The Measured Network Traffic of Compiler Parallelized Programs

Peter A. Dinda
Northwestern University, Computer Science
http://www.cs.northwestern.edu/~pdinda

Brad M. Garcia
Laurel Networks

Kwok-Shing Leung
Magma Design Automation

Original Study Done At Carnegie Mellon University

Overview

- Analysis of packet traces of representative HPF-like codes running on a shared Ethernet
- Traffic very different from typical models
  - Simple packet size+interarrival behaviors
  - Correlation between flows
  - Periodicity within flows and in aggregate
- Implications for prediction and QoS models

Current focus

Caveats: Data is old, shared media network

Outline

- Why?
- CMU Fx Compiler
- Communication Patterns and programs
- Methodology
- Results
- Implications for prediction and QoS
- Conclusions and future work

Why Study Traffic of Parallel Programs?

- Networking provisioning
  - Source models, Aggregated traffic models
- Adaptive applications
  - Measurement and prediction
  - Network Weather Service, Remos, RPS
- Resource reservation and QoS systems
  - Source models, framing QoS requests
  - Intserv, ATM
- Computational Grids may introduce lots of such traffic
- …

CMU Fx Compiler

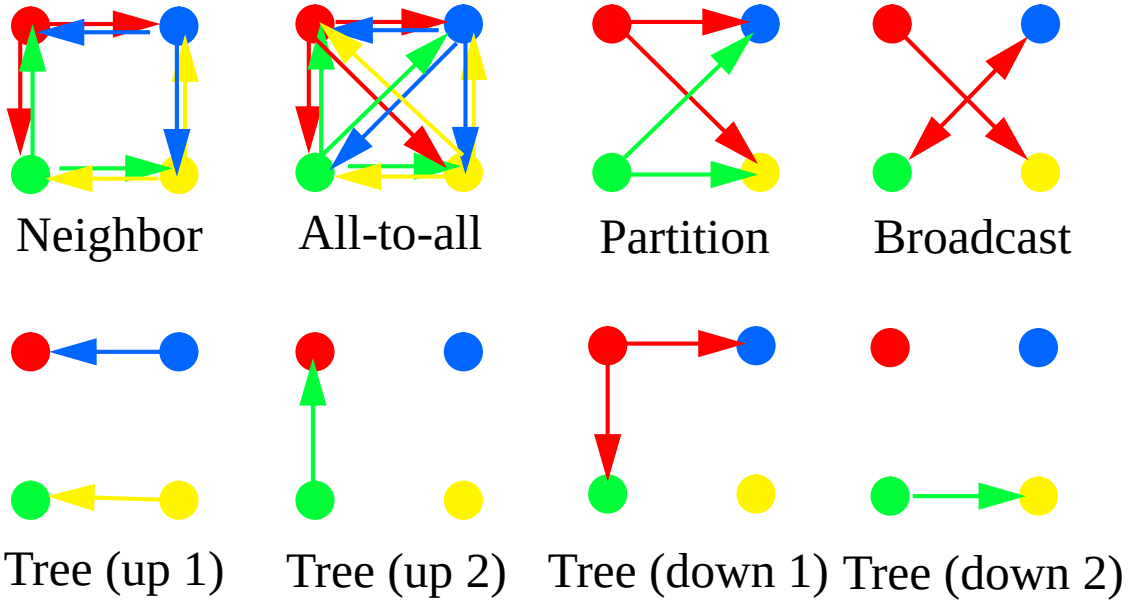
- Variant of High Performance Fortran
- *Both* task and data parallelism
- Sophisticated communication generation
  - Compile-time and run-time
- Multiple target platforms
  - Custom communication back-ends
    - iWarp (original target), Paragon, T3D, T3E, MPI
  - PVM, MPI on many platforms
    - Alpha/DUX, Sun/Solaris, I386/Linux, …

http://www.cs.cmu.edu/~fx

Genesis of Communication Patterns in Fx Programs

- Parallel Array assignment
- Parallel Loop iteration input and output
- Parallel prefix and loop merge
- Inter-task communication
  - Tasks can themselves be task or data parallel
- Parallel I/O
- Distribution of sequential I/O

Communication Patterns in Study



Pattern	Kernel	Description
Neighbor	SOR	2D Successive Overrelaxation
All-to-all	2DFFT	2D Data Parallel FFT
Partition	T2DFFT	2D Task Parallel FFT
Broadcast	SEQ	Sequential I/O
Tree	HIST	2D Image Histogram

Run-time communication

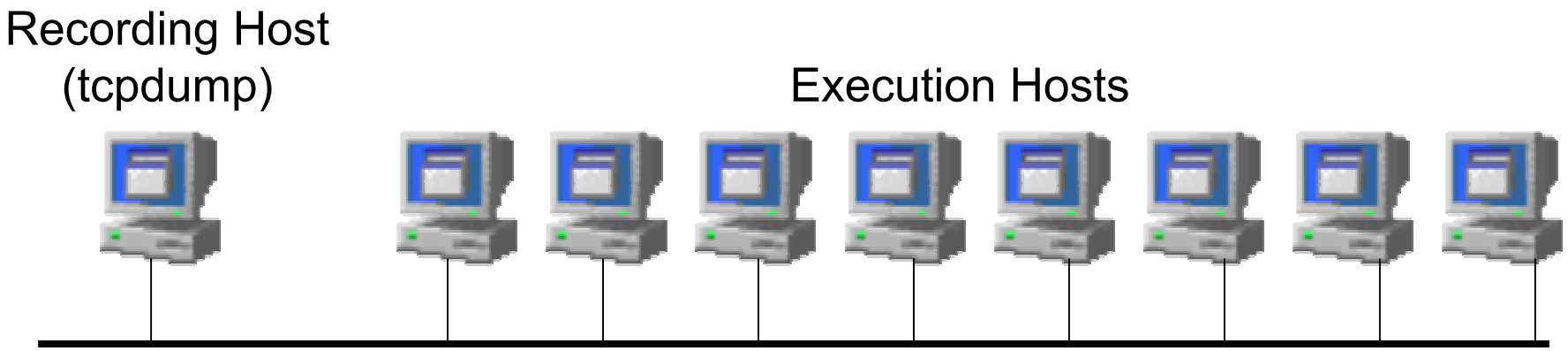
Beyond Kernels: AIRSHED

- Air quality modeling application
  - Used Fx model created by app developers
- Coupled chemistry and wind simulations on 3D array (layers, species, locations)

```
Data input and distribution
Do I=1,h
   Pre-processing
   do j=1,k
      Chemistry/vertical transport
      Distribution transpose
      Horizontal transport
      Distribution transpose
```

Environment

- Nine DEC 3000/400 workstations
  - 21064 at 133 MHz, OSF/1 2.0, DEC's tcpdump
  - Fx 2.2 on PVM 3.3.3, TCP-based communication
- 10 mbps half-duplex shared Ethernet LAN
  - Not private: experiments done 4-5 am



Why is this still interesting?
(study was done ~1996)

- Compiler and run-time still representative
- Communication patterns still common
- Hard to do a study like this today
  - Modern Ethernet is switched
  - Can only see an approximation of the aggregate traffic (SNMP, Remos) not the actual packets
- Artificial synchronization? BSP anyway
- Implications for network prediction and QoS models are important

Methodology

- Run program, collecting all packets
  - Arrival time and size (including all headers)
- Classify packets according to “connections”
  - One-way flow of data between machines
  - All packets from machine A to machine B
  - Includes TCP ACKs for symmetric connection
- Aggregate and per-connection metrics

Metrics

- Packet size distributions
- Packet interarrival time distributions
- Average bandwidth consumed
- Instantaneous bandwidth consumed
  - 10 ms sliding window
  - Time and frequency domain (power spectrum)

Results for 4 node versions of programs

Packet Size Distributions

- Trimodal
  - MTU-sized packets
  - “leftovers” (n byte message modulo MTU size)
  - ACKs for symmetric connection
- T2DFFT generates many different sizes
  - Run-time communication system uses multiple PVM pack calls per message

Typical LAN traffic has wide range of sizes

Packet Interarrival Time Distribution

- Bursty, but only at only a few timescales
  - Most are within a burst: closely spaced
  - Few are between bursts: farther apart
- Quite deterministic on-off sources
  - Synchronized communication phases in Fx

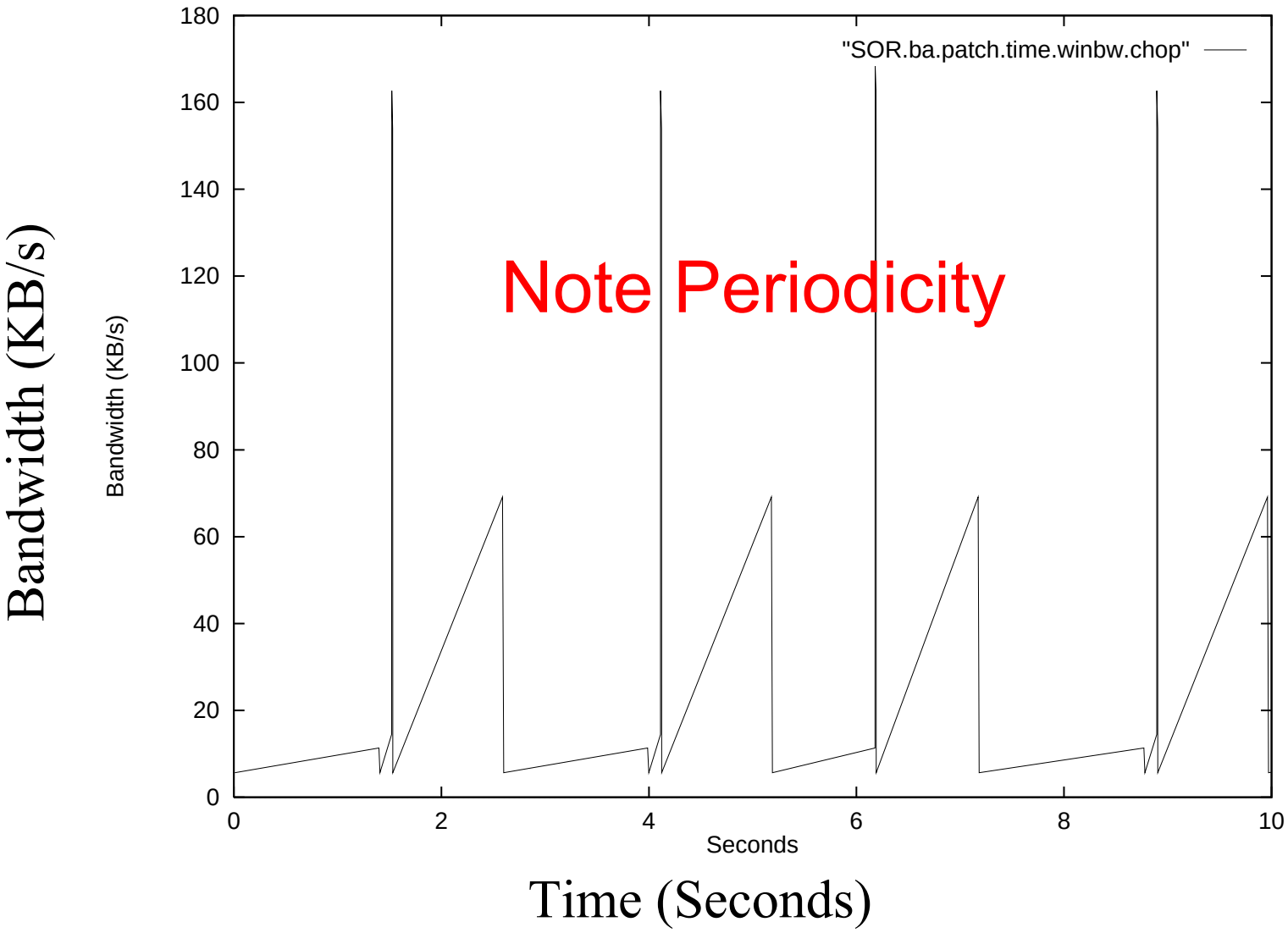
Typical network source model is heavy-tailed stochastic on-off which mix giving self-similarity

Long-term Average Bandwidth

Program	KB/s (Aggregate)	KB/s (Connection)
SOR	5.6	0.9
2DFFT	754.8	63.2
T2DFFT	607.1	148.6
SEQ	58.3	-
HIST	29.6	-
AIRSHED	32.7	2.7

Resource demands are often quite light

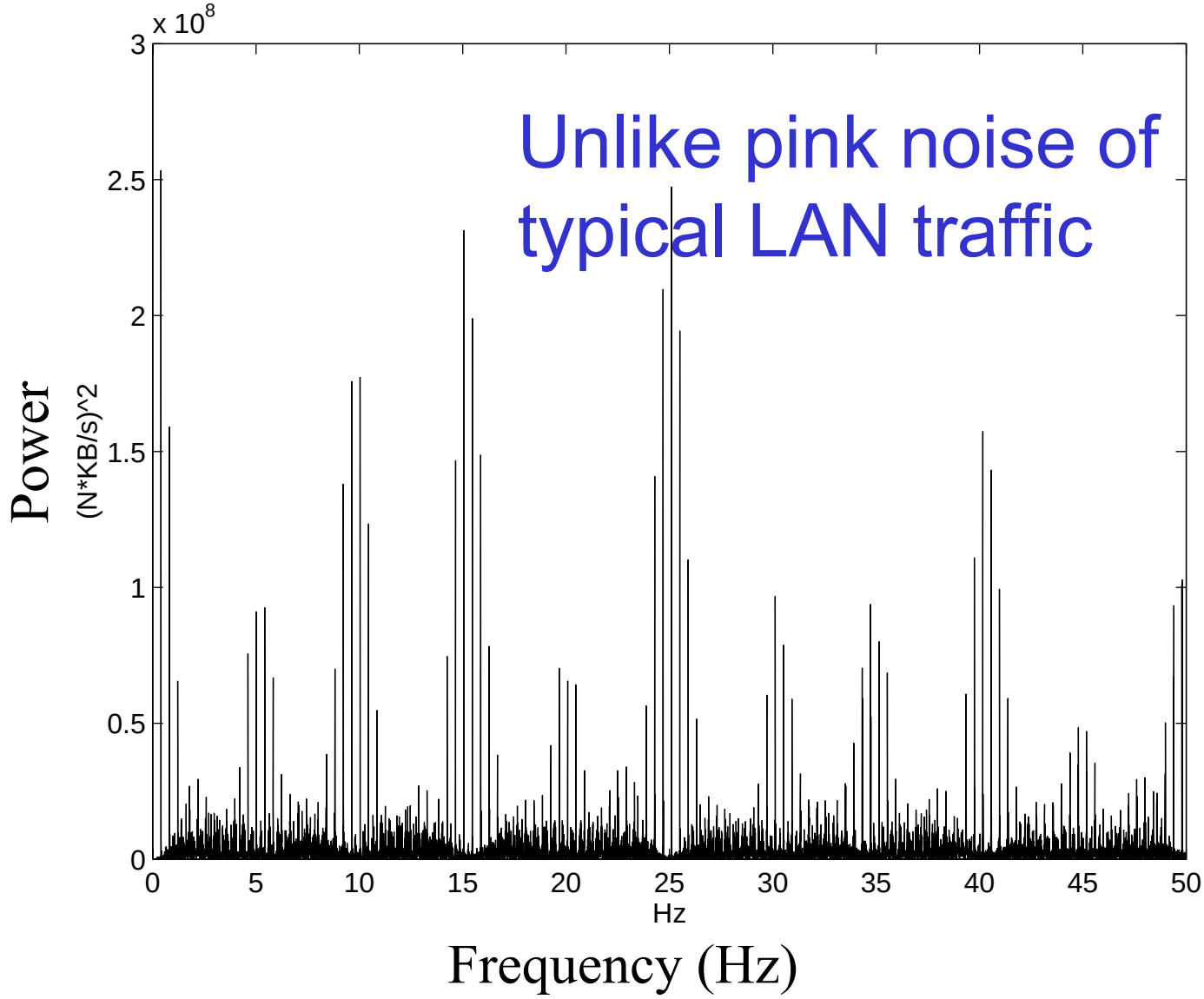
SOR: connection, time domain



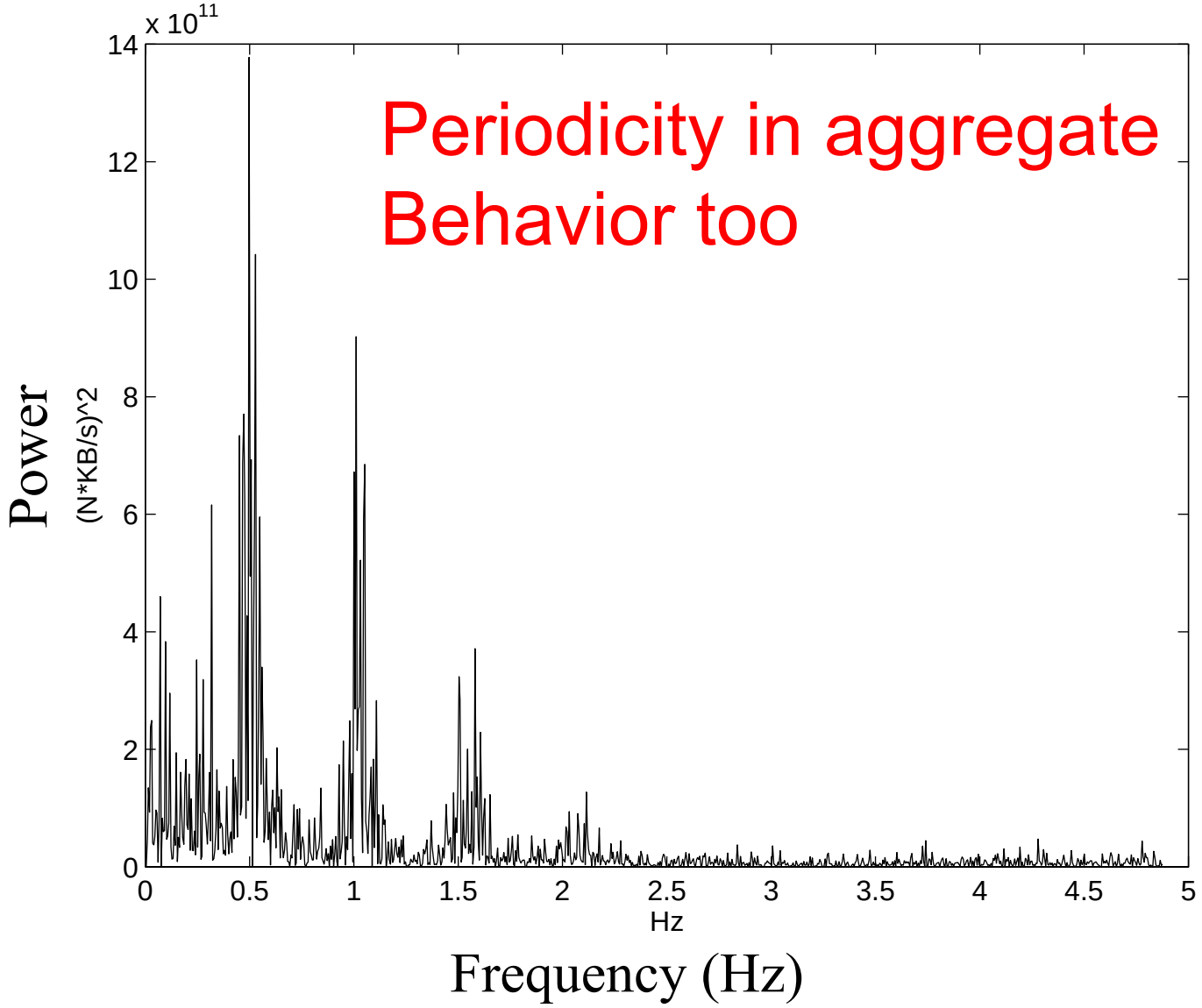
The Power Spectrum

- Frequency domain view of signal
- Density of variance as function of frequency
  - Power = variance
- Excellent for seeing periodicities
  - Periodically appearing feature in the signal turns into a spike in the power spectrum

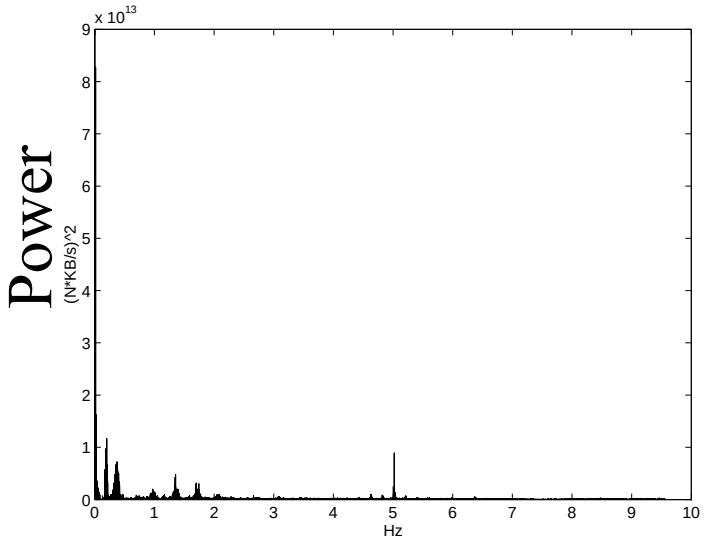
SOR: connection, power spectrum



2DFFT: aggregate, power spectrum



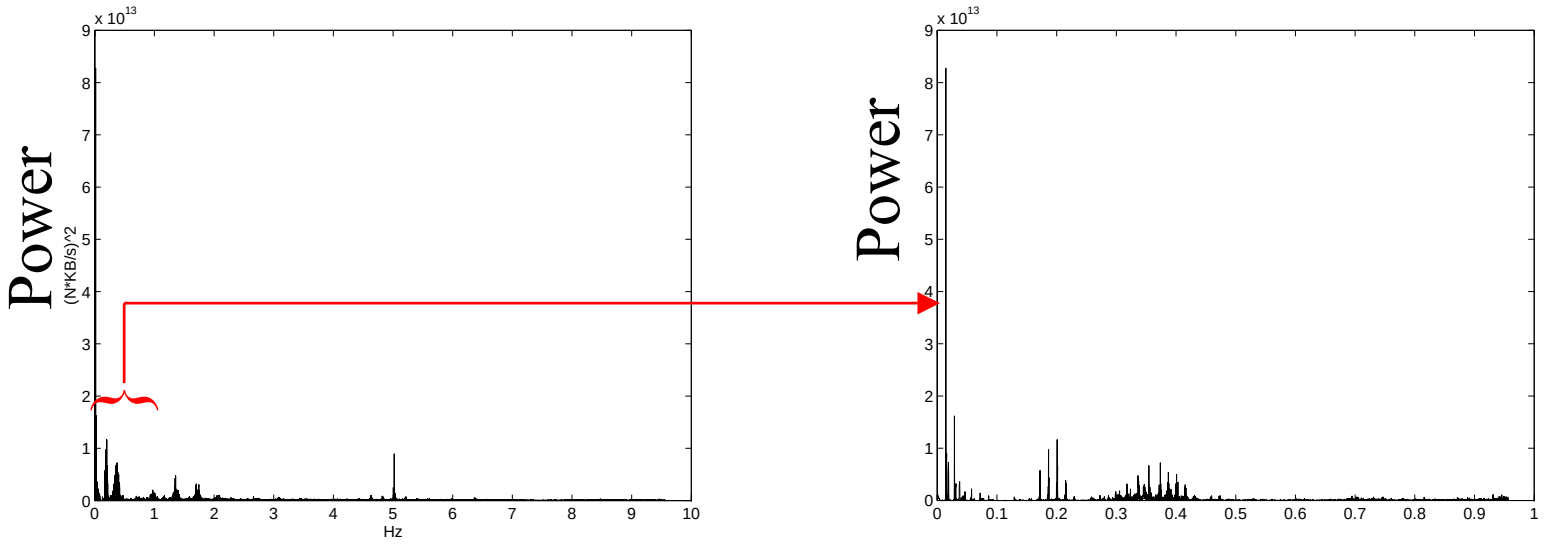
Airshed: aggregate, power spectrum



Frequency (0-10 Hz)

Periodicity at a limited number of timescales

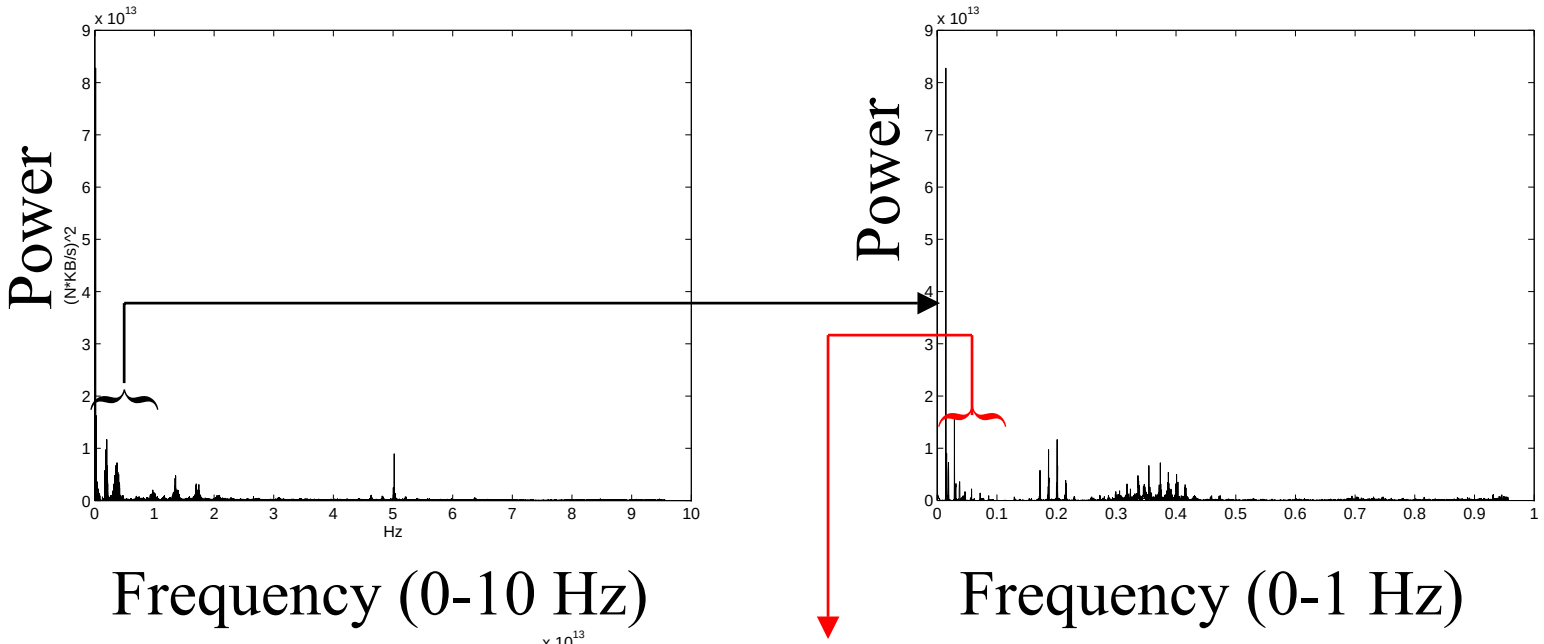
Airshed: aggregate, power spectrum

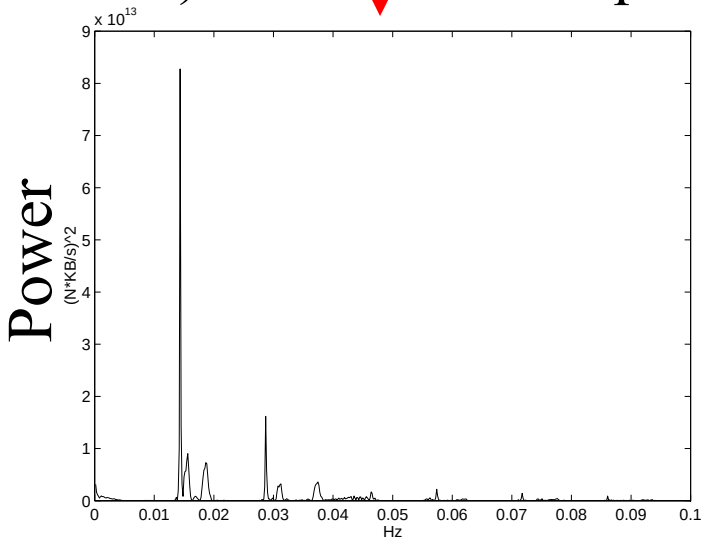


Frequency (0-10 Hz)

Frequency (0-1 Hz)

Airshed: aggregate, power spectrum





Implications for Network Prediction

- Networks with significant parallel workloads may be more predictable
  - Typical LAN and WAN look like pink noise
  - Mixing effects unclear, however
- Source model must be different
  - More deterministic, no heavy tails
- Parallel applications appear detectable

Implications for QoS Models

- Pattern should be conveyed
  - Show traffic correlation along multiple flows
- Source model more deterministic
  - Deterministic burst size
  - Deterministic burst interval that depends on proffered bandwidth
- More degrees of freedom to expose
  - Number of nodes
  - Closer coupling of network's and app's optimization problemss

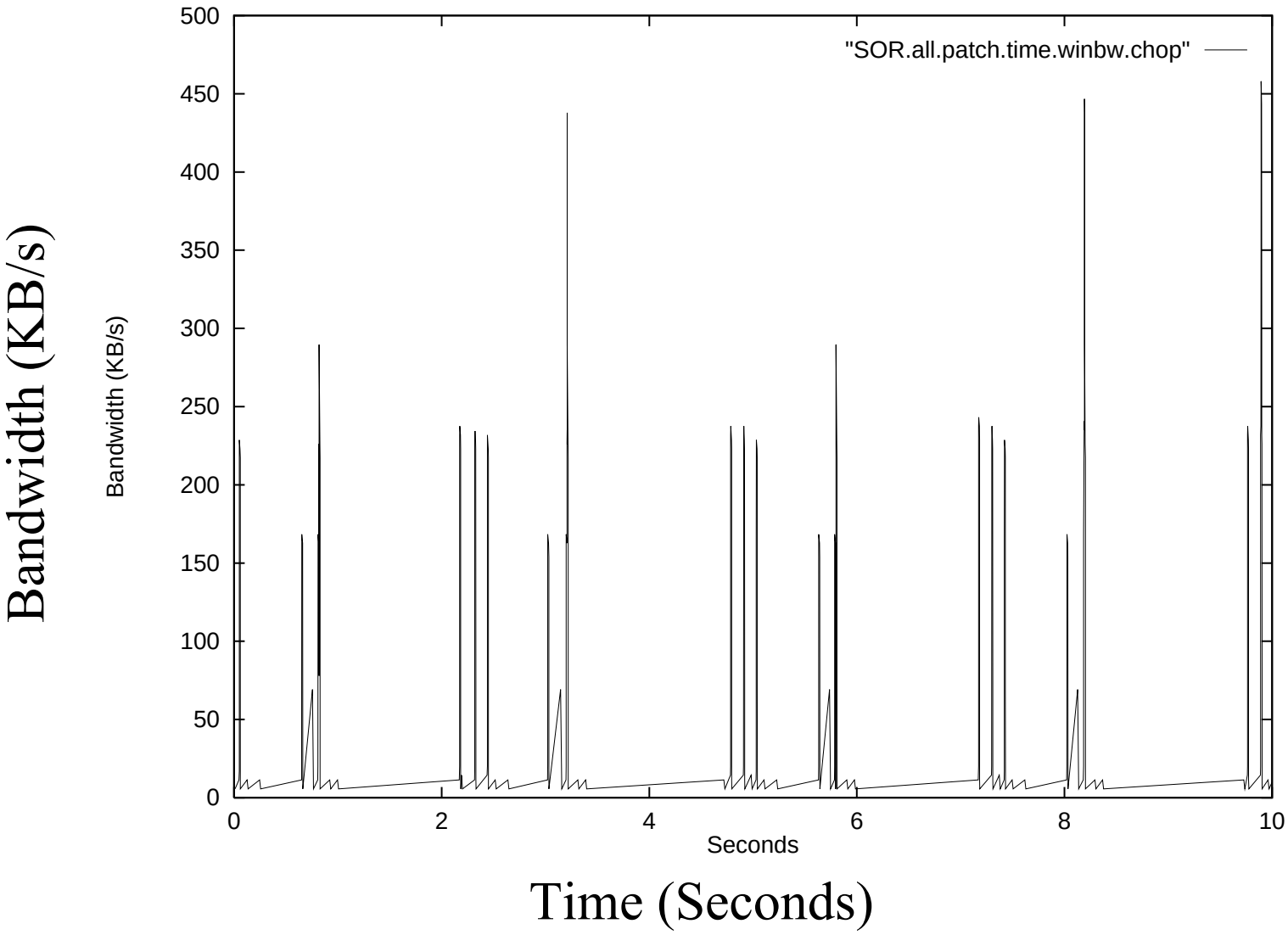
Conclusions and Future Work

- Analysis of packet traces of Fx codes running on a shared Ethernet using 1996 data
- Traffic very different from typical models
  - Simple packet size+interarrival behaviors
  - Correlation between flows
  - Periodicity within flows and in aggregate
- QoS models should allow and exploit these differences
- Parallel traffic appears more predictable than common traffic

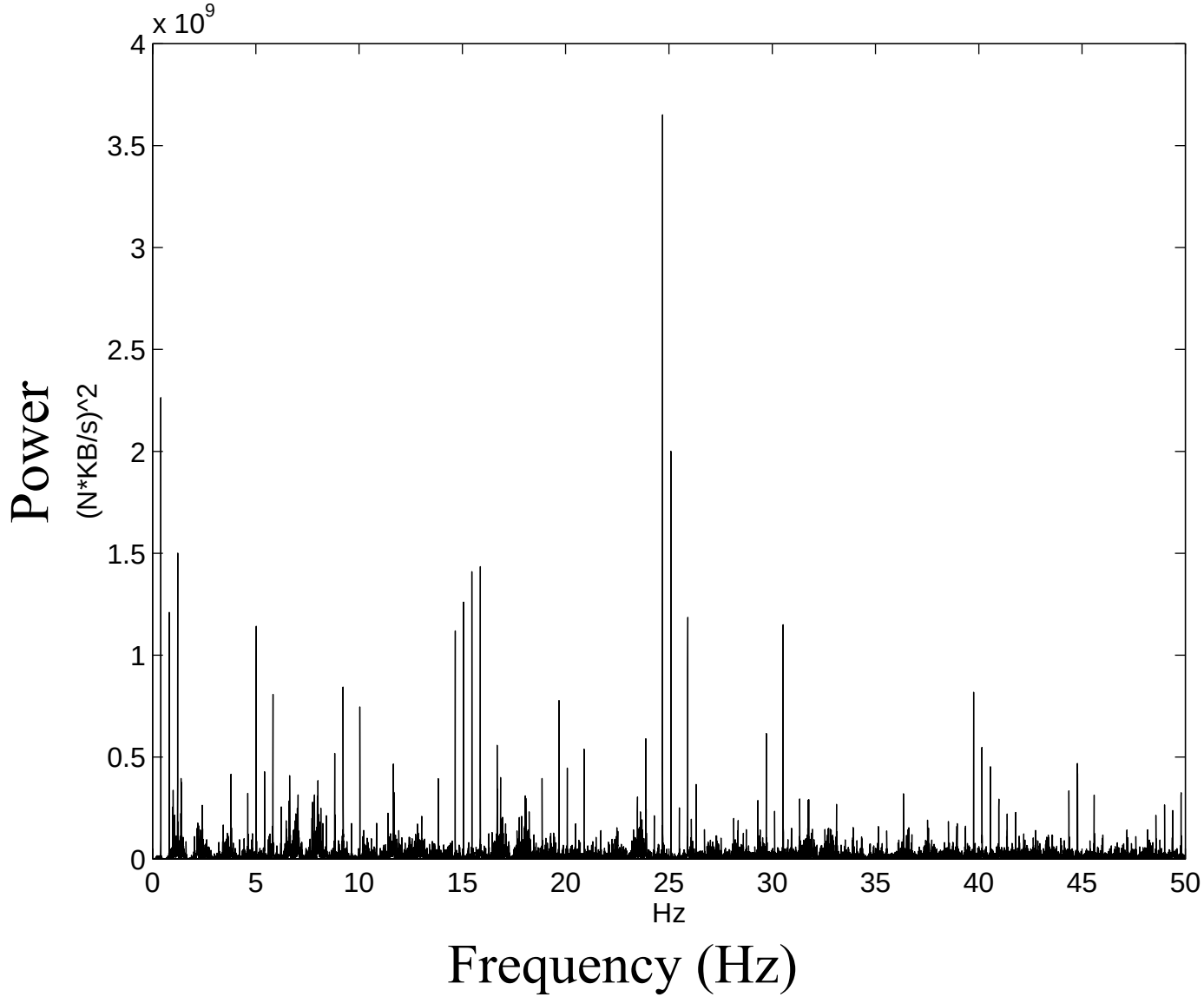
For More Information

- http://www.cs.northwestern.edu/~pdinda
- Resource Prediction System (RPS) Toolkit
  - http://www.cs.northwestern.edu/~RPS
- Prescience Lab
  - http://www.cs.northwestern.edu/~plab
- Fx and AIRSHED
  - http://www.cs.cmu.edu/~fx
  - http://www.cs.cmu.edu/afs/cs.cmu.edu/project/gems/www/hpcc.html

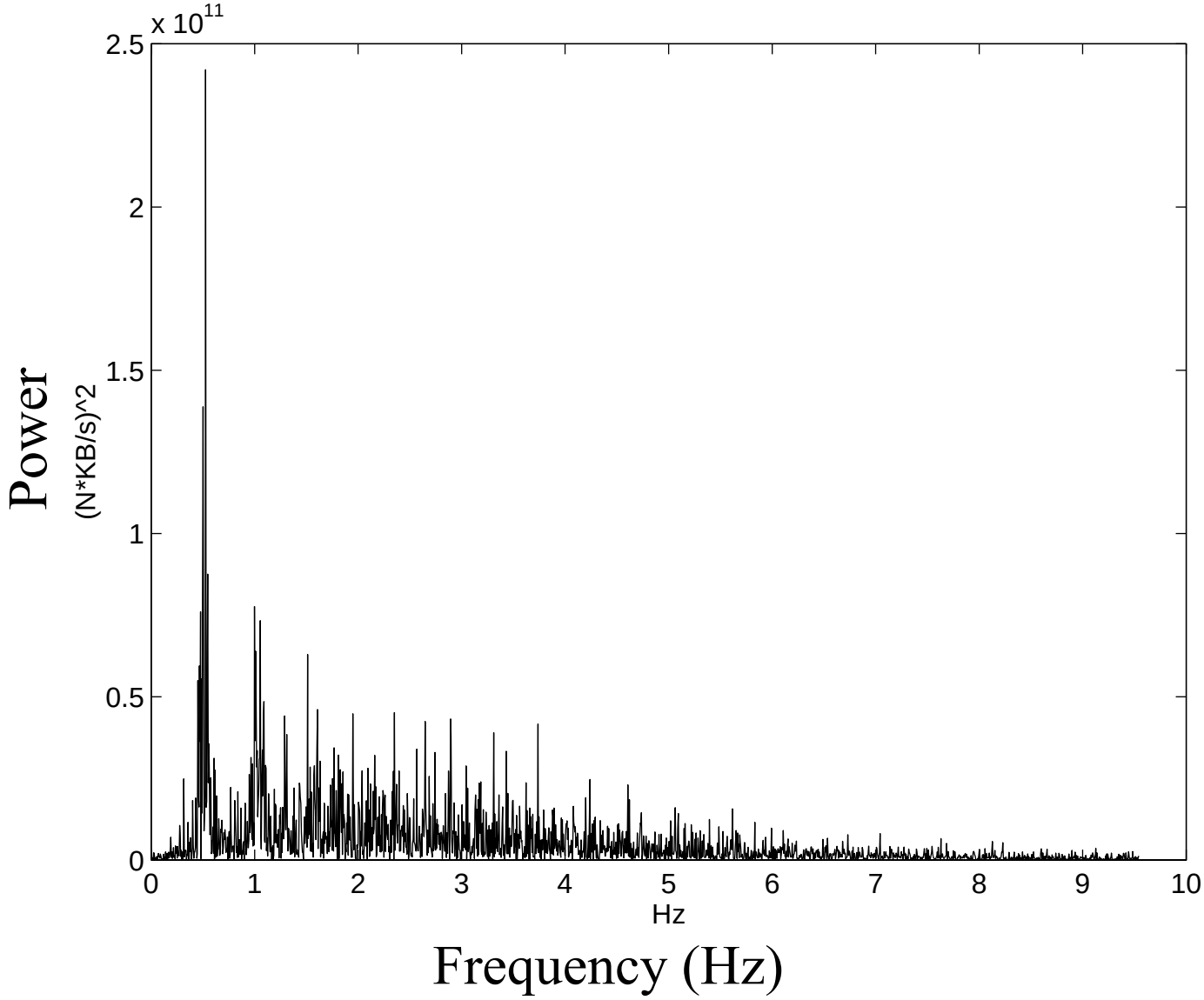
SOR: aggregate, time domain



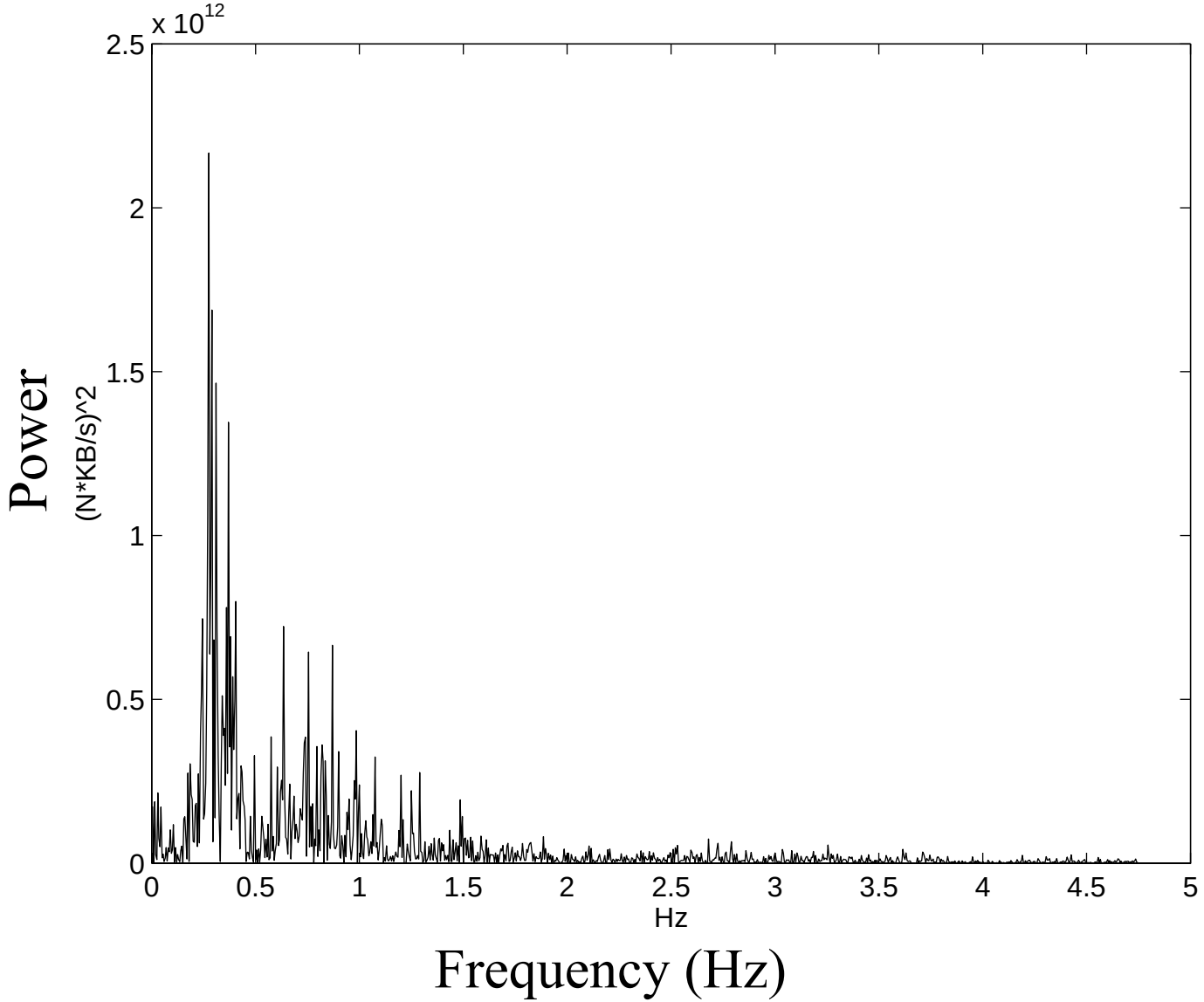
SOR: aggregate, freq domain



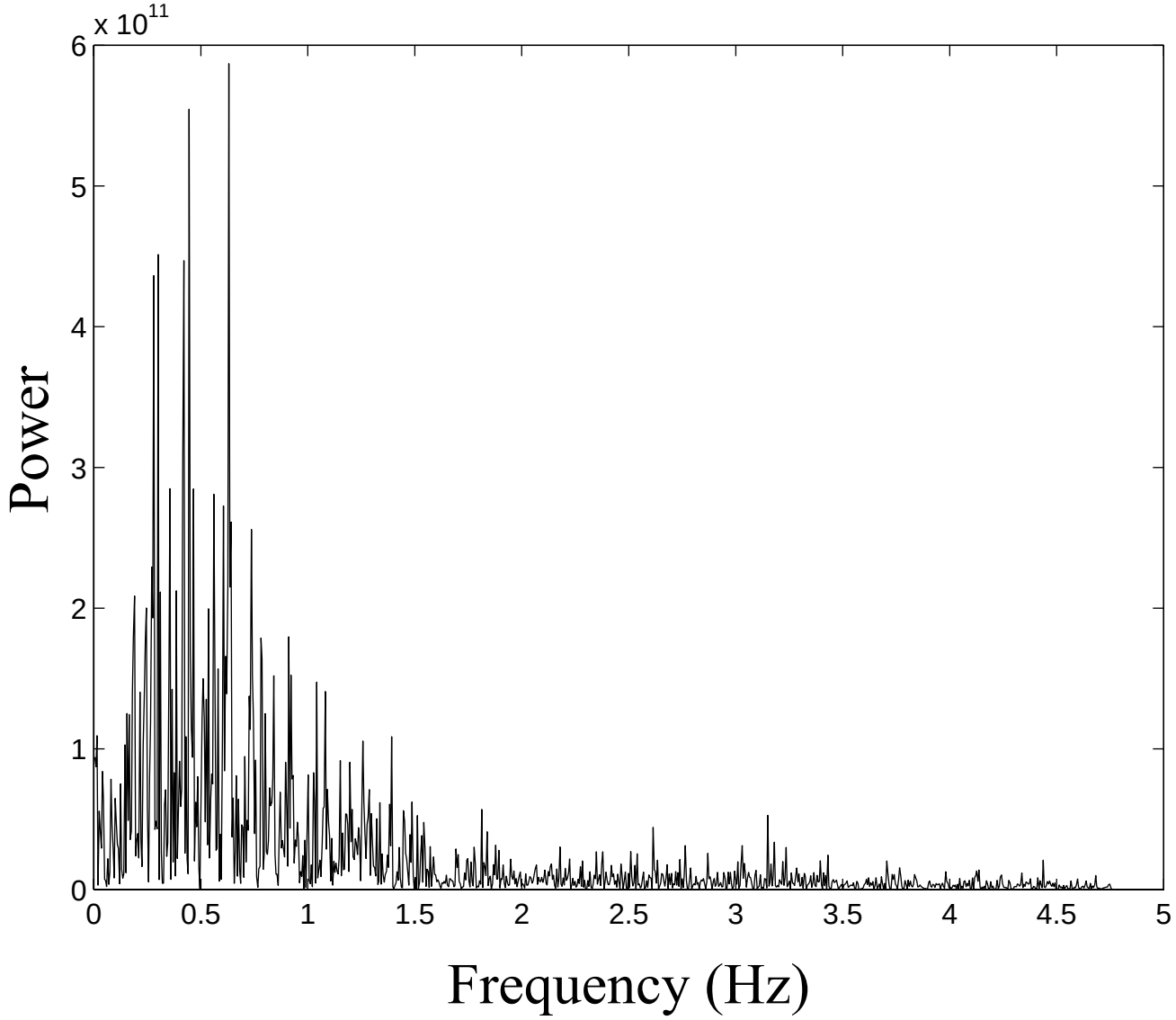
2DFFT: connection, freq domain



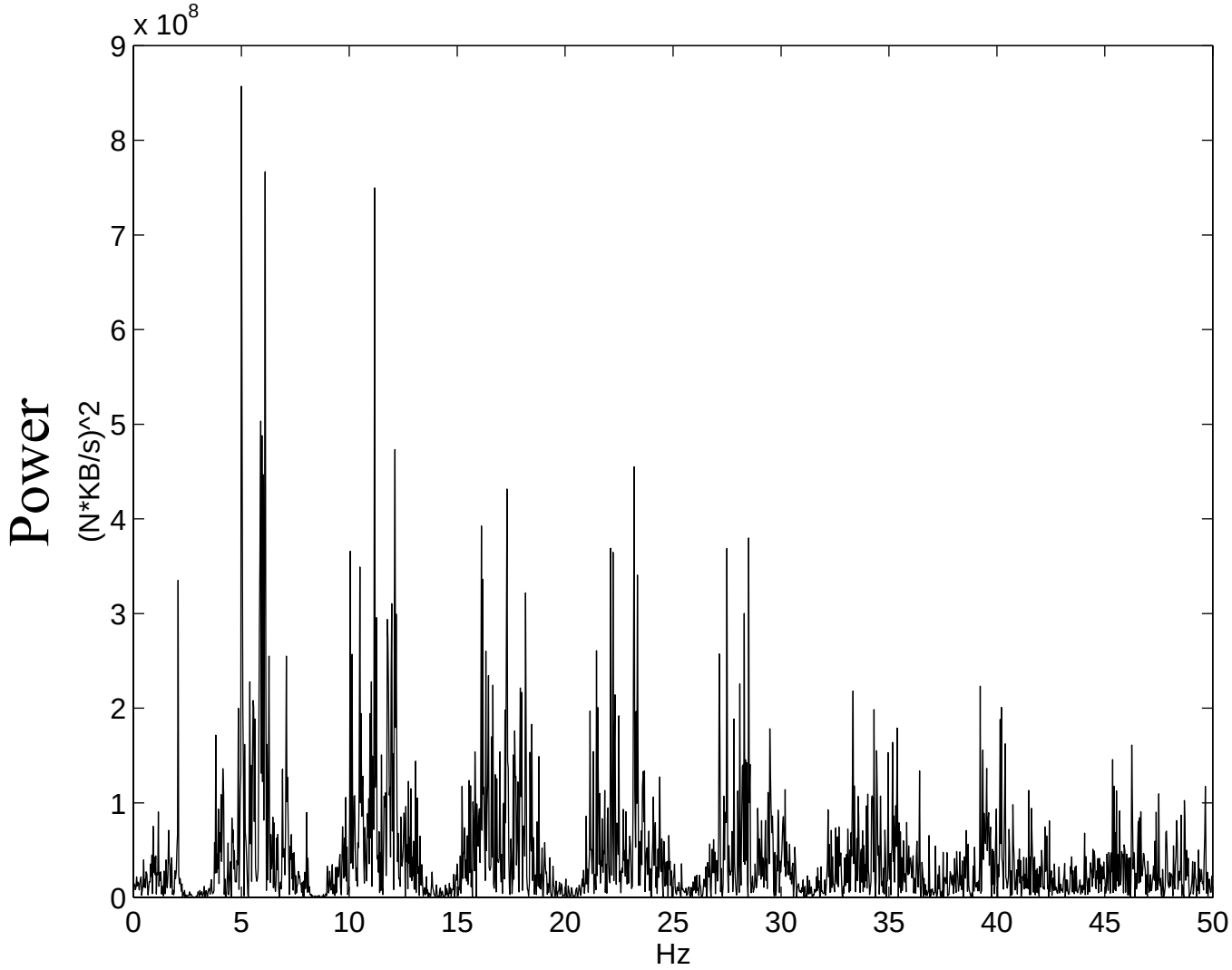
T2DFFT: aggregate, freq domain



T2DFFT: connection, freq domain



HIST: aggregate, freq domain



Frequency (Hz)

SEQ: aggregate, freq domain