Panoramica sugli algoritmi Multicanonici e nuovo algoritmo migliorato, con applicazione alla PMD

A. Bononi, N. Rossi, A. Orlandini

Dipartimento di Ingengeria dell'Informazione
Università di Parma

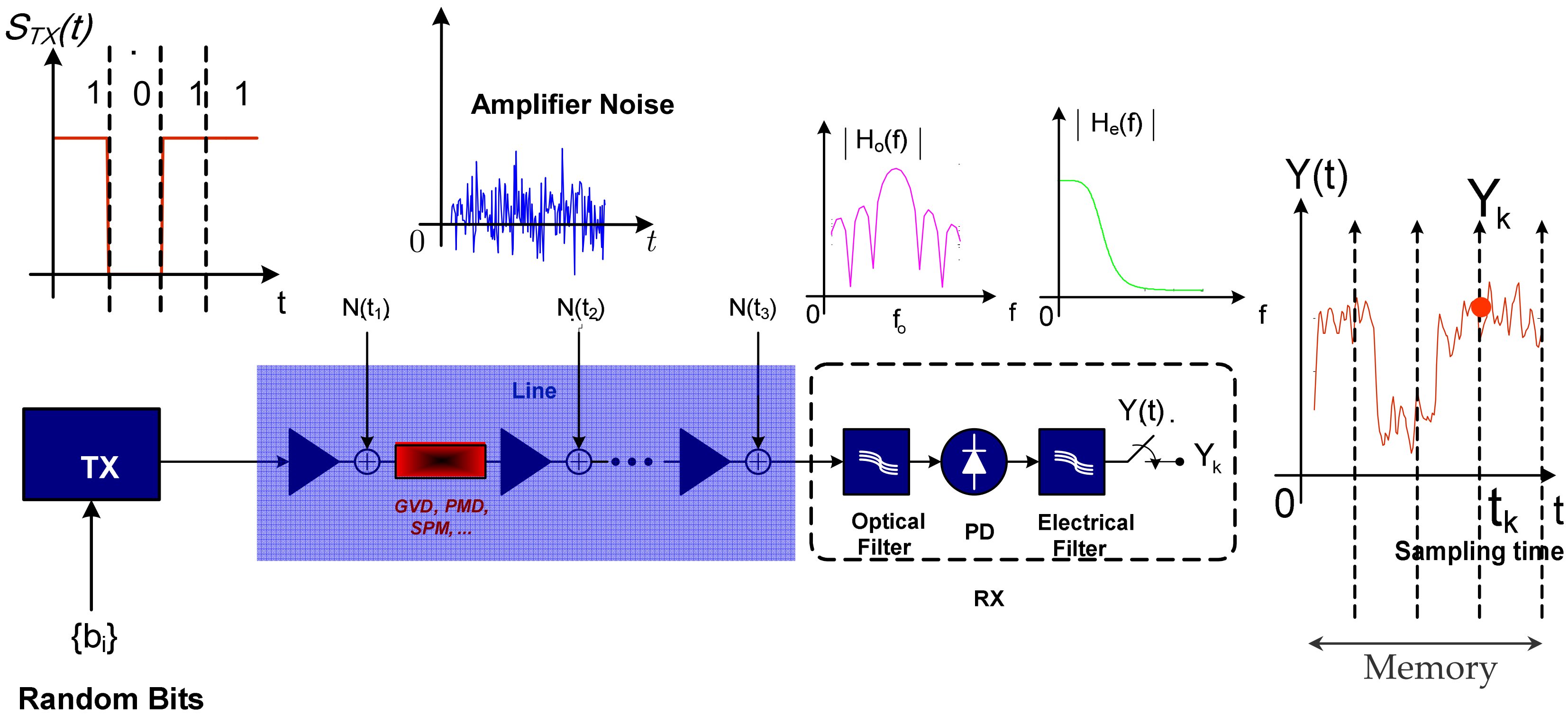
Outline

- Motivation
- Monte Carlo (MC)
- Importance Sampling (IS)
- Flat Histogram (FH) Methods
    - Multicanonical Monte Carlo (MMC)
    - Wang Landau (WL)
    - Fast MMC
- Example: a PMD problem
- Conclusions

Motivation

In telecommuincations, we often need to estimate the probability density function (PDF) of a random variable (RV) of interest

Motivation

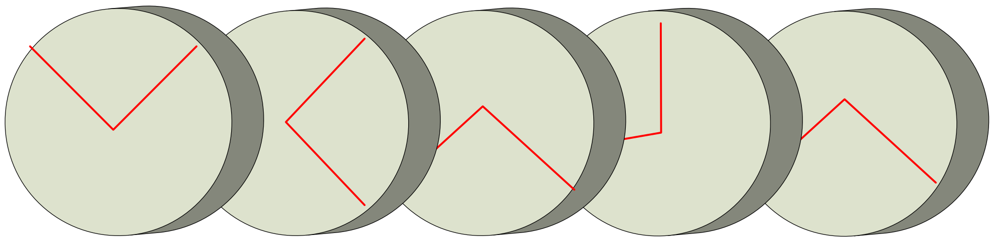


Decision Variable

$$Y=g(\underline{X})$$

"state" $\underline{X}$=(set of noise samples along the line
and of random bits adjacent to bit of interest
falling within system memory)

Motivation



When calculating the differential group delay (DGD) of a fiber with PMD:

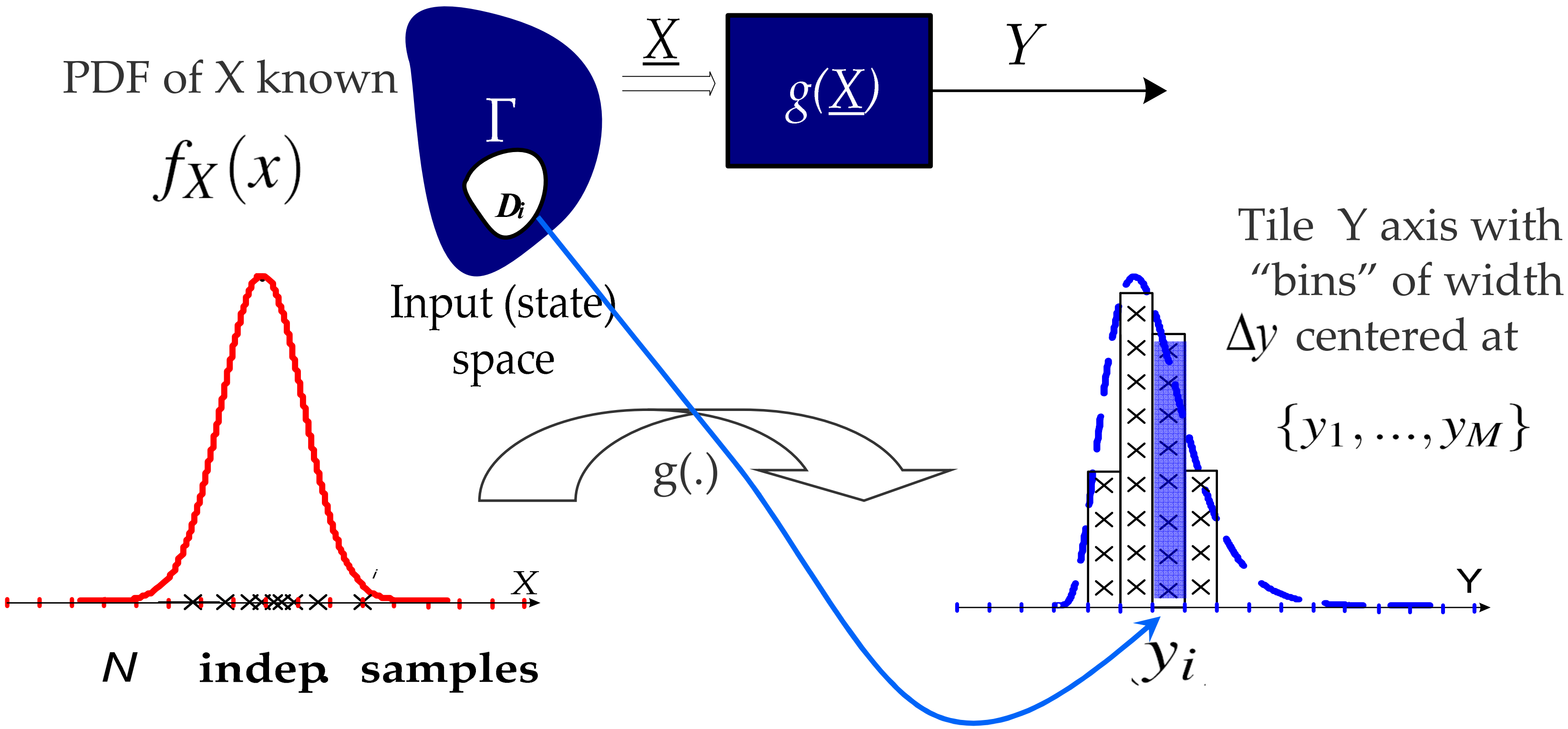
$$Y=g(\underline{X})$$

"state" $\underline{X}$=(set of PSP orientations and DGD of each waveplate)

Outline

- Motivation
- **Monte Carlo (MC)**
- Importance Sampling (IS)
- Flat Histogram (FH) Methods
    - Multicanonical Monte Carlo (MMC)
    - Wang Landau (WL)
    - Fast MMC
- Example: a PMD problem
- Conclusions

Monte Carlo (MC)



Thereby estimate the probability mass function (PMF) of discretized Y:

$$P_i \equiv P(y_i) \triangleq P\{Y \approx y_i\}$$

and evaluate PDF from PMF as $f_Y(y) \simeq P_i/\Delta y$

Monte Carlo

Define $$I_i(Y) = \begin{cases} 1 & \text{if}\,\{Y \approx y_i\} \\ 0 & \text{else} \end{cases}$$ Indicator of i-th bin visit (FLAG)

Probability that a sample falls in bin i:

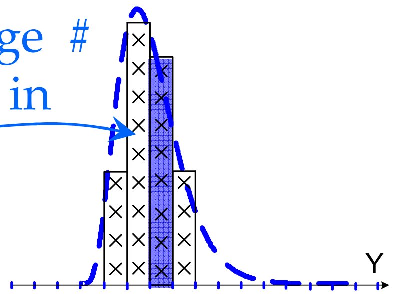
$$P_i = \int_{D_i} f_X(x)dx = \int_\Gamma I_i(g(x))f_X(x)dx = E[I_i(g(X))]$$

Sample mean of RV $I_i(g(X))$ $$\hat{P}_i^{MC} \triangleq \frac{1}{N}\sum_{j=1}^{N} I_i(g(X_j)) = \frac{N_i}{N}$$

Define $$\varepsilon_i \triangleq Var[\hat{P}_i]/P_i^2$$ quadratic relative error of estimate

$$\varepsilon_i^{MC} = \frac{1-P_i}{NP_i}$$

Average # points in

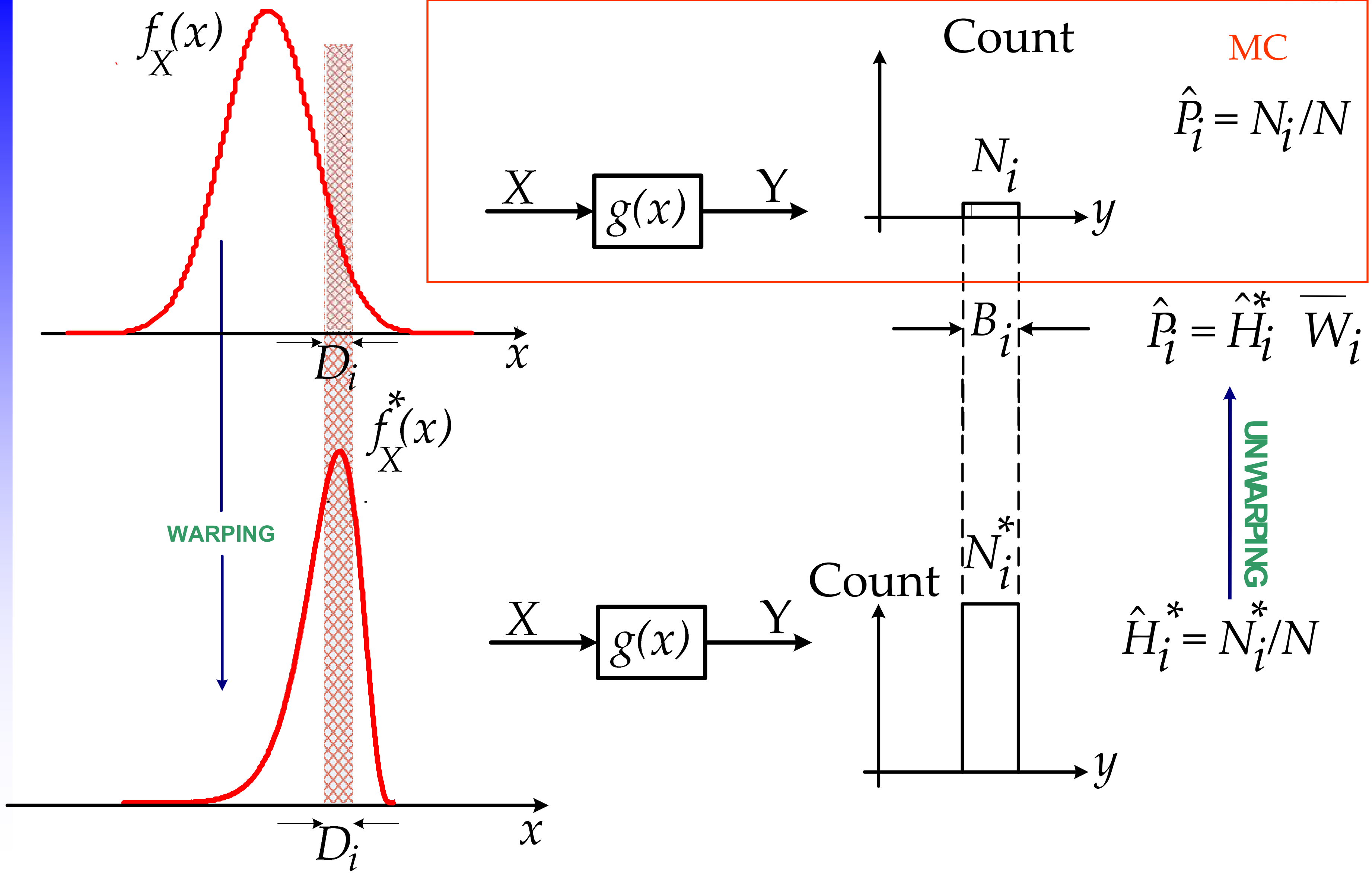


...but on tails get few or no samples.......

Outline

- Motivation
- Monte Carlo (MC)
- **Importance Sampling (IS)**
- Flat Histogram (FH) Methods
    - Multicanonical Monte Carlo (MMC)
    - Wang Landau (WL)
    - Fast MMC
- Example: a PMD problem
- Conclusions

Importance Sampling (IS)



Importance Sampling: Analysis

Coefficient $\overline{w}_i$ to unwarp formally found as follows:

$$P_i = \int_\Gamma I_i(g(x)) \left[\frac{f_X(x)}{f_X^*(x)}\right] f_X^*(x)dx = E^*\left[I_i(g(X))w(X)\right]$$

IS weight (known function, hopefully << 1 on tails)

Form sample mean of RV $I_i(g(X))w(X)$

$$\hat{P}_i^{IS} \triangleq \frac{1}{N}\sum_{j=1}^{N} I_i(g(X_j))w(X_j) = \underbrace{\left(\frac{N_i^*}{N}\right)}_{\hat{H}_i^*} \underbrace{\left[\frac{1}{N_i^*}\sum_{n=1}^{N_i^*} w(X_n)\right]}_{\overline{w}_i}$$

$\hat{H}_i^*$: Histogram of visits

$\overline{w}_i$: Average weight on bin i

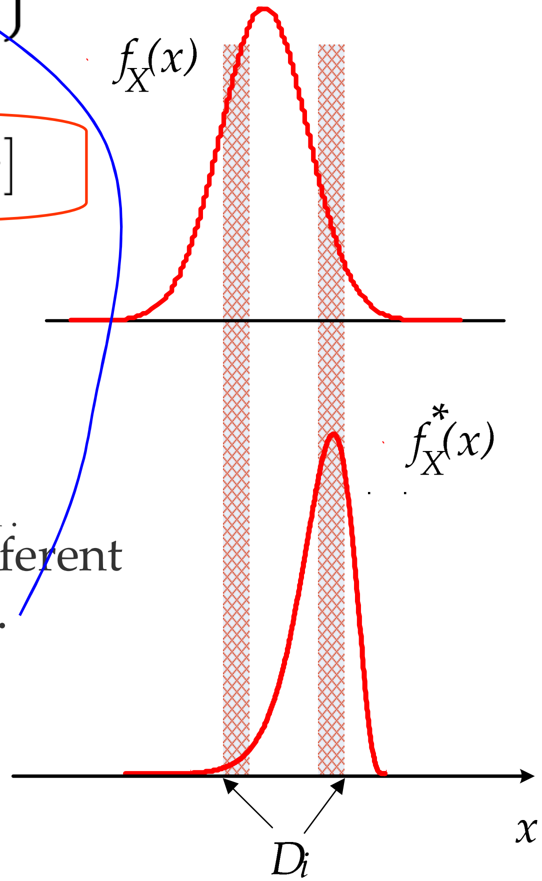
Importance Sampling: Precision

$$E^*[\hat{P}_i^{IS}] = P_i$$ unbiased, like MC

$$\varepsilon_i^{IS} = \frac{1}{N}\left\{\frac{1}{H_i}\left(\frac{Var^*[w(X)|X \in D_i]}{(E^*[w(X)|X \in D_i])^2}+1\right)-1\right\}$$

$$H_i = E^*[\hat{H}_i]$$

$$E^*[w(X)|X \in D_i] = E^*[\overline{w}_i]$$



True limit of IS is our a priori ignorance of domains D_i. Hence may get the wrong warping, assigning widely different weights over the same D_i, thus increasing the ....

Uniform Weight IS (UWIS)

$E^*[\hat{P}_i^{IS}] = P_i$ unbiased, like MC

$$\varepsilon_i^{IS} = \frac{1}{N}\left\{\frac{1}{H_i}\left(\frac{Var^*[w(X)|X \in D_i]}{(E^*[w(X)|X \in D_i])^2} + 1\right) - 1\right\}$$

$Var^*[w(X)|X \in D_i] \to 0$

Best warpings give uniform weight over whole D_i (UWIS = uniform weight IS)

$$\varepsilon_i^{UWIS} = \frac{1}{N}\left\{\frac{1}{H_i} - 1\right\}$$

$$\varepsilon_i^{MC} = \frac{1-P_i}{NP_i}$$

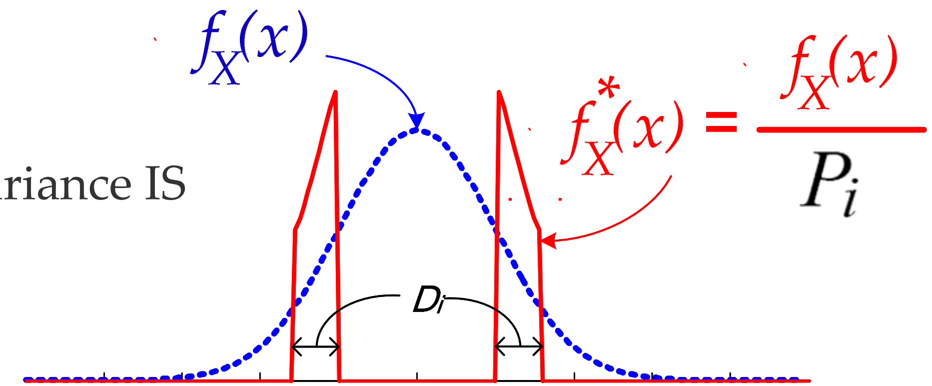
If $P_i \ll H_i \ll 1$ gain over MC is evident....

Zero Variance IS (ZVIS)

$$\varepsilon_i^{UWIS} = \frac{1}{N}\left\{\frac{1}{H_i} - 1\right\}$$

If $H_i = 1$

Zero Variance IS



This is optimal (exclusively for estimate of bin i).
All samples fall within D_i !
Not realizable, as requires knowledge of P_i, ie, of what we wish to estimate...

Flat Histogram IS (FHIS)

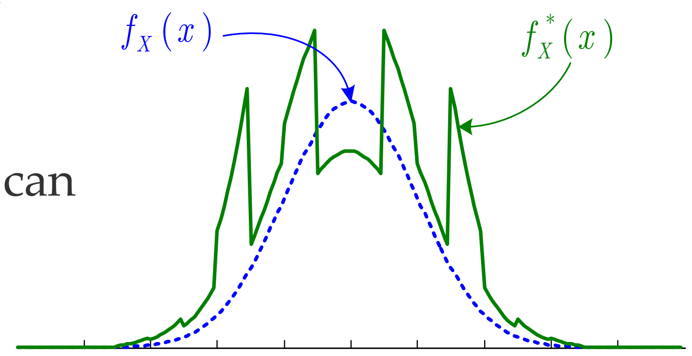
An "optimal" UWIS exists for the estimate of whole PMF $\{P_1, P_2, \ldots, P_M\}$ of Y

$$f_X^*(x) = \frac{f_X(x)}{MP(g(x))}$$

$P(g(x))$ is probability of the bin where $y = g(x)$ falls

It is UWIS since $w(x) = MP_i$ for every $x \in D_i$

Get it by adding up ZVIS of all bins and renormalizing. Non realizable, as ZVIS, but can be approximated, as we will see...



Properties:

1) $H_i = E^*[\hat{H}_i] = 1/M$ **on all bins: flat histogram (FH) on average**

2) $\varepsilon_i^{FHIS} = \frac{M-1}{N}$ **on all bins: same relative precision!**

That's the best that one can do with N samples !!!

Outline

- Motivation
- Monte Carlo (MC)
- Importance Sampling (IS)
- **Flat Histogram (FH) Methods**
    - Multicanonical Monte Carlo (MMC)
    - Wang Landau (WL)
    - Fast MMC
- Example: a PMD problem
- Conclusions

Flat Histogram (FH) Methods

A family of algorithms (MMC, Wang-Landau and others), which, starting from known PDF of X, $f_X(x)$, build a sequence of warped PDFs

$$f_X^{(n+1)}(x) = \frac{f_X(x)}{c_n \Theta_n(g(x))},\ n = 0, 1, 2, \ldots \quad \text{of UW type}$$

(where $\underline{\Theta}_n \triangleq \{\Theta_n(y_i)\}_{i=1}^{M}$ is an estimate of PMF of Y at cycle n

and c_n its normalization constant) from which we draw samples to form a new estimate $\underline{\Theta}_{n+1}$ of PMF of Y, up to convergence to FH:

$$f_X^*(x) = \frac{f_X(x)}{MP(g(x))}$$

At convergence (empirically verified by a flat visits histogram on average) have:

$$c_n \rightarrow M \quad \underline{\Theta}_n \rightarrow \underline{\mathsf{P}} \triangleq \{P_i\}_{i=1}^{M}$$

Algorithms differ in their update law $\underline{\Theta}_n \rightarrow \underline{\Theta}_{n+1}$

Outline

- Motivation
- Monte Carlo (MC)
- Importance Sampling (IS)
- Flat Histogram (FH) Methods
  - **Multicanonical Monte Carlo (MMC)**
  - Wang Landau (WL)
  - Fast MMC
- Example: a PMD problem
- Conclusions

Multicanonical Monte Carlo (MMC)

- **First FH method, invented by physicist Berg in 1992.**
- **Update law based on a UWIS estimate:**

1. At step (or cycle) *n+1*, *N* samples drawn from $f_X^{(n+1)}(x) = \frac{f_X(x)}{c_n \Theta_n(g(x))}$
2. For every sample calculate $Y_j = g(X_j)$
3. from these evaluate visits histogram $\hat{H}_i^{(n+1)} \triangleq \hat{H}_{n+1}(y_i) = N_i^{(n+1)}/N$
4. updated IS estimate of PMF of Y is finally given by

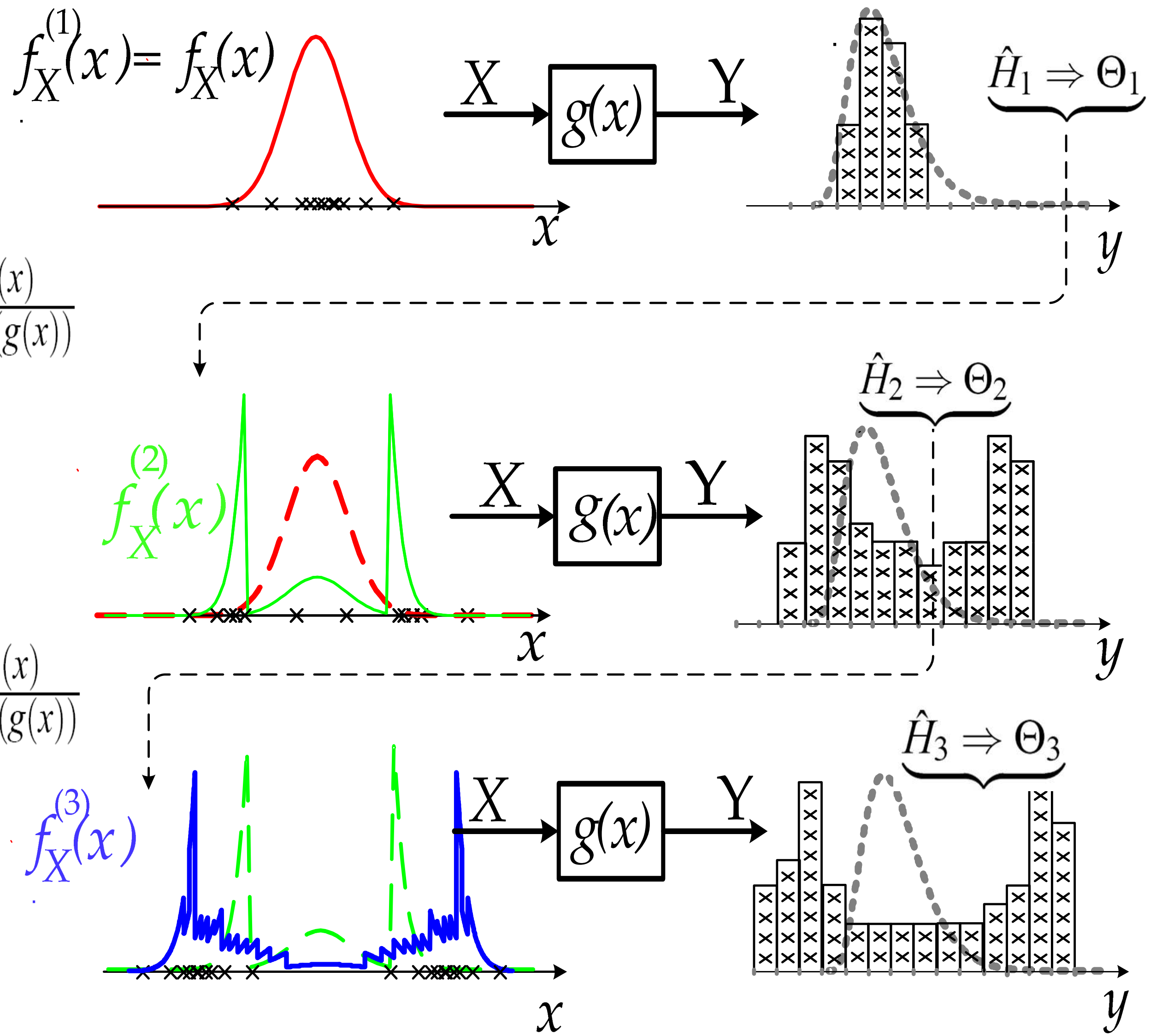
$$\Theta_{n+1}(y_i) = \underbrace{\left(\frac{N_i^{(n+1)}}{N}\right)}_{\hat{H}_{n+1}(y_i)} \underbrace{\left[\frac{1}{N_i^{(n+1)}} \sum_{n=1}^{N_i^{(n+1)}} w(X_n)\right]}_{c_n \Theta_n(y_i)}$$

UWIS

Intuition on MMC

Θ_0
Uniform

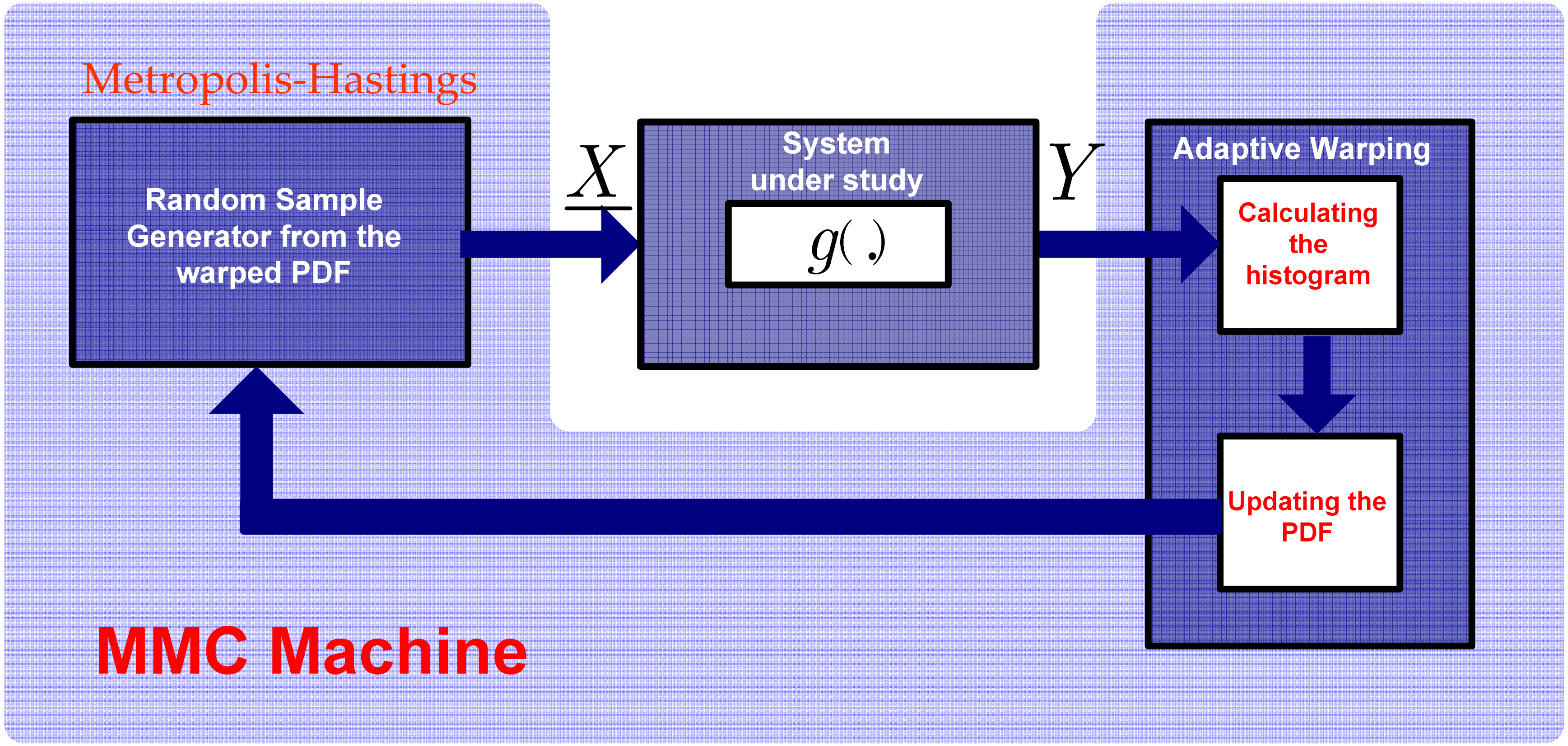
$$f_X^{(n+1)}(x) = \frac{f_X(x)}{c_n \Theta_n(g(x))}$$



$$f_X^{(n+1)}(x) = \frac{f_X(x)}{c_n \Theta_n(g(x))}$$

1 cycle is pure MC

MMC Block-diagram



MMC "engine": Metropolis-Hastings (MH)

MH is an algorithm that produces correlated samples $\{X_1, X_2, \ldots, X_N\}$ as a reversible Markov Chain whose steady-state distribution is the desired PDF $f^*_X(x)$.

At each time step t, if $X_{t-1}=x_p$ is the previous state, a next state is proposed as

$$x_n = x_p + U_t$$

where typically U_t is a uniform RV used to "explore" the state space around x_p..
The odds ratio is formed as

$$R = \frac{f_X^*(x_n)}{f_X^*(x_p)}$$

Then the proposal is accepted with probability max(1,R) and we set $X_t = x_n$.
Else the proposal is rejected and we keep the previous value: $X_t = x_p$.

MMC "engine": Metropolis-Hastings (MH)

Hence in cycle (*n+1*) of MMC, MH generation uses the odds ratio

$$R = \frac{f_X^{(n+1)}(x_n)}{f_X^{(n+1)}(x_p)} = \frac{f_X(x_n)}{\cancel{c_n}\Theta_n(g(x_n))} \cdot \frac{\cancel{c_n}\Theta_n(g(x_p))}{f_X(x_p)}$$

We make 2 important points:

1) the constant c_n cancels out and need not be computed

2) The UW warped PDF can be generated without knowledge of the domains D_i. R evaluated by computing $g(x_p)$, $g(x_n)$ and checking the bin they fall into.

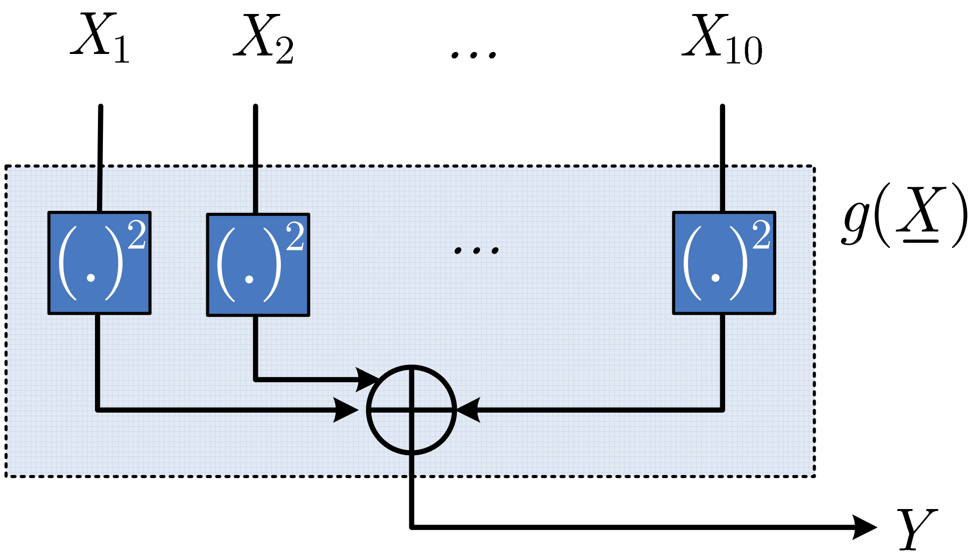
Example 1: Nonlinear Memoryless System

Input: a 10-dimensional vector of normal i.i.d RV's:

$$\underline{X} = \begin{bmatrix} X_1 & X_2 & \dots & X_{10} \end{bmatrix}$$

$$X_i \sim N(0,1)$$

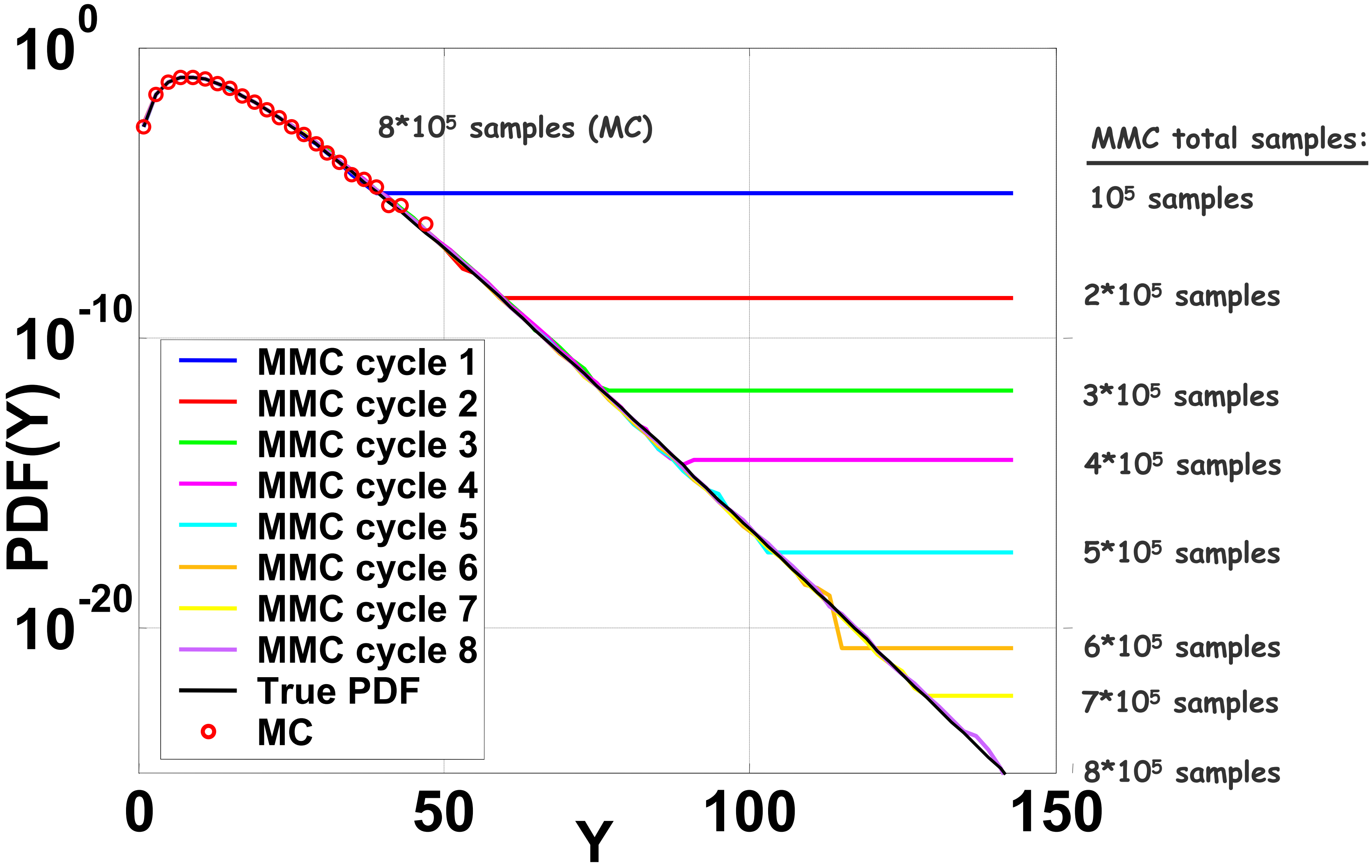
System: $g(\underline{X}) = \sum_{i=1}^{10} X_i^2$



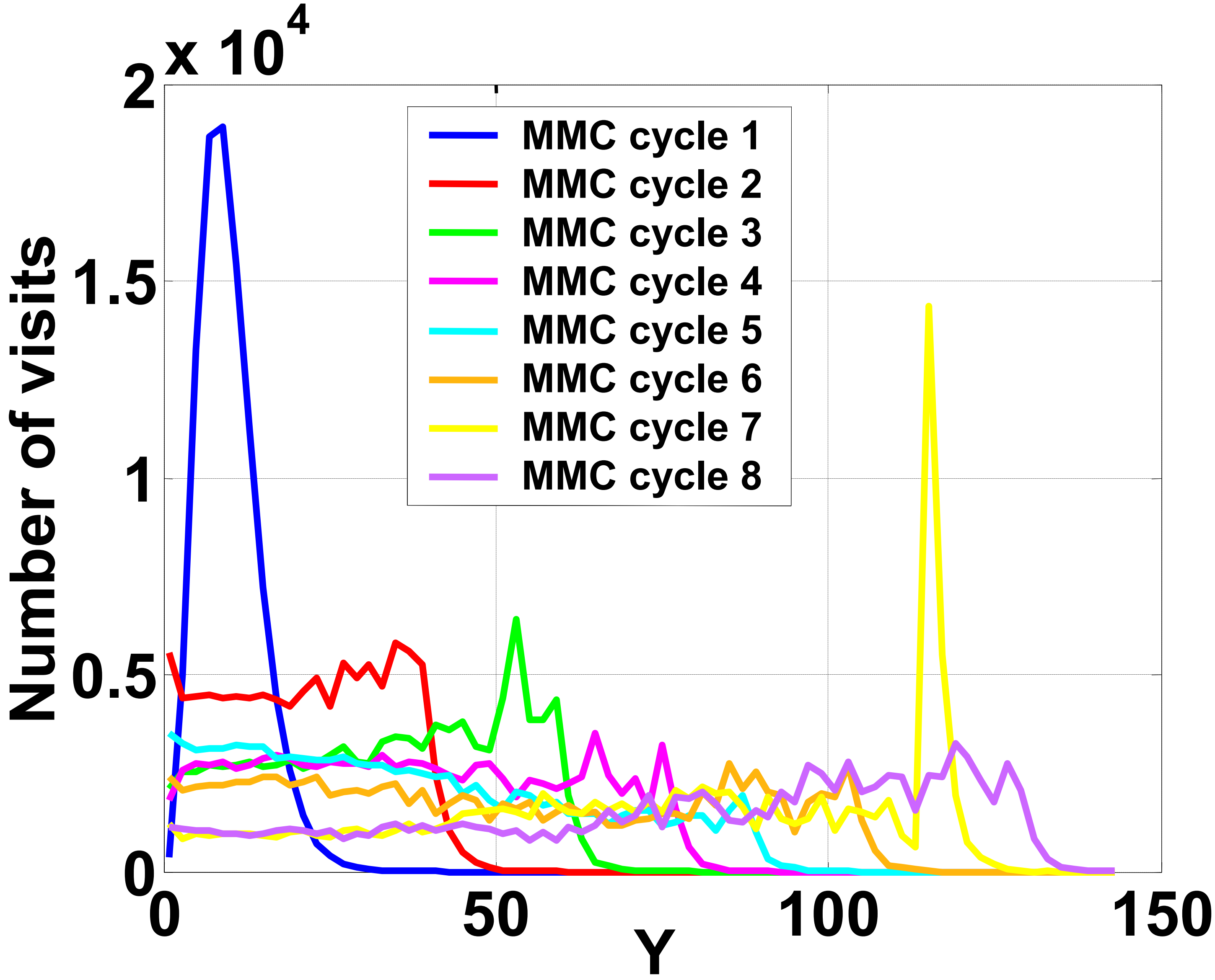
Output: $Y = g(\underline{X})$ of which we know the statistics:

$$Y \sim \chi^2(10)$$

Example 1: Standard MMC estimate



Example 1: Standard MMC Visits Histogram



Outline

- Motivation
- Monte Carlo (MC)
- Importance Sampling (IS)
- Flat Histogram (FH) Methods
    - Multicanonical Monte Carlo (MMC)
    - **Wang Landau (WL)**
    - Fast MMC
- Example: a PMD problem
- Conclusions

Wang-Landau (WL)

Wang-Landau updates at every time sample. Like MMC, it works in cycles, of variable duration. It uses a starting cycle parameter value $f_0 > 1$

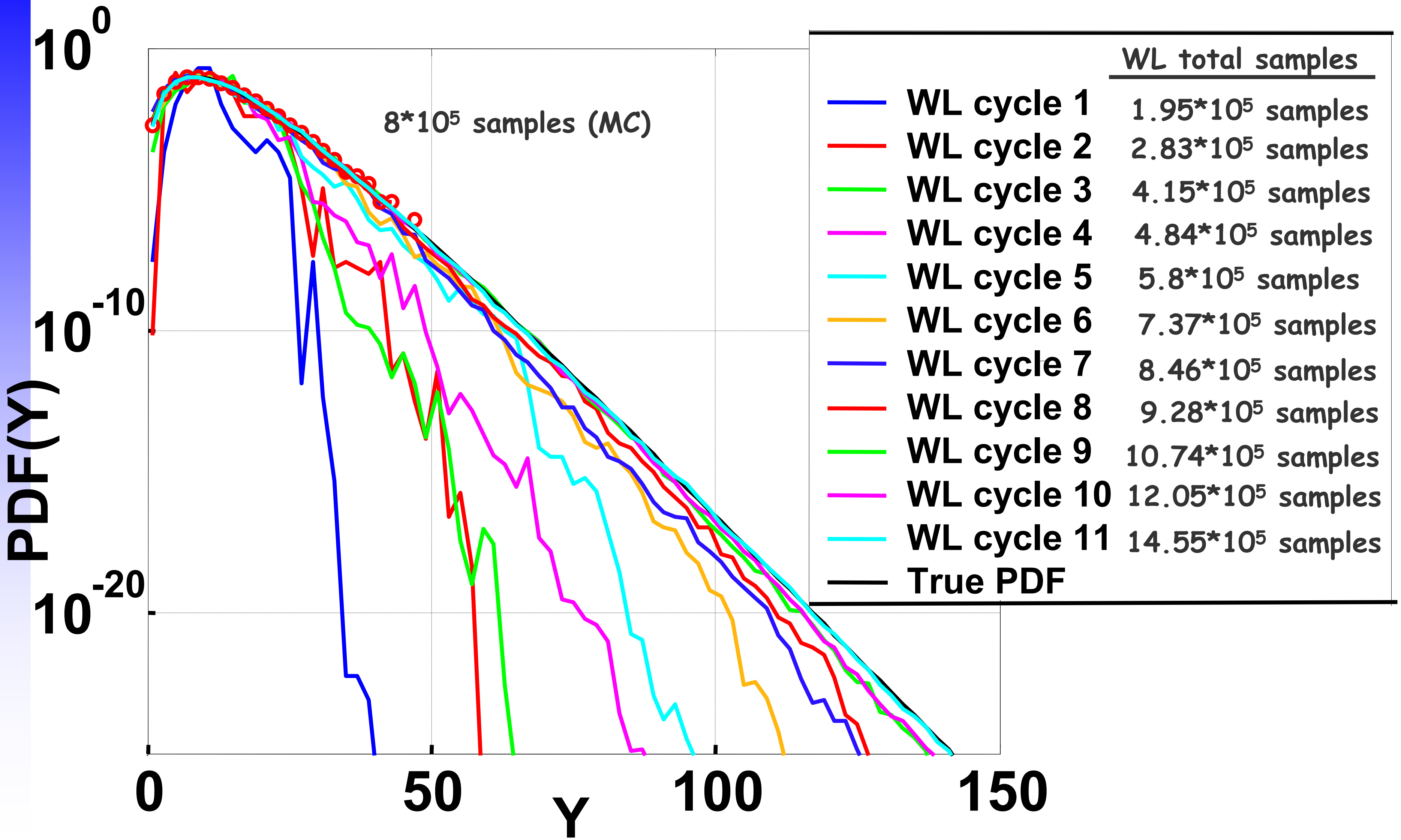
- **WL Algorithm**

0. At beginning of cycle m, reset the visits count and update the cycle precision parameter: $f_m = \sqrt{f_{m-1}}$
1. At time n of cycle m, draw a sample from $f_X^{(n)}(x) = \frac{f_X(x)}{c_{n-1}\Theta_{n-1}(g(x))}$
2. update estimate of PMF of Y as

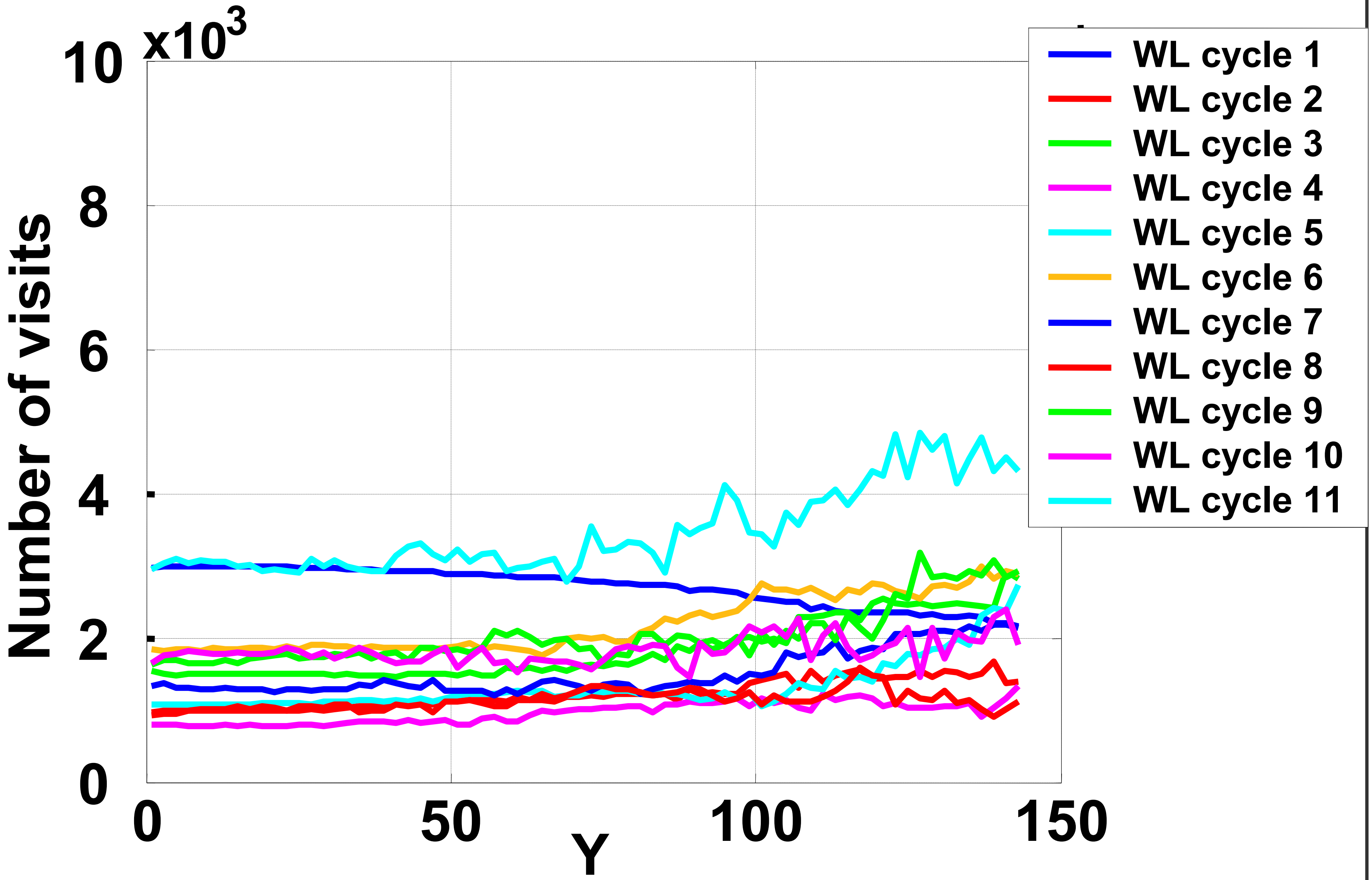
$$\Theta_n(y_i) = \begin{cases} f_m \cdot \Theta_{n-1}(y_i) & \text{if } g(X_n) \approx y_i \\ \Theta_{n-1}(y_i) & \text{else} \end{cases}$$

3. Update the visits histogram: increment by 1 count in visited bin.
4. If visits count is flat within desired tolerance (20%) go to next cycle m+1. else increment time and goto 1.

Example 1: WL estimate



Example 1: WL Visits Histogram



Outline

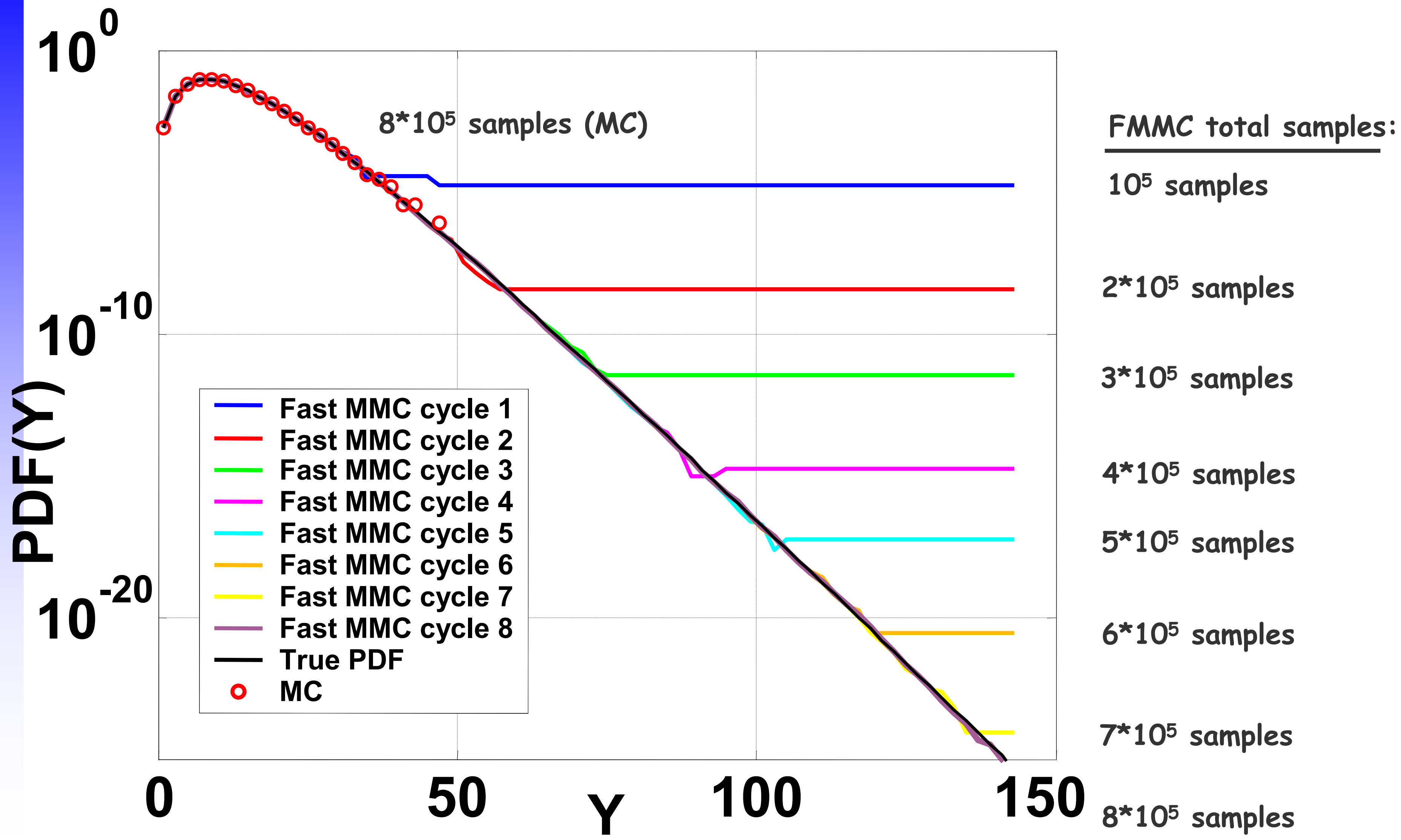
- Motivation
- Monte Carlo (MC)
- Importance Sampling (IS)
- Flat Histogram (FH) Methods
    - Multicanonical Monte Carlo (MMC)
    - Wang Landau (WL)
    - **Fast MMC**
- Example: a PMD problem
- Conclusions

Fast MMC

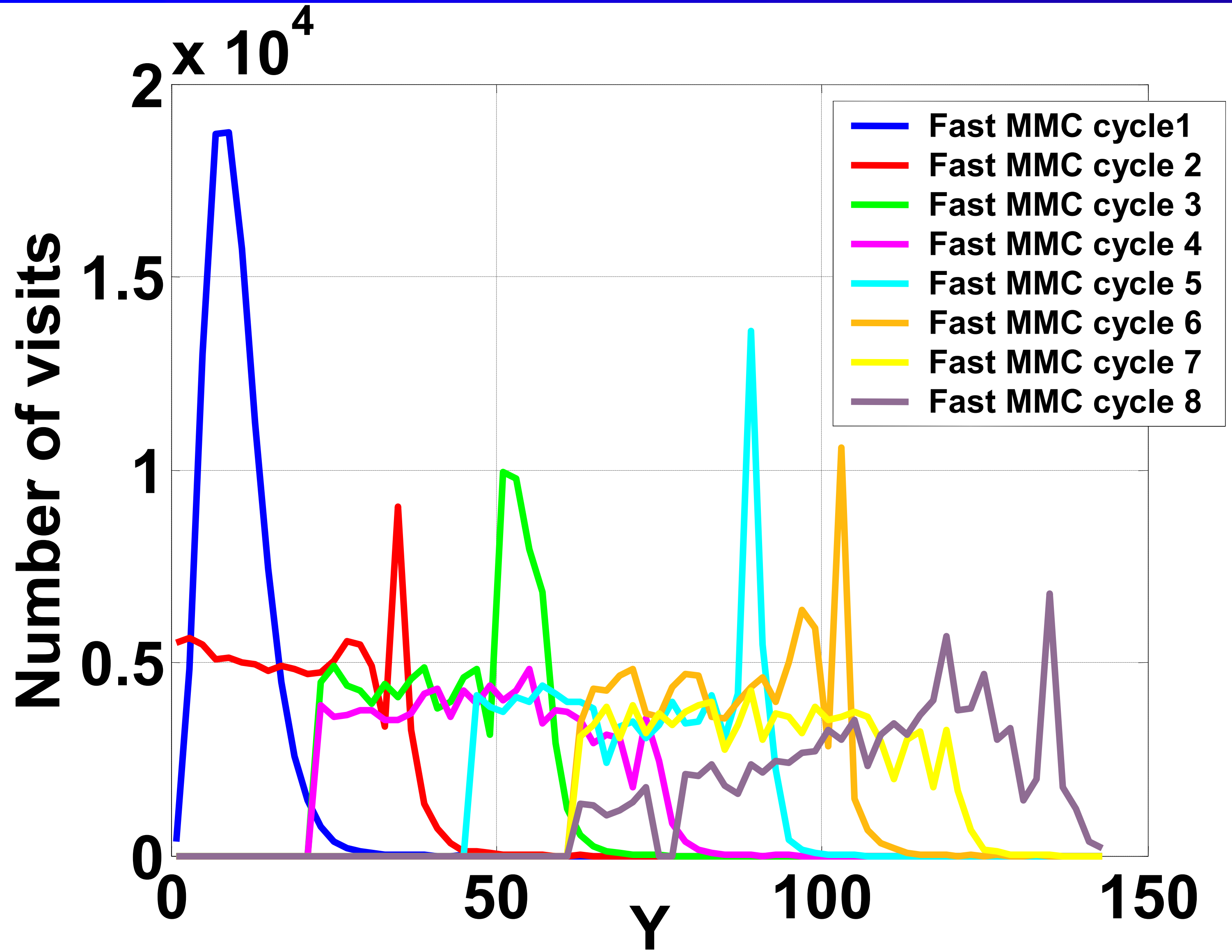
MMC has the evident drawback that modal region is visited and thus re-estimated at every cycle ⇒waste of samples!

We propose a novel algorithm (details in the proceedings) that prevents MMC from visiting regions of Y range over which a prescribed estimation precision has already been achieved in previous cycles.

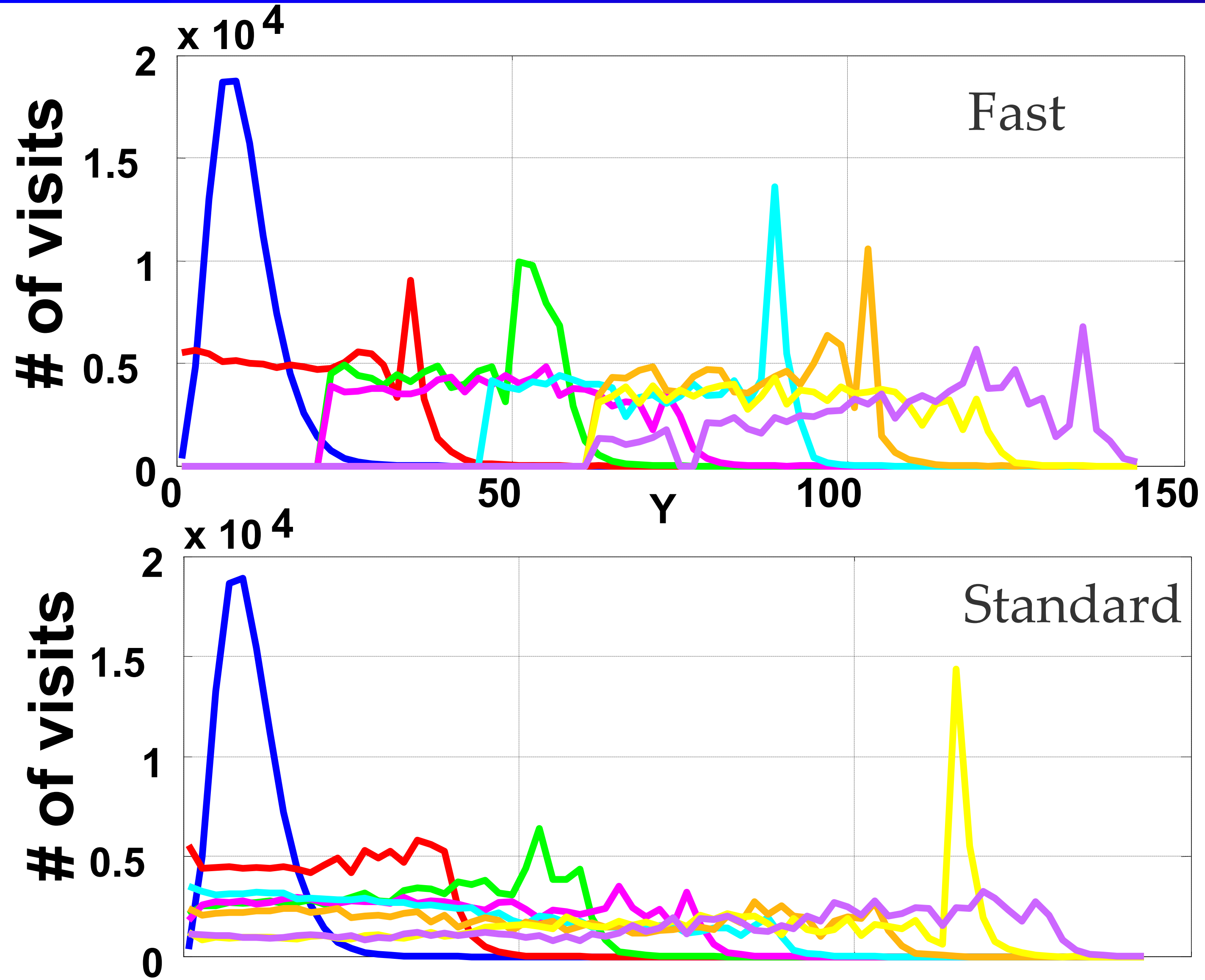
Example 1: Fast MMC estimate



Example 1: Fast MMC Visits Histogram



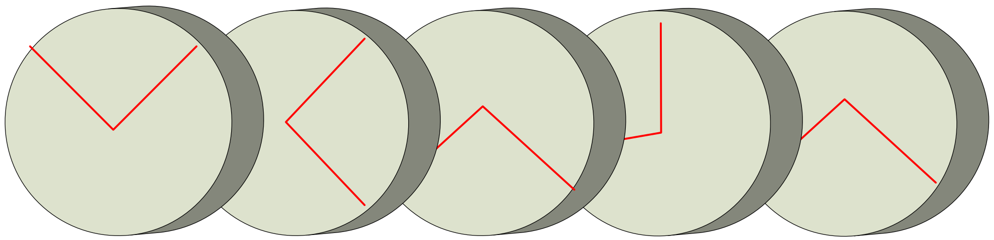
Example 1: Fast MMC vs Standard MMC



Outline

- Motivation
- Monte Carlo (MC)
- Importance Sampling (IS)
- Flat Histogram (FH) Methods
    - Multicanonical Monte Carlo (MMC)
    - Wang Landau (WL)
    - Fast MMC
- **Example: a PMD problem**
- Conclusions

Example 2: a PMD problem

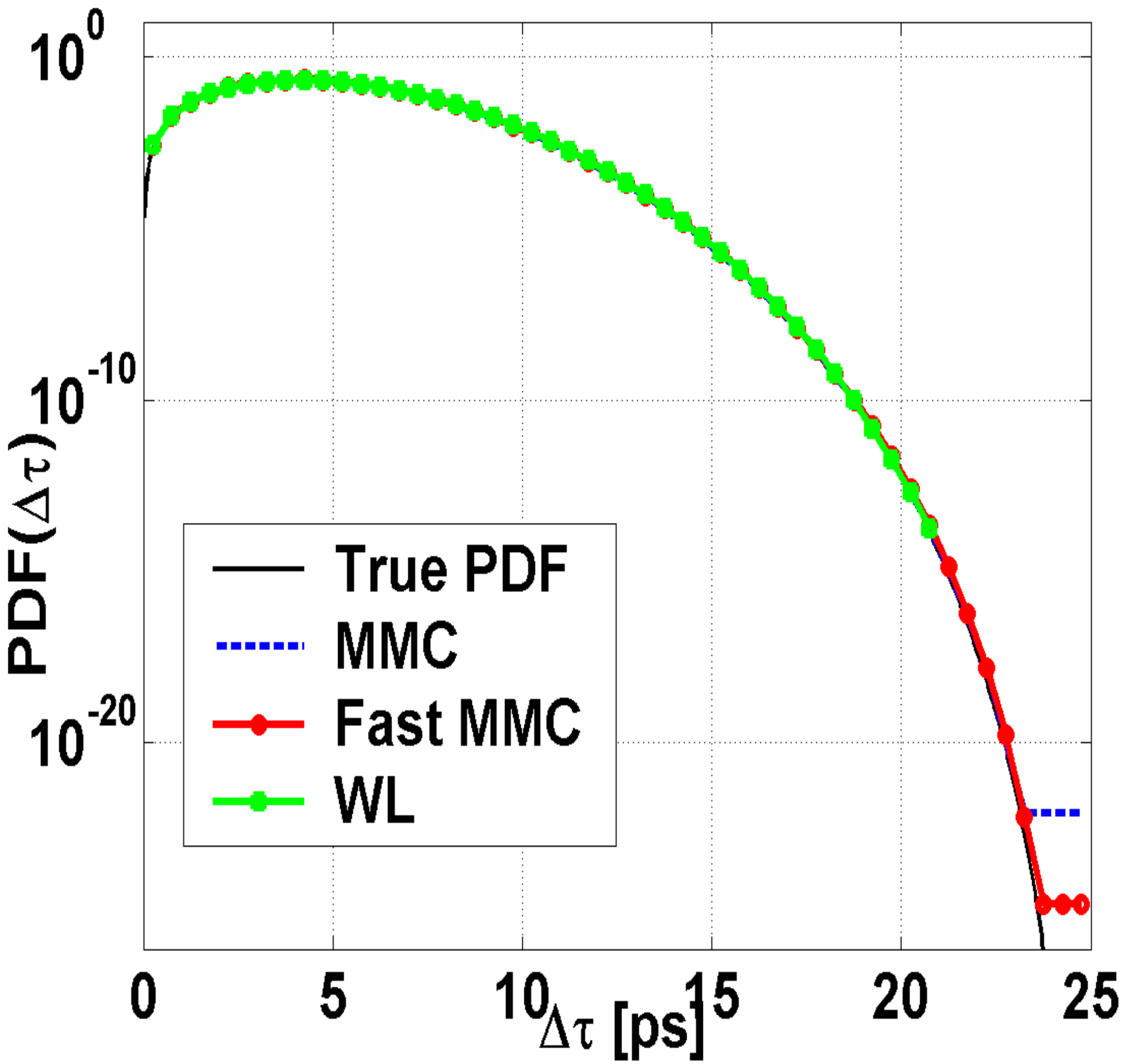


PMD fiber emulated with 25 PMF waveplates, each with a local DGD of 1 ps and uniformly distribted PSP over Poincarè sphere.

Exact PDF of global DGD is known in this problem.

We compare estimates from MMC/ fast MMC and WL over DGD range [0...25] (ps) using M=50 uniform bins for all algorithms.

Example 2: a PMD problem



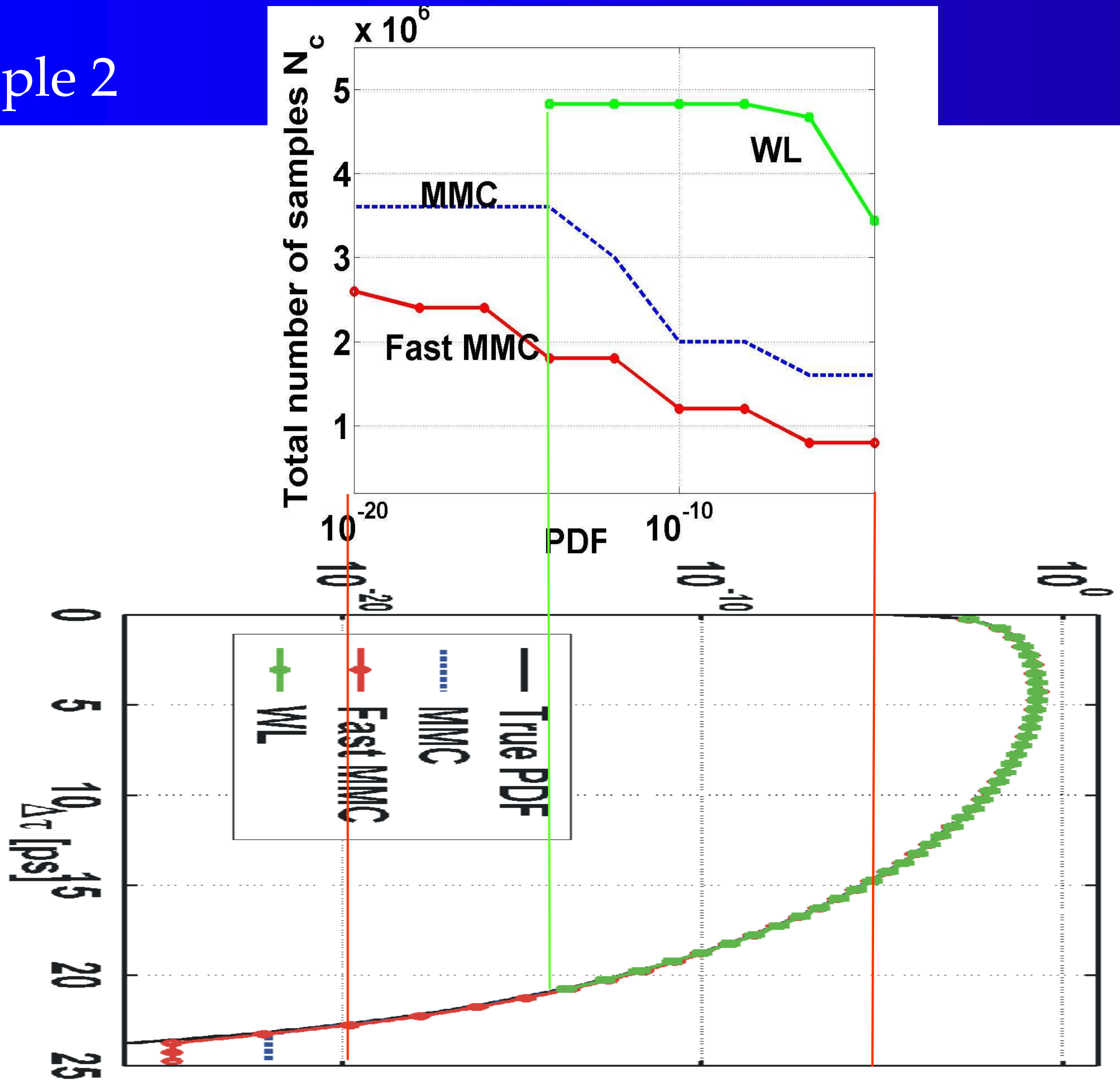
Here objective was to reach a relative error of < 5% down to 10^{-20}

MMC techniques use cycles of $2 \bullet 10^5$ samples.

We see that all 3 methods are accurate within specs, but WL could not go below 10^{-14} with this uniform binning

Next I show how many samples were needed to get these curves

Example 2



x 10^6
Total number of samples N_c
5
4
3
2
1
WL
MMC
Fast MMC
10^-20
10^-10
PDF
True PDF
MMC
Fast MMC
WL
Δτ [ps]
0
5
10
15
20
25

Conclusions

- MMC is an adaptive IS-based Flat Histogram algorithm.

  Doesn't need almost any knowledge of specific physical problem!

  This is major difference with IS

- Number of runs scales linearly with input dimensions if input RVs are independent. More dramatic scaling with correlated RVs.

- WL doesn't seem to offer any advantages over MMC. It needs lots of tricks (details usually not published) to properly work.

- Our fast MMC is good, but does not reduce runs by more than 50%. Reason is that Metropolis proposes to re-visit forbidden modal regions very often $\Rightarrow$ lots of rejections.

Many more details and lots of references at

www.tlc.unipr.it/bononi/ricerca/seminars/MMCcourse.pdf