Two Kernels (1 power + 1 reduce sum) -> 2 copies CPU <-> GPU

		Matrix Size (s)			
		2000x4000	4000x2000	10000x40000	40000x10000
Thread per Block	32	0.193634 + 0.006150 **0.199784**	0.200827 + 0.006302 **0.207129**	0.707466 + 0.249694 **0.95716**	0.710917 + 0.249815 **0.960732**
	64	0.202225 + 0.006194 **0.208419**	0.206941 + 0.005839 **0.21278**	0.718469 + 0.254589 **0.973058**	0.716244 + 0.254347 **0.970591**
	128	0.192099 + 0.006205 **0.198304**	0.191475 + 0.005970 **0.197445**	0.715279 + 0.273626 **0.988905**	0.704865 + 0.254299 **0.959164**

One single Kernel (power + reduce sum) -> Only one copy CPU <-> GPU

		Matrix Size (s)			
		2000x4000	4000x2000	10000x40000	40000x10000
Thread per Block	32	0.209592	0.203447	0.334091	0.332158
	64	0.205405	0.179097	0.329866	0.335435
	128	0.18238	0.177194	0.362994	0.343992

Two Kernels (1 power + 1 reduce sum) -> Only one copy CPU <-> GPU, sharing Power Matrix ref with reduce sum

		Matrix Size (s)			
		2000x4000	4000x2000	10000x40000	40000x10000
Thread per Block	32	0.187509	0.176152	0.329446	0.336244
	64	0.19906	0.20084	0.335179	0.348333
	128	0.19048	0.194846	0.331039	0.322193