CSE 4351 Notes 1: History, Motivation, and Overview

History of High-Performance Computing (dates based on major publications)

YEAR	GOVERNMENT	ACADEME	INDUSTRY
1955			IBM 704 (5 KFLOPs, Amdahl)
1958			CDC Founded
1964	Energy/Defense Needs	Illiac IV	
1965		Critical Regions (Dijkstra) Multics (MIT/GE/AT&T) FFT	
1967		Data Dependence Defined	Amdahl's Law (IBM)
1970		C.mmp	
1971			Relational Model Proposed (IBM) Intel 4004
1972		Flynn's Taxonomy (SISD, SIMD, MIMD)	Cray leaves CDC to Start Cray Research
1973			Early MRI Work GCD Test (TI)
1974		Brent's Theorem	Tandem Founded
1975		Cm*	
1976			CRAY-1
1977		Conservative PDES	
1978	Ada	PRAM formalized Communicating Sequential Processes	
1979		Garey/Johnson NP-Completeness Book	
1980		Mead & Conway VLSI Book NYU Ultracomputer DOACROSS Randomized Routing Termination Detection	
1981		Dependence Graph Analysis for Vectorization(Illinois) Hypercube Work Commences (Caltech)	DEC VAX
1982			Convex founded Alliant founded
1983	Optimistic PDES (Time Warp)	AKS Sorting Network Snoop Protocols	nCUBE founded Sequent founded Encore founded
1984			Transputers

1985	Japanese 5th Generation Project	Hillis Connection Machine Diss. Linda (Gelernter, Yale) Wormhole Routing (Caltech)	IBM RP3
1986		List Ranking Popular	Commercial Version of X Windowing System
1988		BSP Proposed RAID	
1989			Cray Computer Founded nCUBE/2
1990	PVM		Hypercube Popularity
1991		PRAM Emulation	SUN introduces SMP
1992	Scalable Coherent Interface	Leighton’s Topology Book JaJa’s PRAM Book Omega Test Stanford DASH	
1993	High Perf. FORTRAN MPI ScaLAPACK Released	LogP Proposed Reif’s PRAM Book	CM-5
1995	Accelerated Strategic Computing Initiative announced (SGI, IBM, Intel)	Greenlaw P-Completeness Book NOW	IBM SP-2 CRAY T3E Myrinet
1996	LINPACK TeraFLOP (Intel/Sandia)		
1997	MPI-2 OpenMP HPF-2	CILK Linux Supercomputing	SGI Origin Cray Dies
2000	LINPACK 4.9 TeraFLOPs (IBM/Livermore)		
****	****	****	****

‘‘Challenge’’ Problems

Cosmology

Molecular Biology

Gene Sequencing

Material Science

Data Mining / Warehousing

Image Processing

Climate Modeling

Oil/Seismic Data

Circuit Simulation

Classic Metrics

Input Size = N

Number of Processors = P

Usually bounded by a polynomial in N

$$\text{Speed-Up} = S = \frac{\Gamma}{T} = \frac{T_S}{T_P} = \frac{\text{worst-case time best sequential}}{\text{worst-case time parallel}}$$

Can use either (1) actual time or (2) asymptotic notation

Does not account for number of processors

If speed-up > P , either

(a) better sequential algorithm exists, or

(b) there is hidden processing

Work = W = T*P

Gives basis for discussing efficiency. Work for sequential algorithm is just Γ

(Also called cost by authors that define work as the total number of operations)

$$\text{Efficiency} = E = \frac{\Gamma}{W} = \frac{\Gamma}{T*P} = \frac{S}{P}$$

Desirable that efficiency be θ(1) ("work-optimal"). Efficiency is always O(1).

Amdahl's Law

$$S = \frac{1}{(1-f) + f/P} \leq \frac{1}{1-f}$$

where f is the parallelizable fraction of the original execution time.

If f=1, then S=P. If f=0, then S = 1. If f=0.5, then S = 2P/(P + 1).