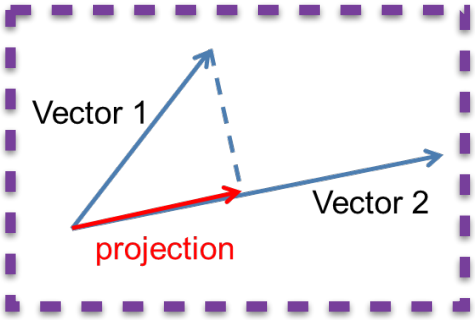
Introduction



Voice Activity Detection



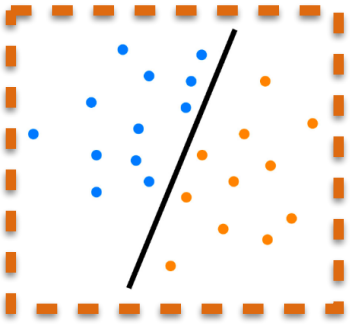
Posterior Extraction from GMMs



Dimensionality Reduction

Raw Speech Vectors

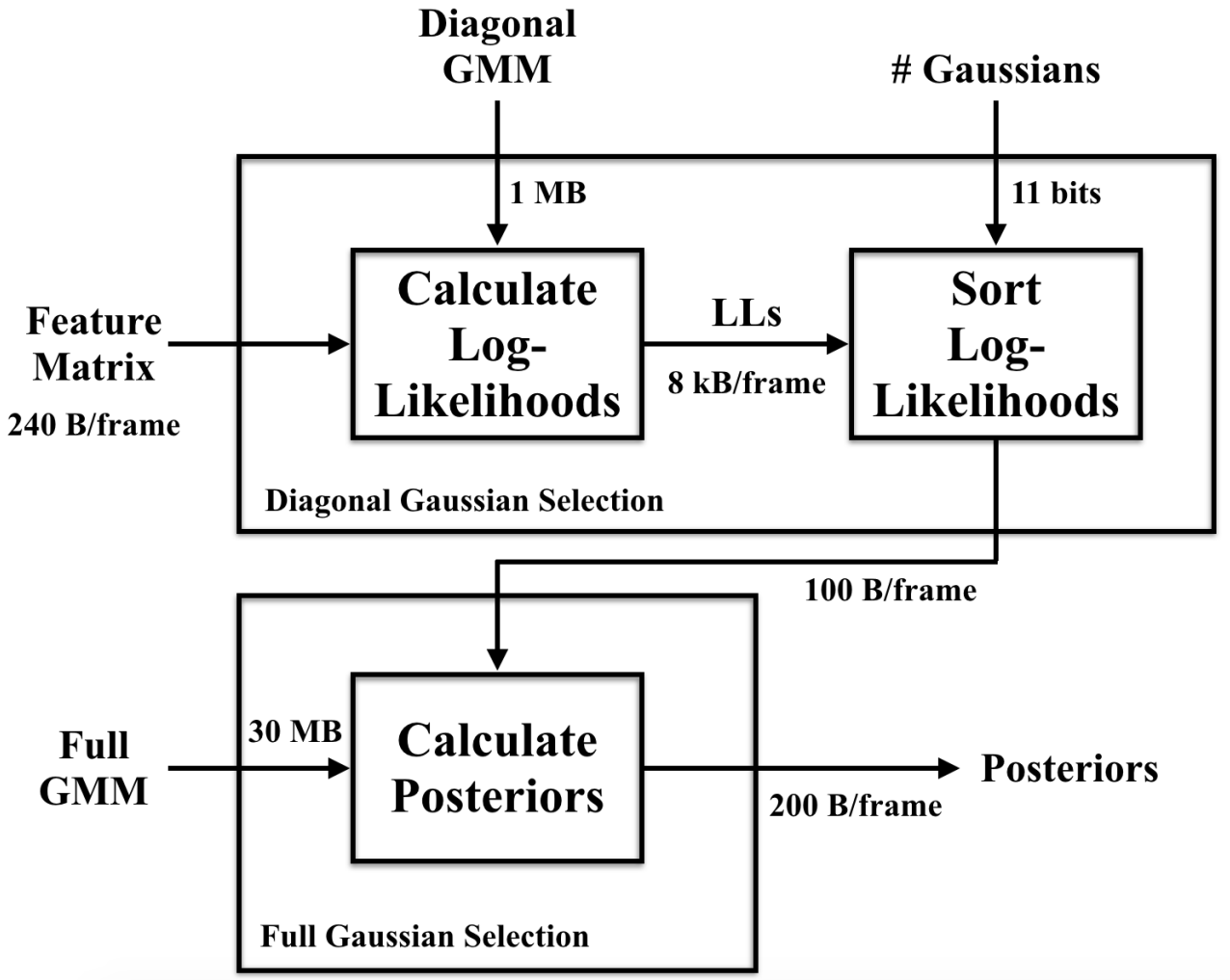
Applications



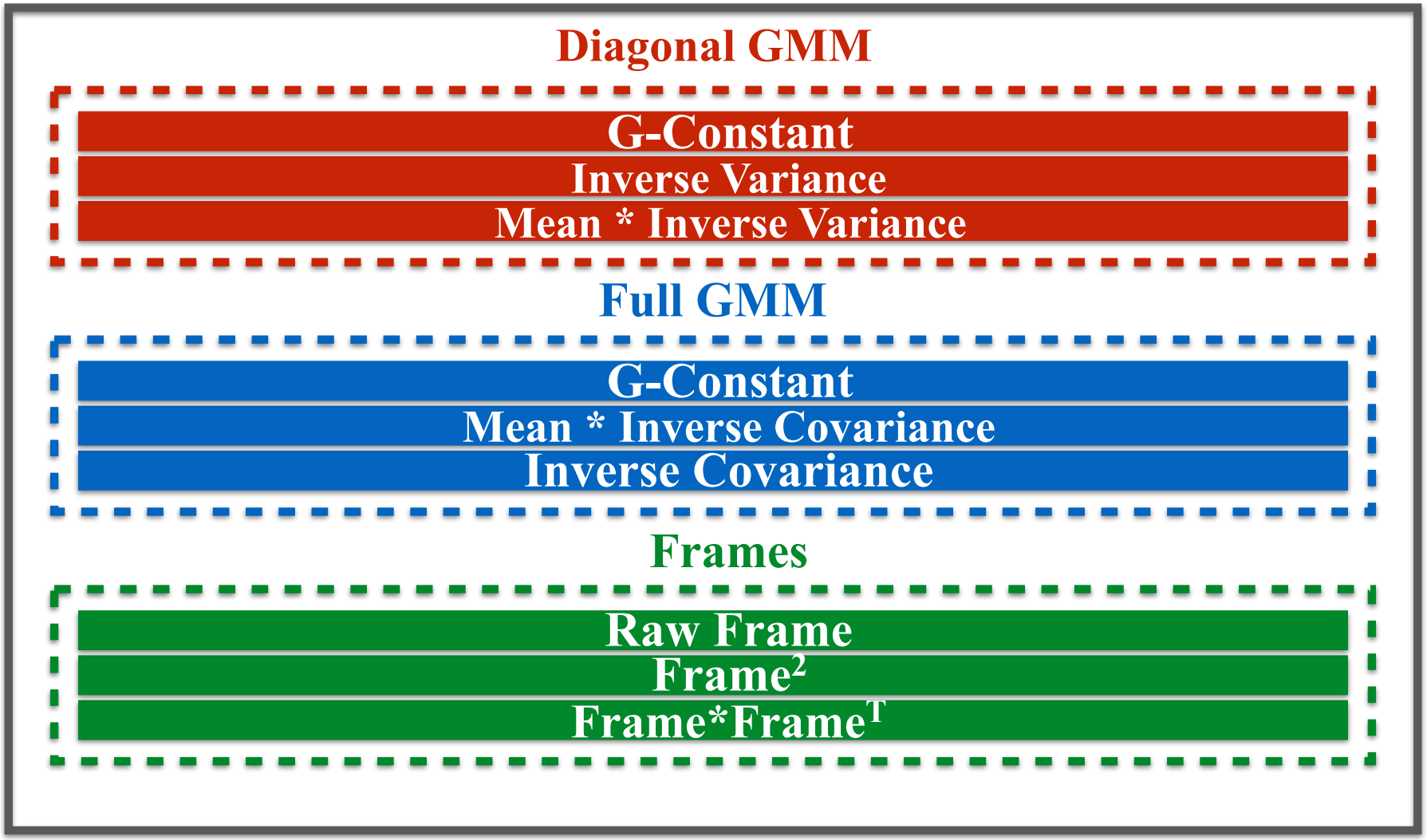
Classification

System Overview

- All computations are performed on 32-bit fixed-point numbers
- GMMs and frame matrix are stored in DRAM

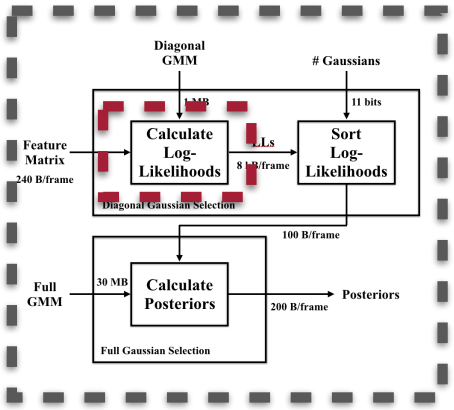


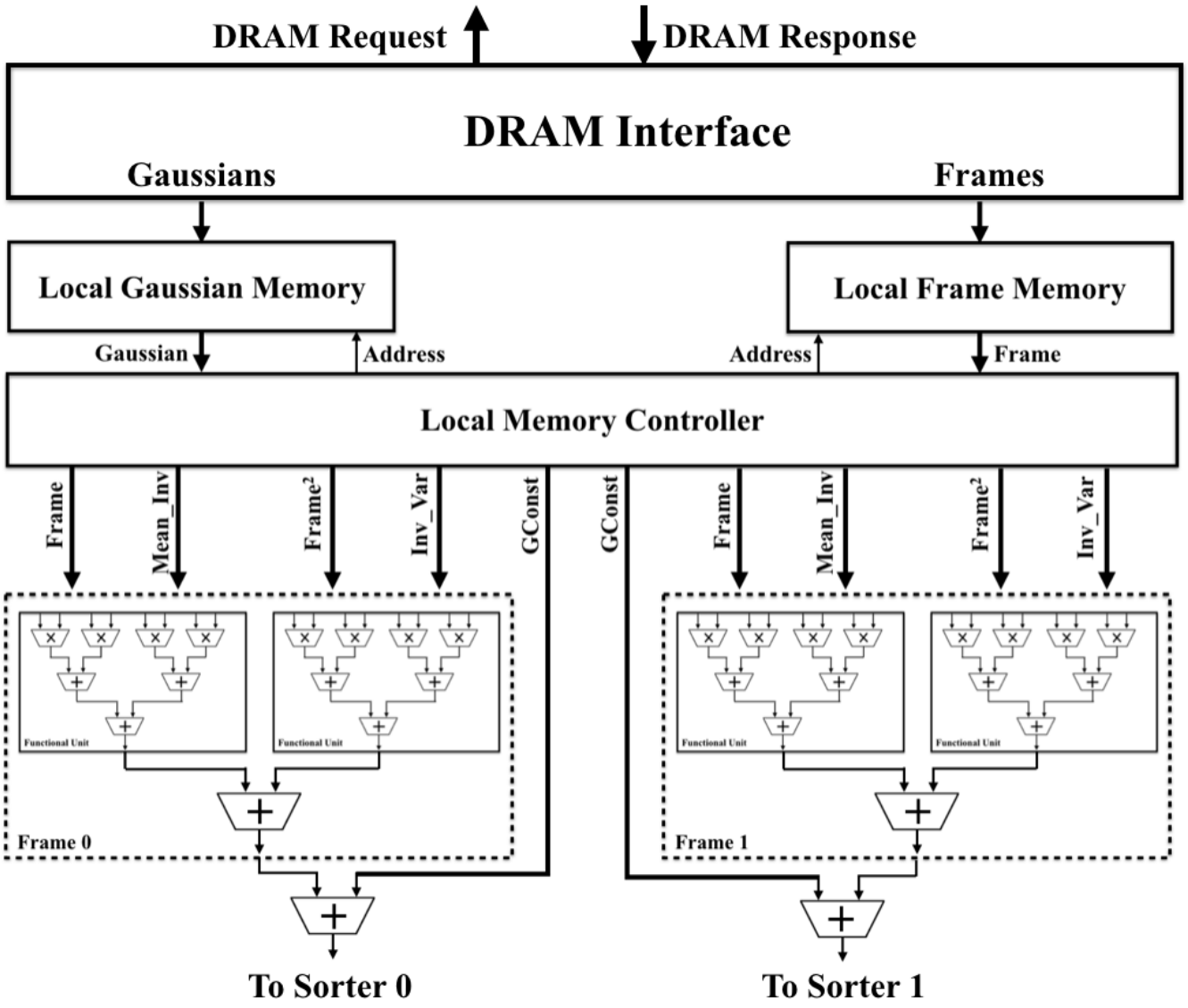
Memory Layout



Diagonal GMM Selection Module

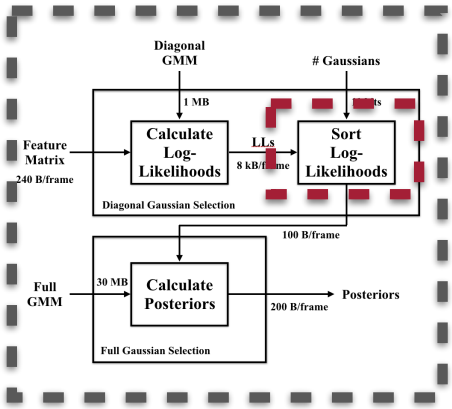
- Computes two dot products
- 60 elements, 32 bits

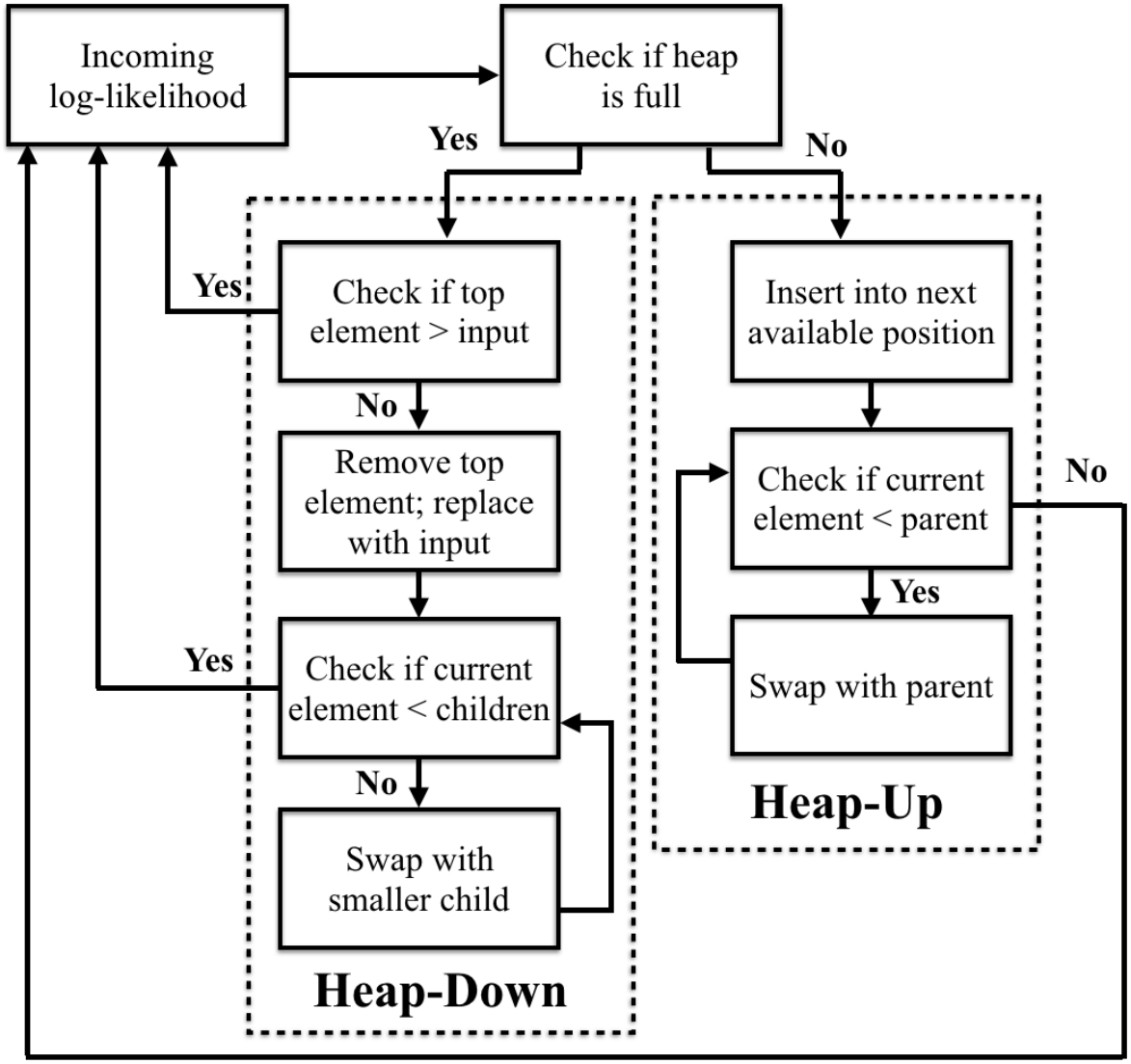




Sorting Module

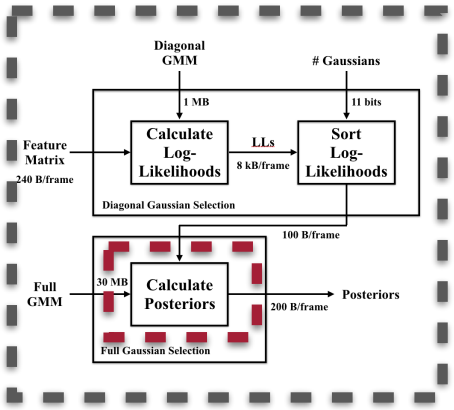
- Chooses *K* largest log-likelihoods

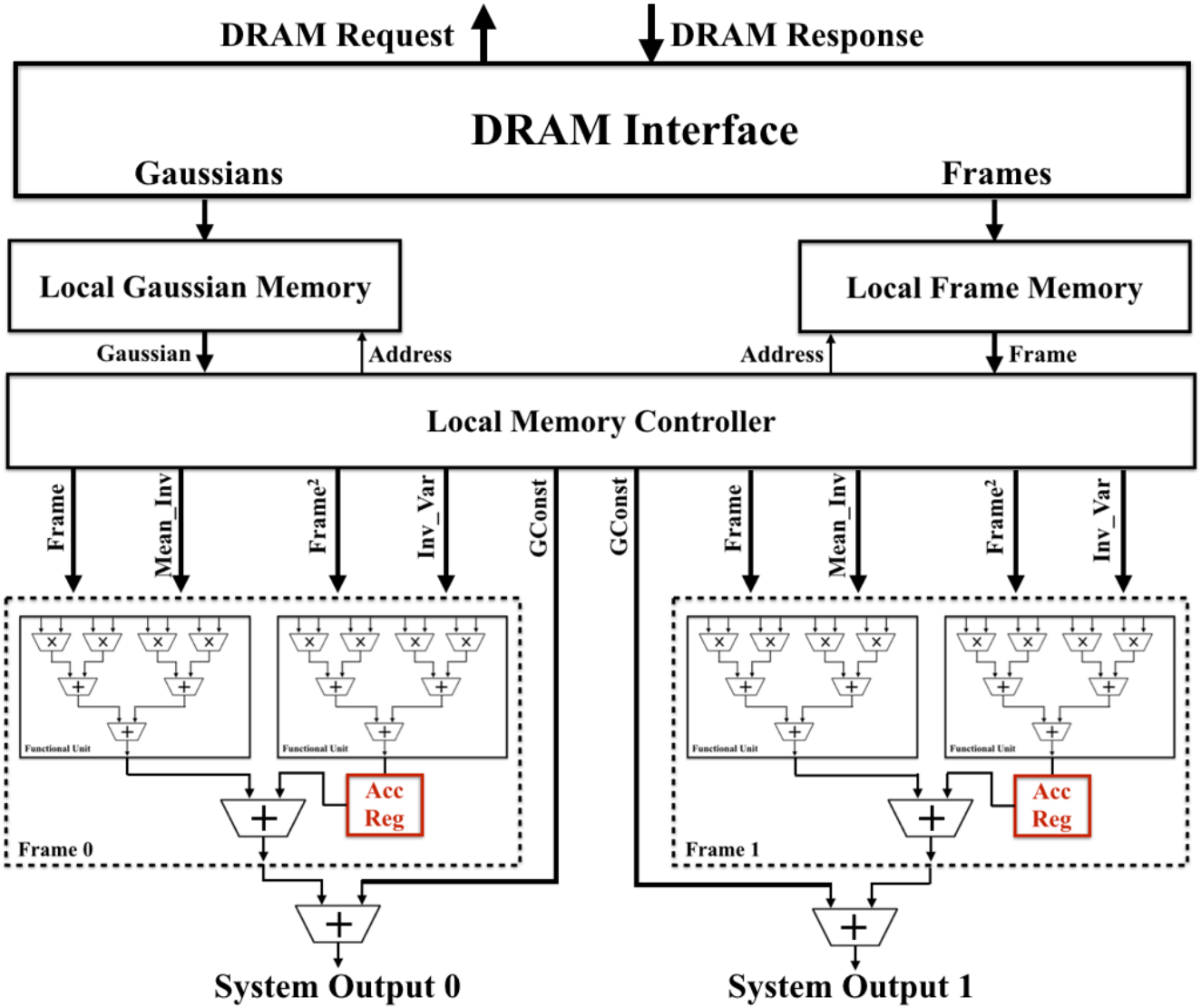




Full GMM Selection Module

- Computes two dot products
- 60 elements, 32 bits
- 1830 elements, 32 bits





Simulation Results

Gaussian ID	Hardware Value	Software Value	Error
0	-156.495	-156.5087	0.00875
4	-161.275	-161.2985	0.0145
5	-195.919	-195.9338	0.00755
6	-181.110	-181.1254	0.00850
8	-192.481	-192.4866	0.00290

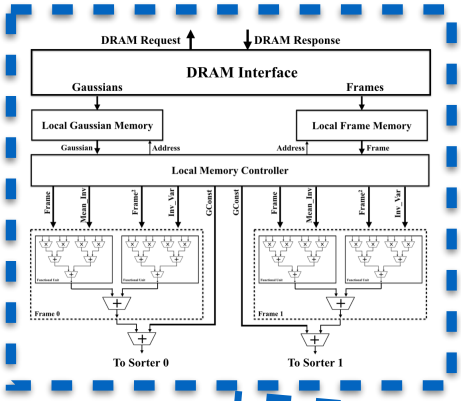
- Setup: 10 Gaussians, 2 frames, 5 selected
- Average error for the frame shown is 0.0085%

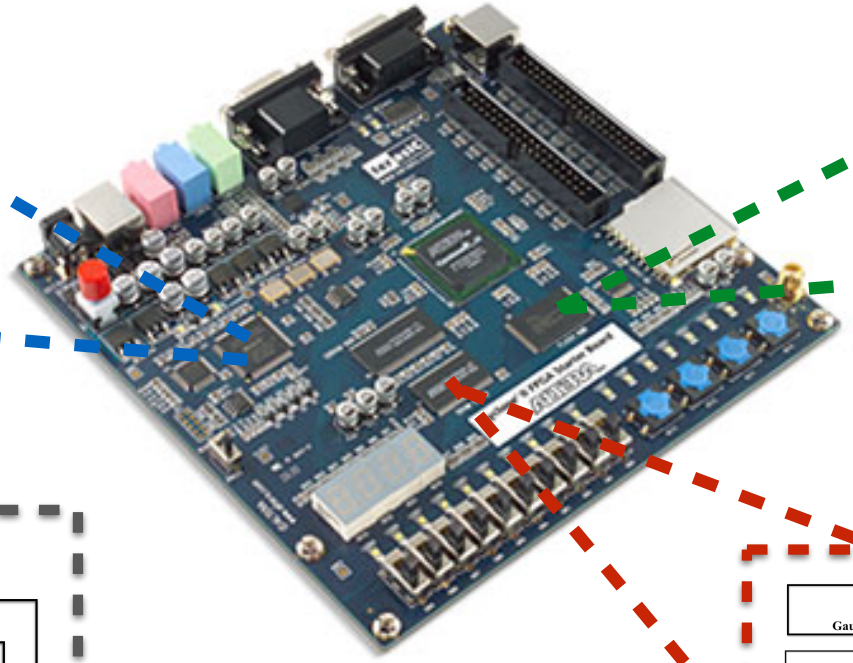
Synthesis Results

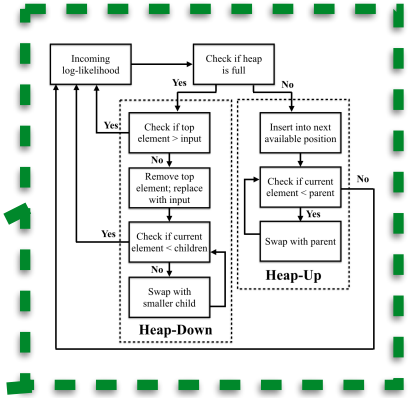
Parameter	Value
Slice LUTs	295450 (97.31% utilization)
Slice Registers	260133 (42.84% utilization)
Block RAM Tiles	354 (34.36% utilization)
DSP Blocks	1920 (68.57% utilization)
Clock Frequency	20 MHz
Critical Path	33.405 ns (In DRAM Control)
Total Negative Slack	0.000 ns
Total Negative Slack Failing Endpoints	277973
Worst Negative Slack	0.077 ns

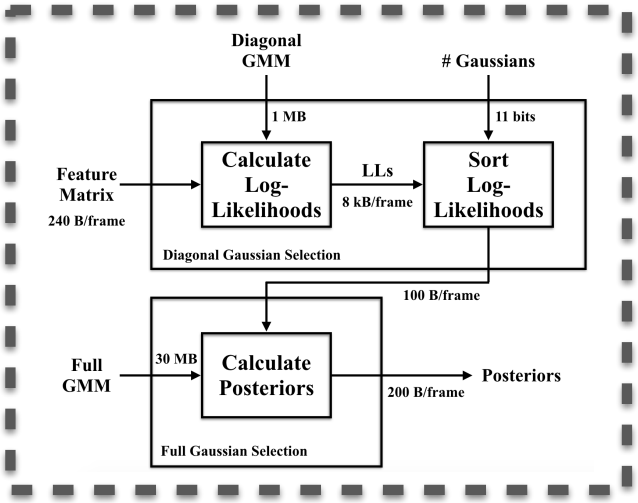
- Area was a significant concern
- Routing difficulties may have led to timing violations
- Could have increased BRAM utilization

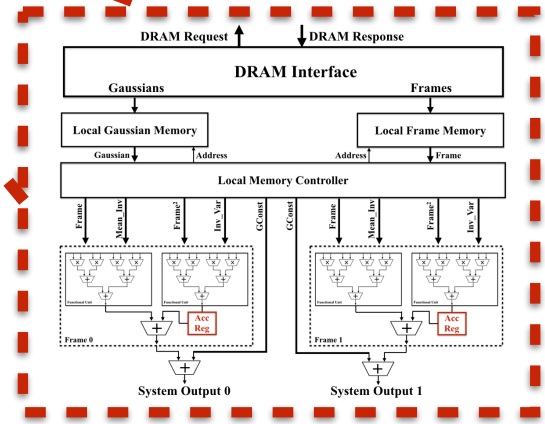
Questions?


DRAM Request
DRAM Response
DRAM Interface
Gaussians
Frames
Local Gaussian Memory
Local Frame Memory
Gaussian
Address
Address
Frame
Local Memory Controller
Frame
Mean_Inv
Frame²
Inv_Var
GConst
Frame 0
Frame 1
To Sorter 0
To Sorter 1




Incoming log-likelihood
Check if heap is full
Yes
No
Check if top element > input
Yes
No
Remove top element; replace with input
Check if current element < children
Yes
No
Swap with smaller child
Heap-Down
Insert into next available position
Check if current element < parent
No
Yes
Swap with parent
Heap-Up


Diagonal GMM
Gaussians
1 MB
11 bits
Feature Matrix
240 B/frame
Calculate Log-Likelihoods
LLs
8 kB/frame
Sort Log-Likelihoods
Diagonal Gaussian Selection
100 B/frame
Full GMM
30 MB
Calculate Posteriors
Posteriors
200 B/frame
Full Gaussian Selection


DRAM Request
DRAM Response
DRAM Interface
Gaussians
Frames
Local Gaussian Memory
Local Frame Memory
Gaussian
Address
Address
Frame
Local Memory Controller
Frame
Mean_Inv
Frame²
Inv_Var
GConst
Acc Reg
Frame 0
Frame 1
System Output 0
System Output 1