Input size (NxN) / Block Size (X,Y)	256x256	512x512	768x768	1024x1024	1280x1280	1536x1536	1792x1792	2048x2048
(16,16)	82	128.74	121.91	130.44	131.62	130.64	131.14	132.47
(32,16)	177.09	241.87	272.43	290.53	284.07	285.32	287.79	292.22
(32,32)	116.79	242.77	275.1	327.01	340.76	341.18	330.28	320.944
(64,16)	189.14	315.58	383.38	392.65	395.70	377.21	373.08	370.41
(64,32)	128.69	426.77	415.77	428.66	433.14	439.14	434.01	420.84
(128,16)	176.35	398.17	393.06	391.75	402.42	399.61	401.1	400.12
(128,32)	134.36	430.36	372.57	445.12	423.63	419.18	421.5	419.27
BLAS with Multicore	9.49	22.5	35.1	41.5	53.2	48.3	56.3	51

Sweet spot : block size (64,32) to (128,32) and input size 1024 to 1792

Figure 1. Matrix multiplication by various sizes

Input size (NxN) / Block size (X,Y)	99	111	128	137	150	178	190	212	237	270	300	350	400	450	550	600	650	700	750	800	850	900	950	1000
(128, 32)	17.73	24.44	37.43	37.22	49.92	71.66	85.74	99.98	108.8	139.7	173	244	294.2	310.2	342.1	290.4	287.6	332.1	350.5	403.2	345.2	331	375.2	411.9

Figure 2. Result for the block size (128,32)