LogP Performance Assessment of Fast NIs with Active Messages

David E. Culler
Rich Martin, Lok Tin Liu, Chad Yoshikawa
U.C. Berkeley

culler@cs.berkeley.edu
new slides – http://www.cs.berkeley.edu/~culler/talks

Hot Interconnects III
Aug. 1995

Motivation

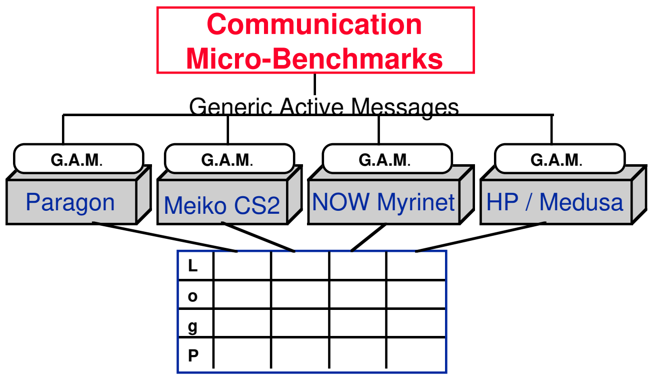
- **Huge space of network interface architectures**

	Controller	Embedded Processor	General Purpose Processor
I/O Bus	Sun ATM	Myrinet LANai FORE SBA-200	IBM SP-2
Graphics Bus	HP Medusa		
Memory Bus	Cray T3D	Meiko CS-2	Paragon

- **Highly idiosyncratic**

=> want a concise, meaningful framework for evaluating design alternatives

Basic Approach



- **Observe a set of GAM "gray box" experiments**
- **Analyze "signatures" to LogP deduce parameters**

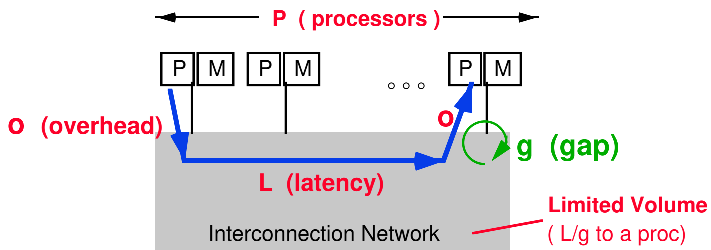
Hot Interconnects III.3

Outline

- **Background**
  - **LogP**
  - **Generic Active Messages**
- **Microbenchmark Method**
- **Short Message Performance**
- **Bulk Transfer Performance**

Hot Interconnects III.4

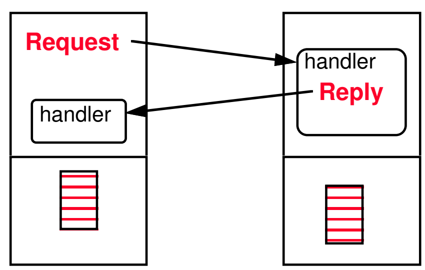
LogP



- **Latency in sending a (small) message between modules**
- **overhead felt by the processor on sending or receiving msg**
- **gap between successive sends or receives** (1/BW)
- **Processors**

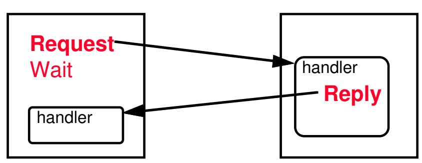
G.A.M.

- **Highly optimized communication layer on a range of HW**
- **Request / Reply small active messages**
- **Bulk-Transfer (store & get)**



Short Message LogP Performance

Time for a Complete Operation



Parameter	Paragon	Meiko CS2	SS20 Myrinet	HP Medusa
Round Trip (μs)	21.3	26.2	32.3	33.8
L + 2o	10.6	13.1	16.1	16.9

Time to Issue a Sequence of Requests

Start Timer

Repeat *M* times

Issue request

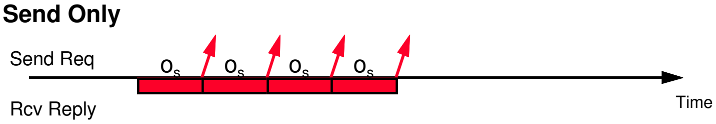
Time(M)

Stop Timer

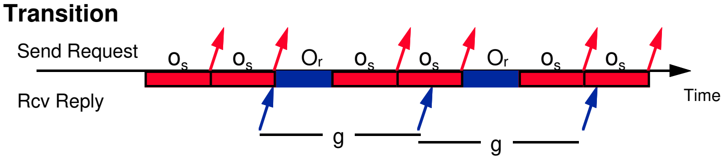
. . . things complete

Time to Issue Sequence

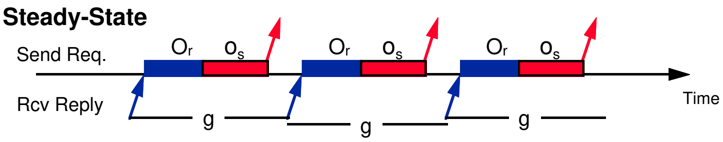
Send Only



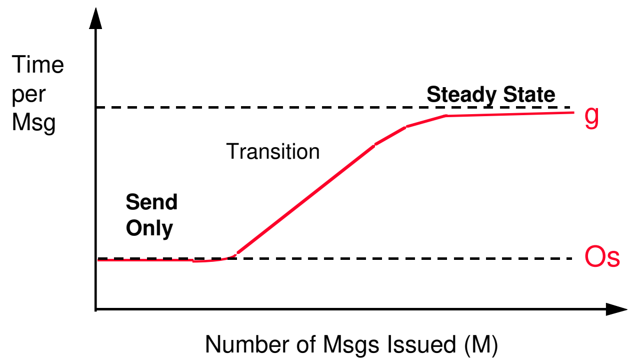
Transition



Steady-State



Analysis of Issue Profile



Add compute interval Δ between requests

For Range of Δ and *M*

Repeat *M* times

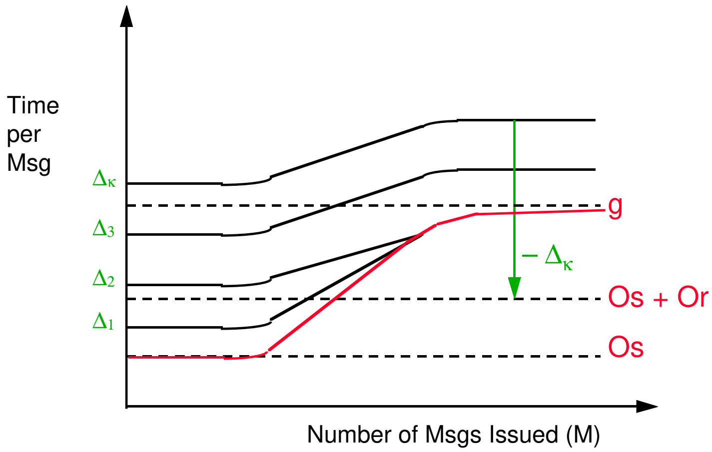
Issue request

Compute for Δ

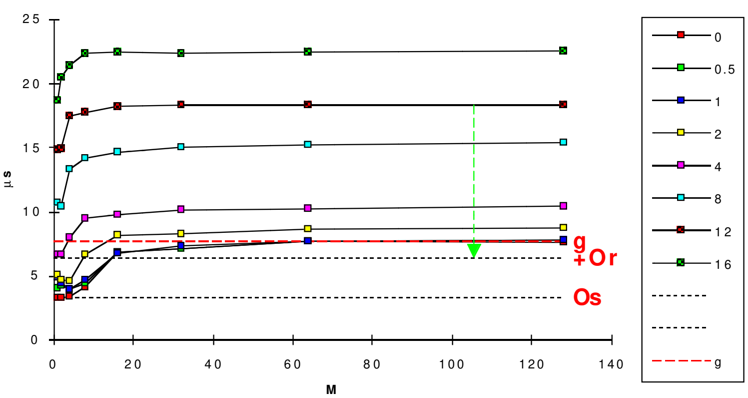
Time(***M***, Δ)

... things complete

Extracting O and Gap from Issue Profile

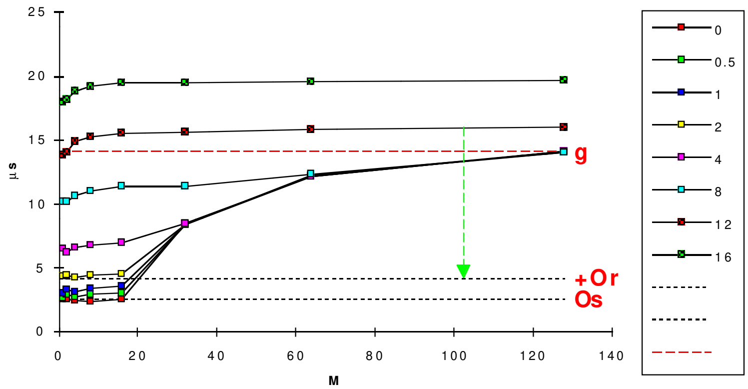


Paragon GAM Issue Profile



L= 3.6 Os= 3.3 Or= 3.1 g= 7.7

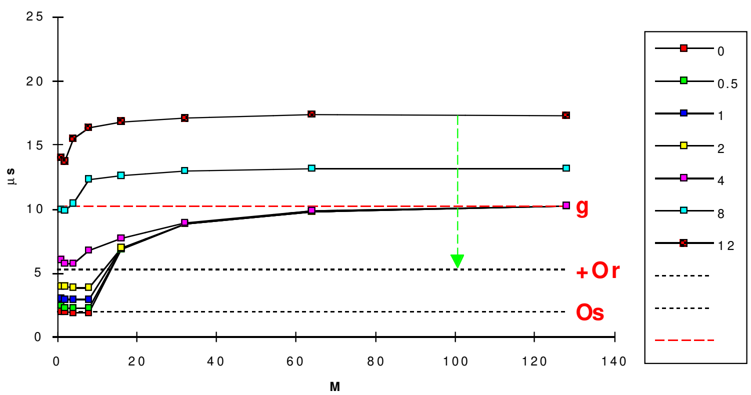
Meiko CS-2 GAM Issue Profile



L= 8.8 Os= 2.6 Or= 1.6 g= 14.1

Hot Interconnects III.15

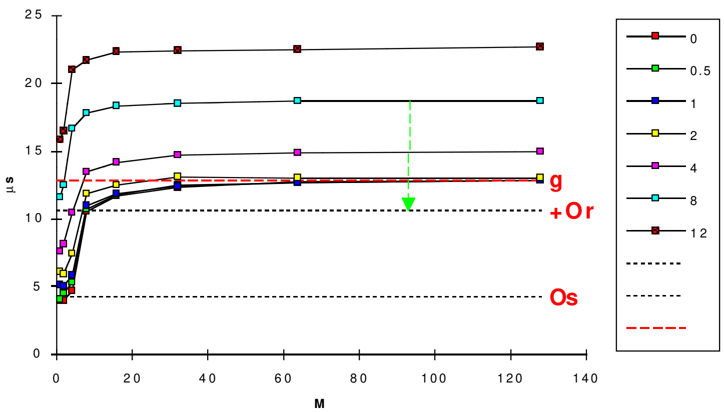
SS20 Myrinet GAM Issue Profile



L= 11.0 Os= 2.0 Or= 3.2 g= 10.3

Hot Interconnects III.16

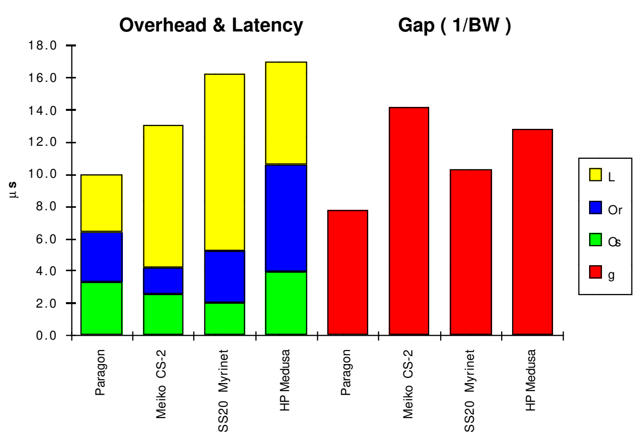
HP Medusa GAM Issue Profile



L = 6.4 Os = 4.0 Or = 6.6 g = 12.8

Hot Interconnects III.17

LogP Summary



Hot Interconnects III.18

Uncontended Validation

- **LogP says 1-1, 1-k, k-k should behave the same as long as no destination contention**

Parameter	Paragon	Meiko CS2	Sun Myrinet	HP Medusa
g (1-1/H-H)	1.00	1.00	1.00	1.00
g (1-1/1-k)	0.96	1.00	1.57	0.99

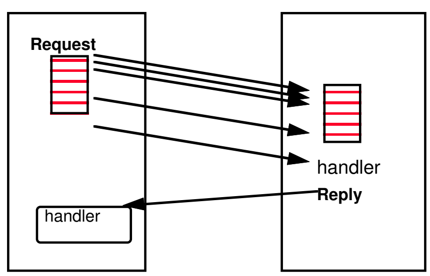
Contended Validation

- **K to 1 should increase g by K**

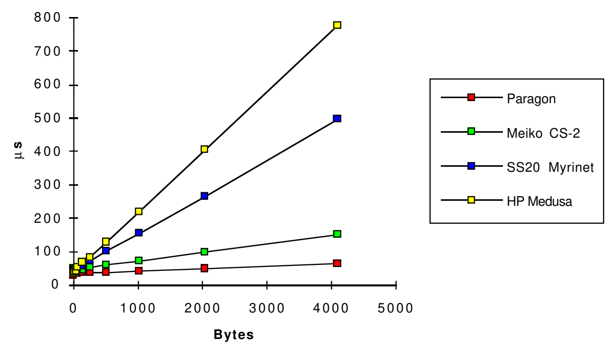
Parameter	Paragon	Meiko CS2	Sun Myrinet	HP Medusa
g (K-1) / g (1-1)	6.0	16.8	3.1	
K	7	15	3	

Bulk Transfer Performance

Time for Complete Transfer Operation

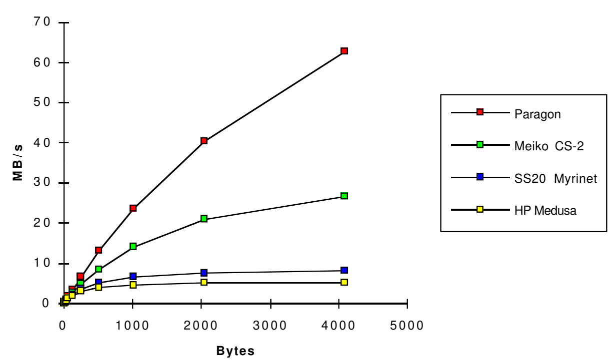


Single Bulk-Transfer Time



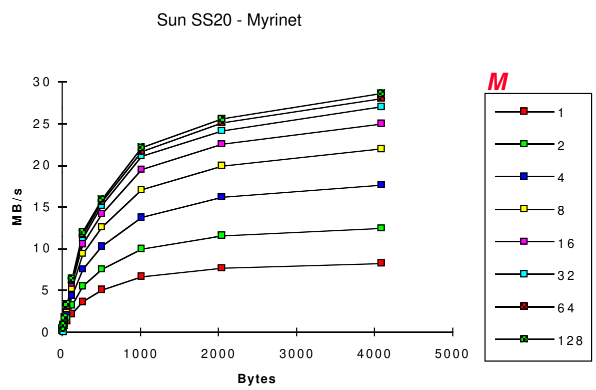
- **Fit to linear model T0 + n/B**
  - **T0 often meaningless**
  - **model is not revealing**

Single Bulk-Transfer Bandwidth



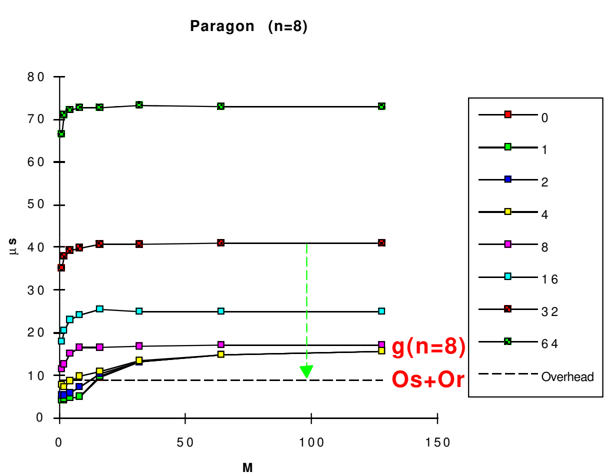
- **BW(n) = n/ T(n) carries no more information**

BW(n) on Sequence of *M* Bulk-Transfers

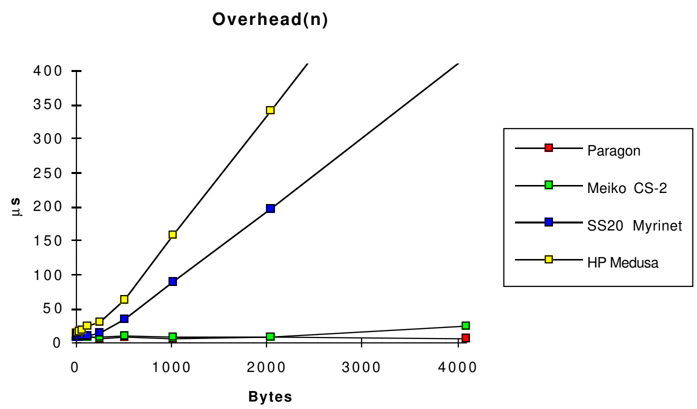


=> Use same µ-benchmark technique to characterize O(n) and g(n)

Overhead & Gap from Bulk Transfer Issue

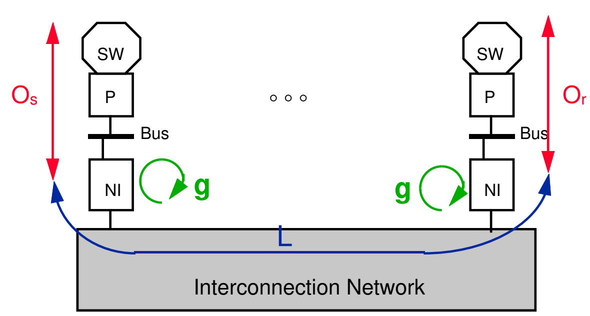


Bulk Issue Overhead Comparison



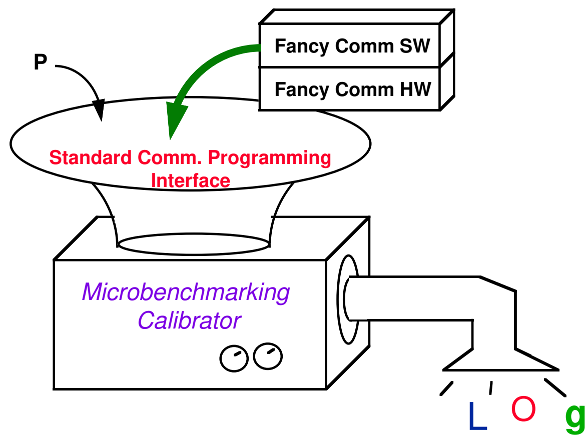
Hot Interconnects III.27

L o g parameters capture HW/SW behavior



Hot Interconnects III.28

Take your block diagrams and . . .



PVM, MPI, UDP, Shared Mem ... Anyone ?

Hot Interconnects III.29