

SAXPY:

vec_size	CPU Runtime	GPU runtime	saxpy_gpu	CUDA memcpy DtoH	CUDA memcpy HtoD	cudaMalloc	cudaMemcpy	cudaLaunchKernel
1	8.80E-08		0.00924309	40.66%	29.67%	29.67%	90.83%	2.78%
32	2.08E-07		0.00901491	39.11%	30.73%	30.17%	90.87%	2.99%
128	1.39E-06		0.00885782	38.50%	32.09%	29.41%	90.98%	2.97%
512	1.31E-06		0.00910658	37.50%	29.69%	32.81%	90.89%	2.94%
2048	4.86E-06		0.00890133	32.05%	27.78%	40.17%	90.97%	2.92%
65536	0.000146018		0.00859364	5.01%	28.95	66.04%	92.19%	1.88%
1,048,576	0.002318114		0.0089	0.48%	49.29%	50.23%	89.33%	0.13%
33,554,432	0.074843582		0.0090152	0.50%	44.08%	55.42%	54.58%	41.66%

more vec size: cuda tasks take most runtime

Monte Carlo:

Generate			2048 Reduce		32				
Sample Size	CPU Runtime	GPU runtime	generatePoints	reduceCounts	[CUDA memcpy DtoH]	cudaMalloc	cudaMemcpy	cudaLaunchKernel	
1	8.75E-05		0.099856671	98.53%	1.00%	0.47%	92.68%	0.17%	6.98%
100	0.00407913		0.100374868	98.58%	0.96%	0.45%	92.76%	0.19%	6.87%
1000	0.040648416		0.100910185	98.98%	0.67%	0.35%	92.27%	0.35%	7.22%
100000	4.068836241		0.11969156	99.97%	0.02%	0.01%	76.30%	17.70%	5.85%
1000000	40.66209855		0.309806004	100.00%	0.00%	0.00%	29.61%	68.17%	2.16%

More sample: cudaMemcpy takes most runtime

Sample Size			1000000 Reduce		32			
Generate	GPU runtime	generatePoints	reduceCounts	[CUDA memcpy DtoH]	cudaMalloc	cudaMemcpy	cudaLaunchKernel	
	32	0.135869661	99.98%	0.01%	0.00%	67.82%	27.47%	4.59%
	256	0.161924089	99.99%	0.01%	0.00%	57.14%	38.94%	3.83%
	1024	0.304442852	100.00%	0.00%	0.00%	30.89%	66.59%	2.46%

More generate thread cuda Memcpy takes more runtime

Sample Size		1000000 Generate		1024			
Reduce	GPU runtime	generatePoints	reduceCounts	[CUDA memcpy DtoH]	cudaMalloc	cudaMemcpy	cudaLaunchKernel
32	0.135869661	99.98%	0.01%	0.00%	67.82%	27.47%	4.59%