Analysis of packet sampling in the frequency domain

Alfredo Grieco

Politecnico di Bari, Italy

Chadi BARAKAT

INRIA Sophia Antipolis, France
Planète research group

Email: Chadi.Barakat@sophia.inria.fr
WEB: http://www.inria.fr/planete/chadi

Motivations

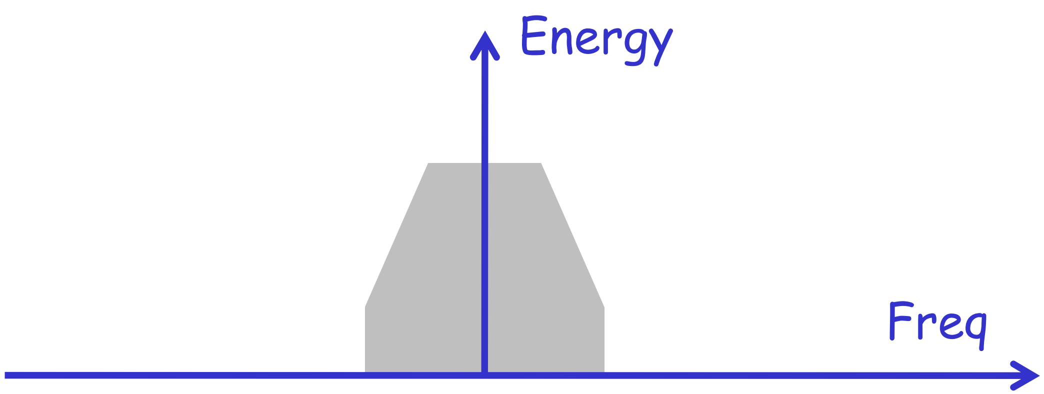
- Packet sampling, a technique to reduce the load on routers by monitoring a subset of packets then inverting sampled measurements
- Many papers have studied the problem with stochastic tools (Duffield et al, Veitch et al, Estan et al, Diot et al, Zseby et al)
  - A snapshot of traffic
  - Sampled randomly then measured
  - Inverted to reduce some error function e.g. MSE
  - Metrics: traffic volume, flow size distribution, heavy hitter statistics, flow counting, etc
- How does packet sampling impact the spectrum of the traffic?
  - What frequencies can we preserve for management applications?

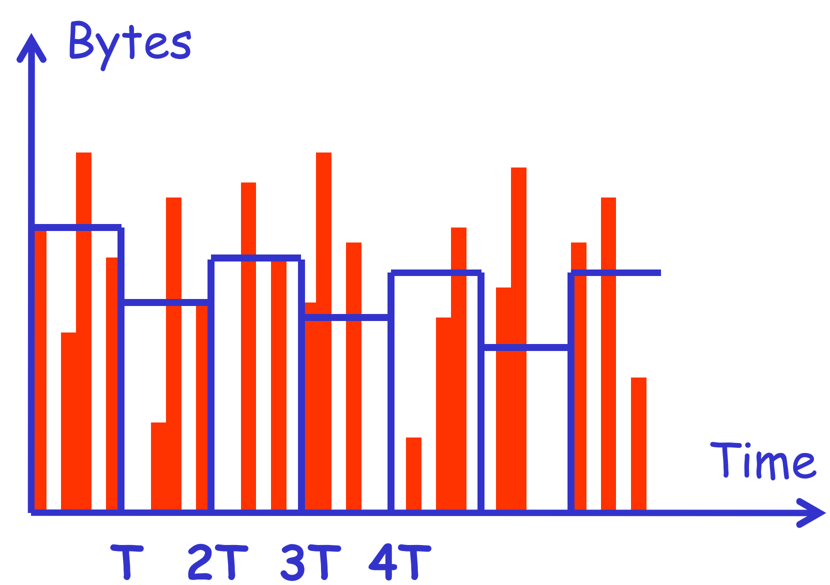
Outline

- Models for traffic and spectrum
- Analysis of packet sampling
- Aliasing noise and its removal by low pass filtering
- The Filter-Bank solution
- Simulation results and conclusions

Traffic model and spectrum

- Traffic: A time series of packets of different sizes d_n
- Measured traffic rate:
  - Divide time into small bins
  - Volume of bytes per bin divided by bin length T
  - The larger the bin the coarser the measurement
- Targeted traffic spectrum:
  - Spectrum of the binned traffic rate
  - Energy of different frequency components



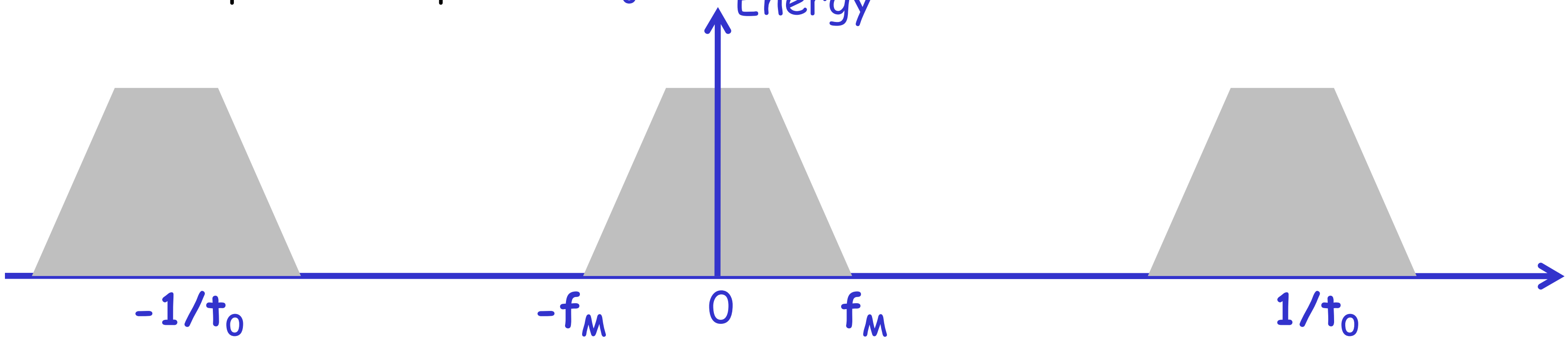


INRIA SOPHIA ANTIPOLIS

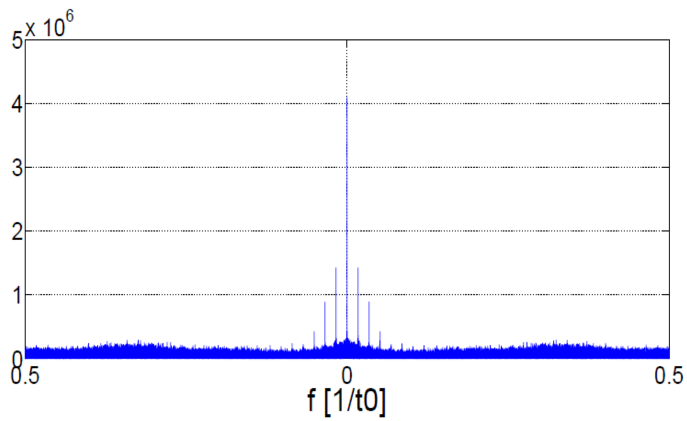
Analysis: No Sampling

- Let $D(f)$ be the spectrum of the original traffic
  - Traffic discretized in tiny time slots t_0
  - Periodic spectrum of period $1/t_0$

$$D(f) = \sum D_0(f+n/t_0)$$

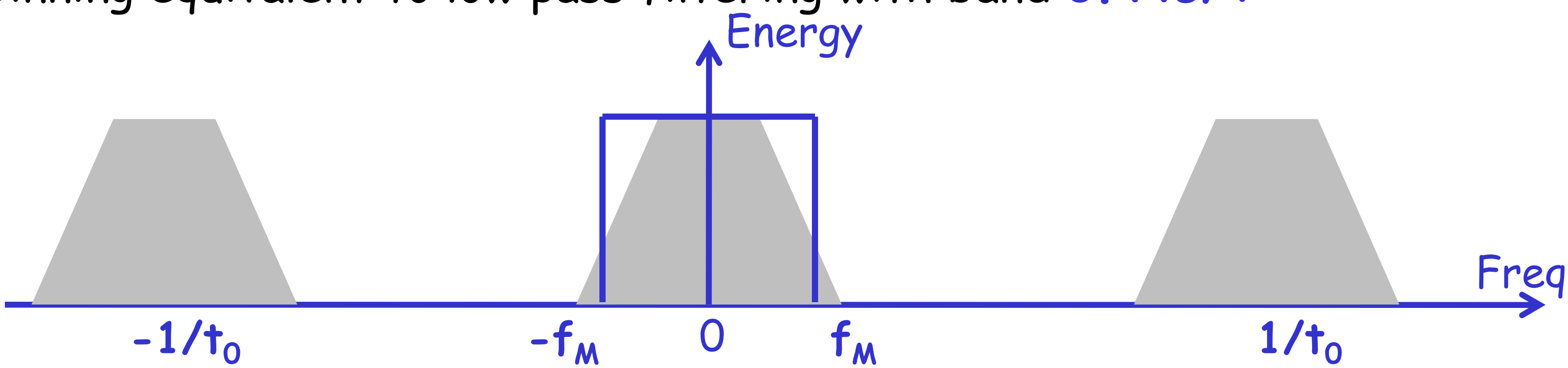


- Suppose the existence of a maximum frequency f_M with $0 < f_M < 1/t_0$
- An example of a real baseband

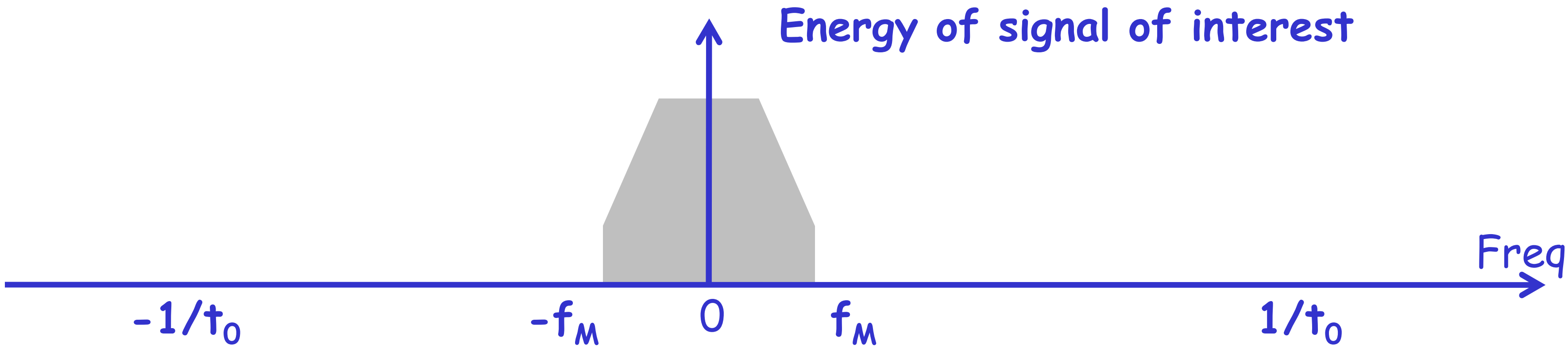


Analysis: No Sampling, With Binning

- Binning equivalent to low pass filtering with band **0.445/T**



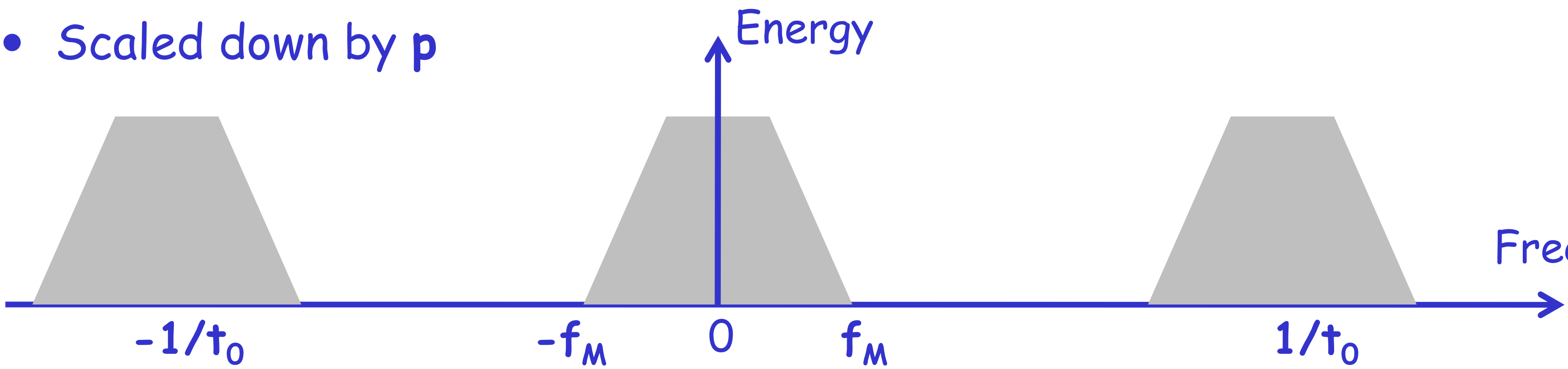
Convolution with a low pass filter of band 0.445/T

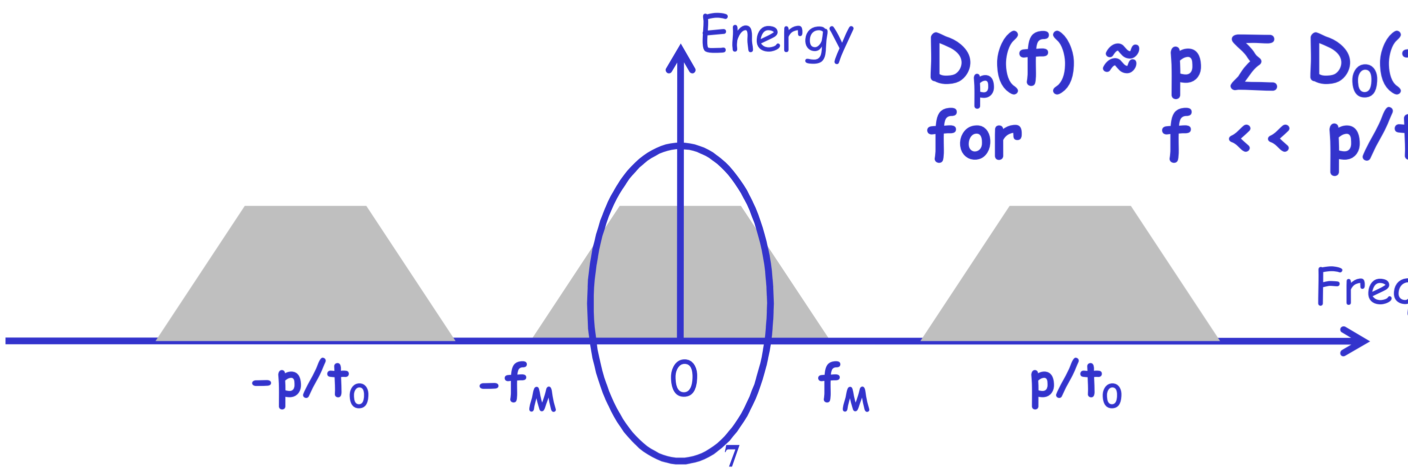


INRIA SOPHIA ANTIPOLIS

Analysis: Sampling

- Traffic sampled with rate **p < 1**
- Let **$D_p(f)$** be the spectrum of the sampled traffic
  - Result: A replication of **$D_0(f)$** with period **p/t_0** in the band of interest
  - Scaled down by **p**



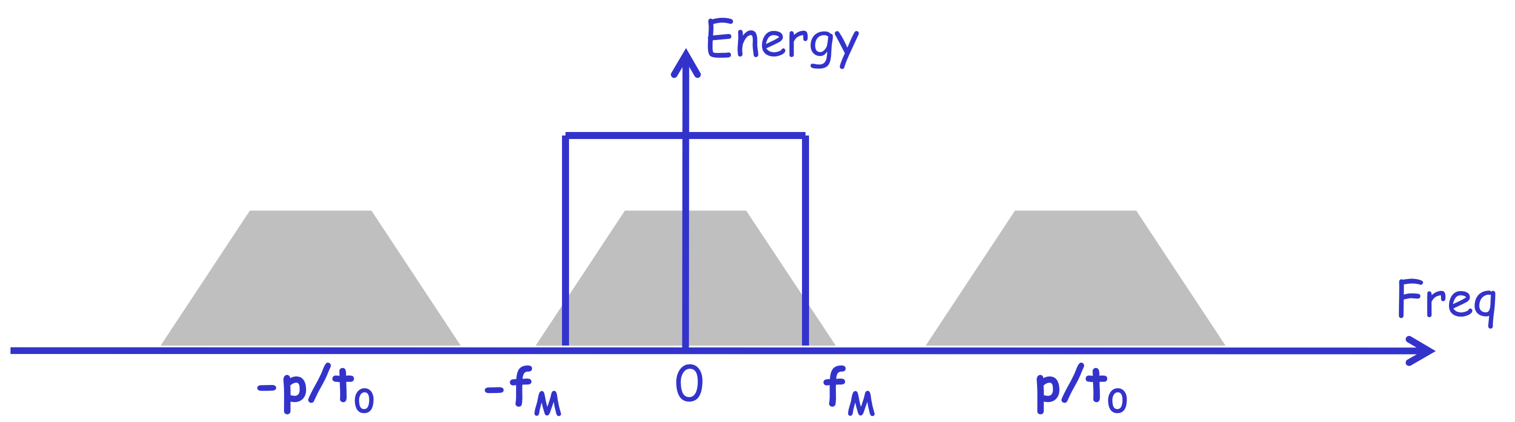


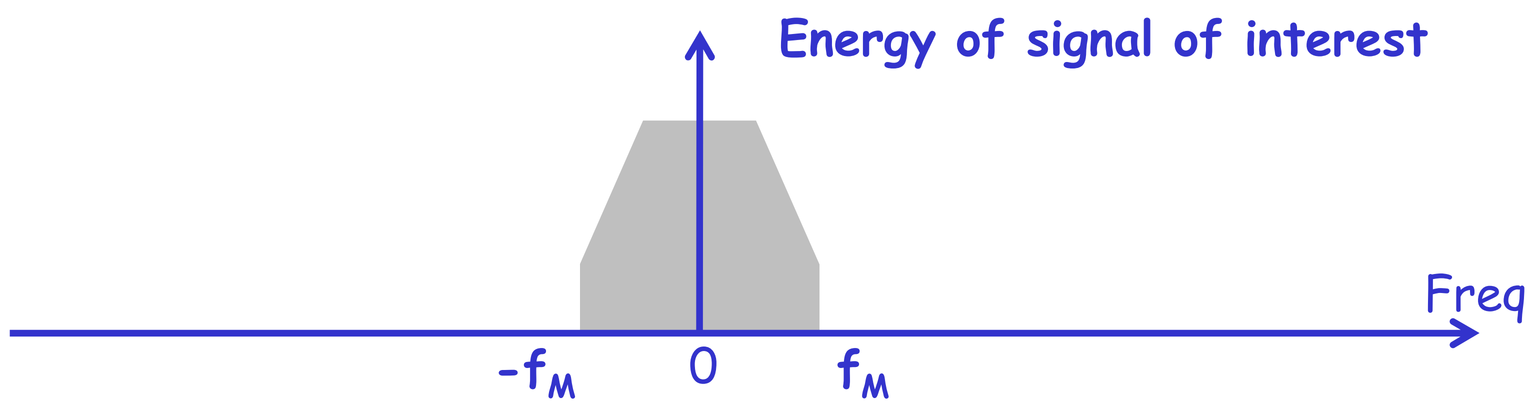
$$D_p(f) \approx p \sum D_0(f+n.p/t_0)$$

for $f << p/t_0$

Analysis: Sampling, With Binning

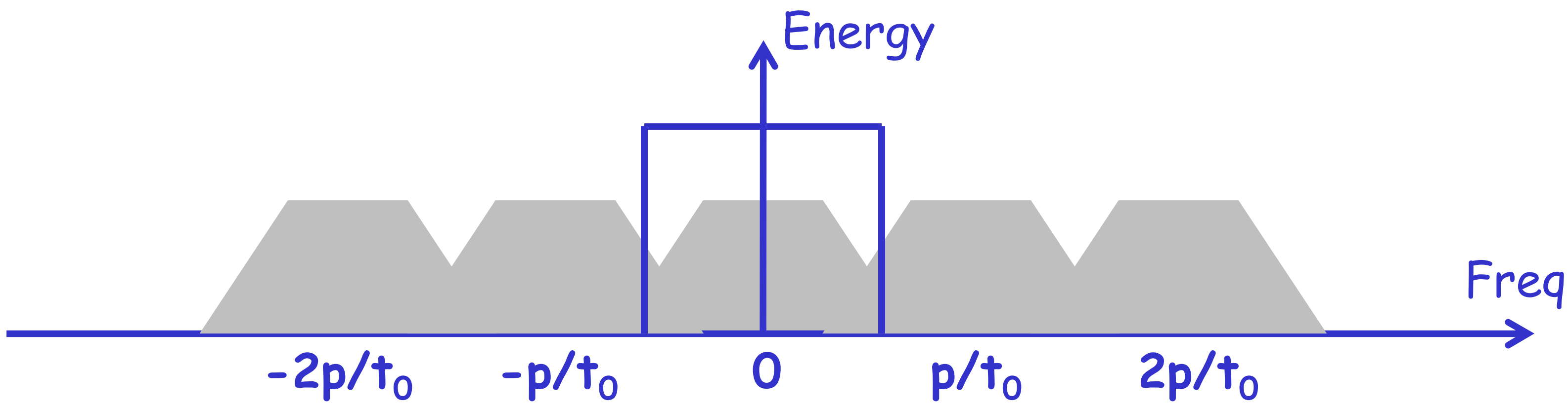
- By binning and scaling up by **1/p**, one can recover the signal of interest



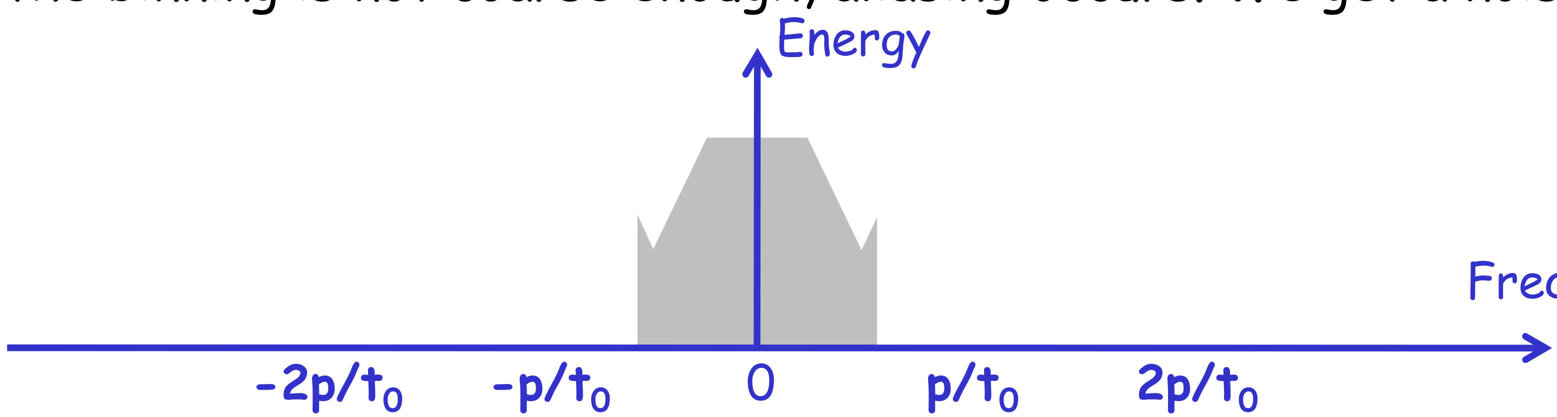


Aliasing for small sampling rates

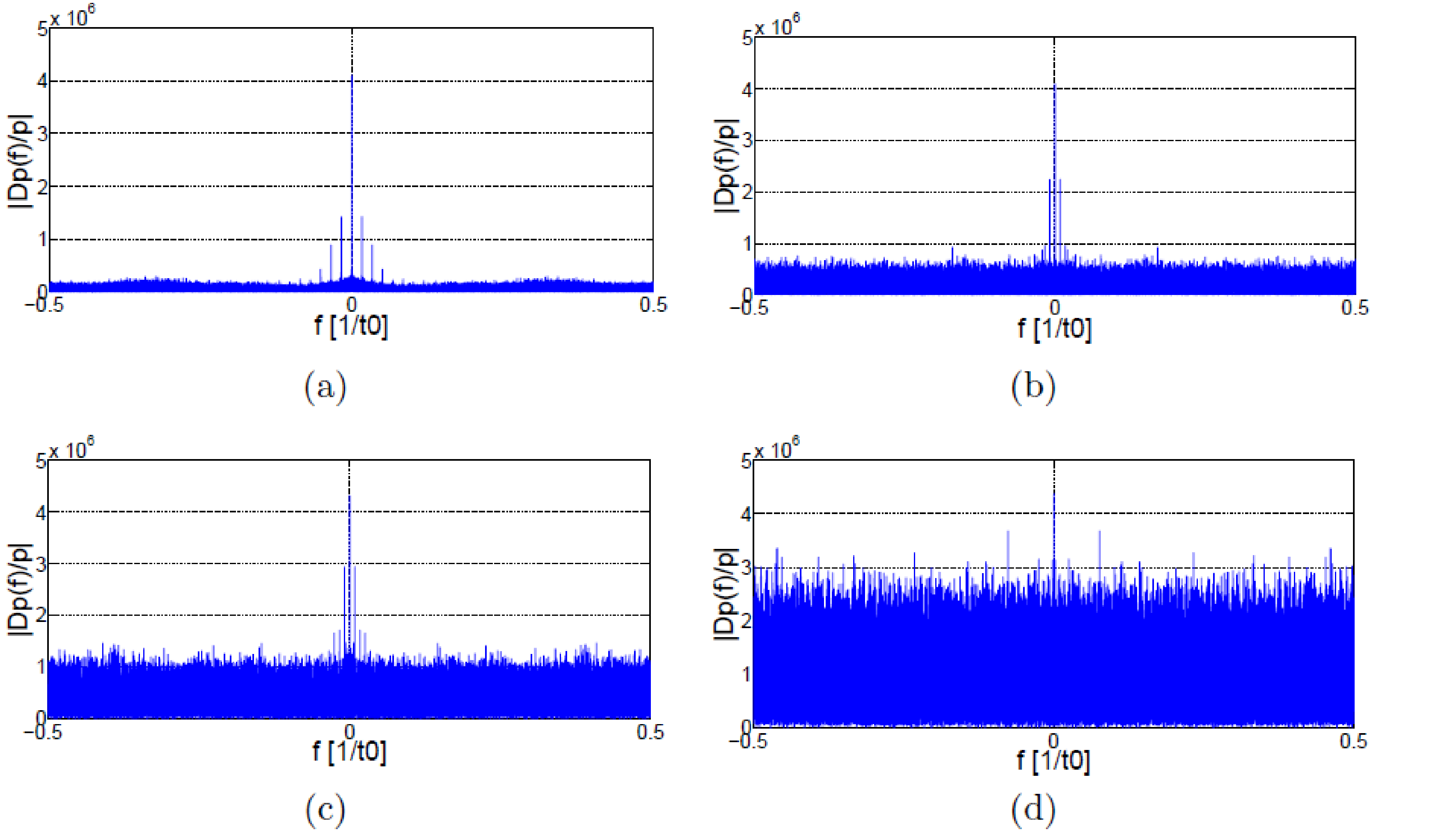
- The smaller the sampling rate, the closer the replicas
  - There is a sampling rate below which they overlap



- If the binning is not coarse enough, aliasing occurs. We get a noisy signal.



Aliasing in the baseband



Baseband component of $D_p(f)/p$: (a) $p = 1$; (b) $p = 0.1$; (c) $p = 0.03$; (d) $p = 0.005$.

INRIA SOPHIA ANTIPOLIS

Aliasing noise elimination

For a traffic of maximum frequency f_M in the baseband

- Either increase the sampling rate to avoid the overlap of replicas in the band of interest
  - Always work
- Or increase the binning interval T
  - Will not work if $p/t_0 < f_M$ (sampling too much)
- General result: Spectrum of the binned traffic rate is preserved upon traffic sampling if and only if

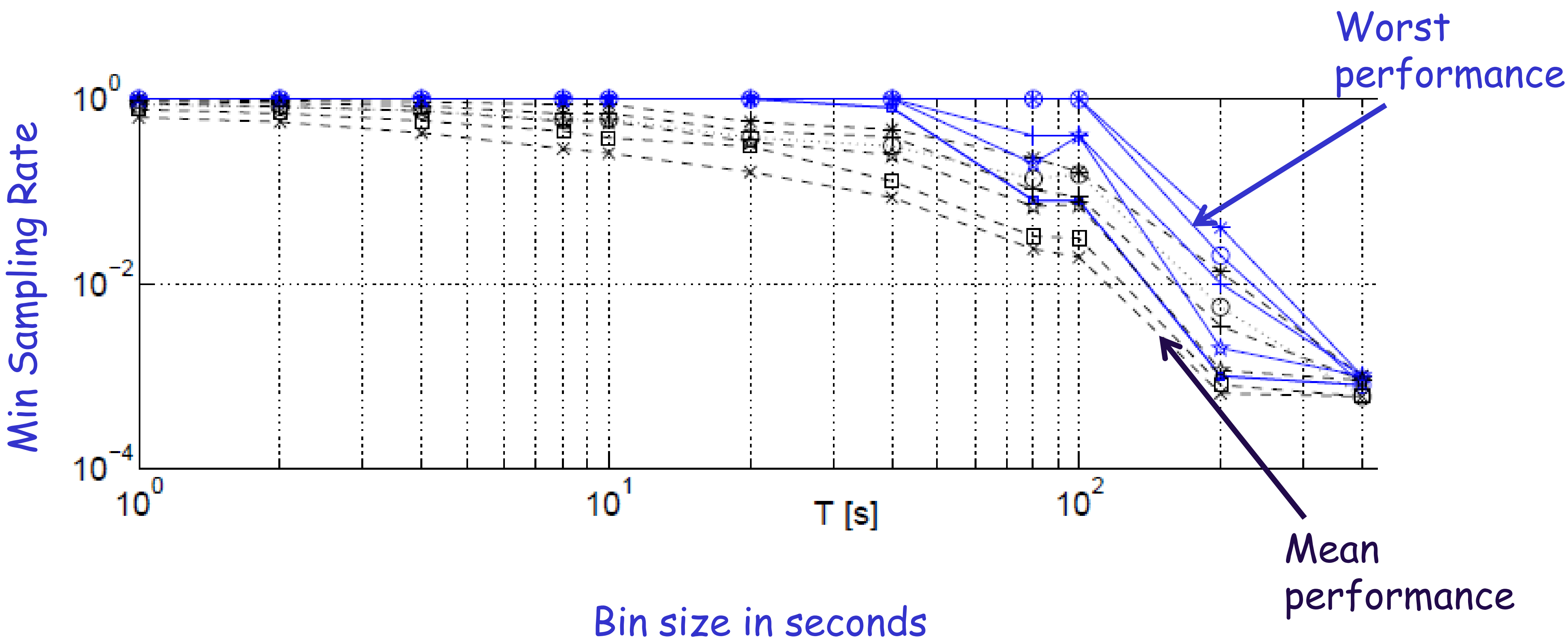
$$0{,}445 / T < p/t_0 - f_M$$

Determining the bin to use

- Fixing the sampling rate and changing the bin is not enough
  - The energy changes with
- One has to fix the bin and change the sampling rate
  - In practice, the traffic is already sampled, so downsampling is not possible. Only upsampling is possible.
- Our solution: Filter-Bank to check Traffic Variance (Energy)
  - Try different bin sizes.
  - For each bin, further increase the sampling rate.
  - If energy (variance) quickly increases, aliasing exists.
  - If energy (variance) slowly increases, the bin size is fine.

Sampling rates vs bin sizes

- Using traces from the Japanese MAWI project cut into pieces



INRIA SOPHIA ANTIPOLIS

Conclusions

- An analysis of packet sampling in the frequency domain
- An expression relating:
  - Sampling rate
  - Maximum frequency in the baseband
  - Minimum binning interval

  in order to avoid aliasing and sampling noise
- Future plans:
  - Estimate the amount of noise caused by aliasing
  - Further study of traffic spectrum and the origins of its components