# Adaptive channels for data analysis and importance sampling

André van Hameren

*Institute of Nuclear Physics, NCSR Demokritos, Athens, Greece*

andrevh@inp.demokritos.gr

Amsterdam, 2003

hep-ph/0301036

# For automatic importance sampling

Objective:

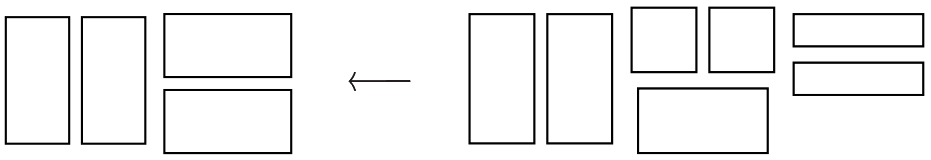
- usually only the channel *weights* are optimized in adaptive multi channeling[Kleiss, Pittau, 1994];
- we would like to adapt also the *channels*, for more flexability.

Means:

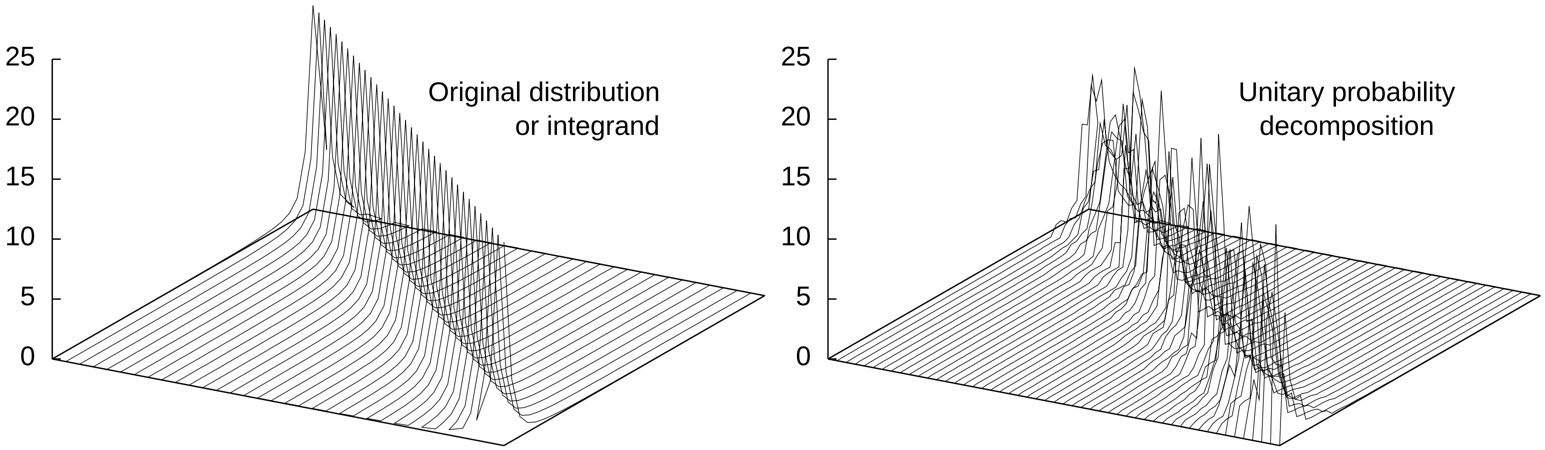
1. map the integration problem to the $s$-dimensional hypercube $[0,1]^s$ and dissect it into adaptive pieces (VEGAS[Lepage 1978], FOAM[Jadach, 2000]);
2. each piece corresponds to a channel with a weight to be optimized;
3. after each optimization step, dissect the piece with highest weight into new pieces.

Implementation: PARNI

- the pieces are boxes of the type $[a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_s, b_s]$;
- after each optimization step, the box with the highest weight is replaced by *all* overlapping pairs of boxes obtained by dissection along the Cartesian directions:



- restrict the number of channels by merging boxes with low weights.




| method | all $10^5$ data points | | | last $10^3$ data points | | |
|---|---|---|---|---|---|---|
| | result | error | eff. | result | error | eff. |
| variance optimization | 0.9977 | 0.0031 | 0.0156 | 1.004 | 0.014 | 0.259 |
| without optimization | 1.0015 | 0.0089 | 0.0588 | 1.093 | 0.093 | 0.059 |
| from u.p.d. | 0.9980 | 0.0016 | 0.0865 | 0.996 | 0.016 | 0.167 |

# For data analysis

Objective:

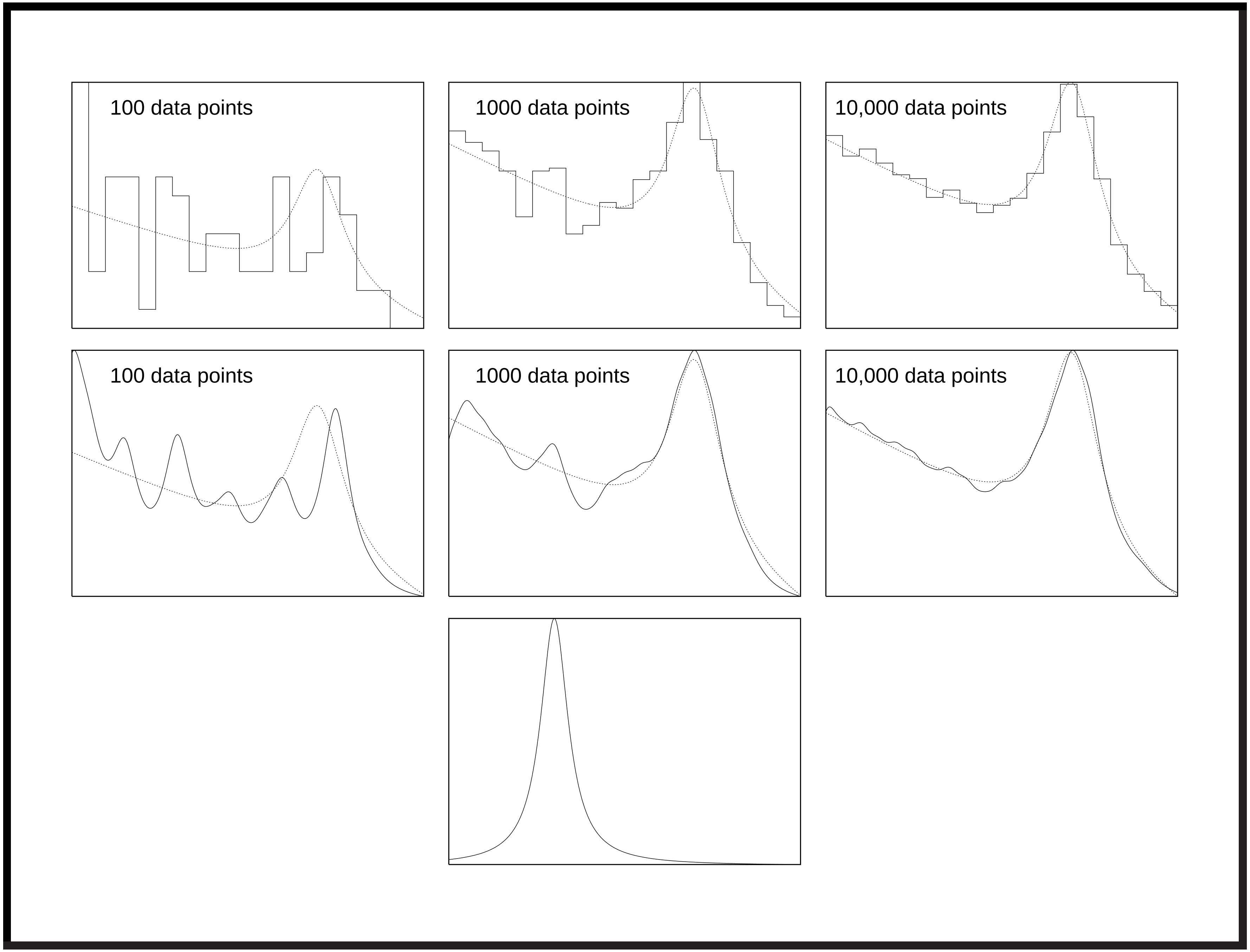
- derive a probability distribution from a sample of data;

Means:

- use channels $\longrightarrow$ 'generalized histogram' or unitary probability decomposition (u.p.d.);
- for importance sampling $\longrightarrow$ optimize the variance;
- for data analysis $\longrightarrow$ optimize the entropy $H = \int_\Omega P(\omega) \log G_x(\omega)\, d\omega$, with

$$P(\omega) = \textit{a priori} \text{ probability density} \quad , \quad G_x(\omega) = \text{u.p.d.} = \sum_{i=1}^{n} x_i g_i(\omega) \ .$$

- equivalent with maximum likelihood if $H \leftarrow \frac{1}{N} \sum_{j=1}^{N} \log G_x(\omega_j)$;
- solution: all $\langle g_i / G_x \rangle_P = \int_\Omega \frac{g_i(\omega)}{G_x(\omega)} P(\omega) d\omega$ should be equal;
- solve numerically, like for variance optimization:
  1. $y_i \leftarrow x_i \langle g_i / G_x \rangle_P$ for all $i = 1, \ldots, n$
  2. $x_i \leftarrow \frac{y_i}{\sum_{j=1}^{n} y_j}$ for all $i = 1, \ldots, n$


100 data points
1000 data points
10,000 data points
100 data points
1000 data points
10,000 data points

# The combination

Using u.p.d.s in integration problems

- with the Metropolis algorithm[Kharraziha, Moretti, 2000], one can calculate

$$\langle f \rangle_A := \frac{\int_\Omega f(\omega)\, A(\omega)\, d\omega}{\langle A \rangle}$$

  for any quadratically integrable $f$ and any integrable positive bounded $A$, but one cannot calculate

$$\langle A \rangle := \int_\Omega A(\omega)\, d\omega$$

- observe that

$$\langle A \rangle = \langle A^{1-\alpha} \rangle_{A^\alpha} \langle A^\alpha \rangle$$

  for any $\alpha \in [0, 1]$, and that the function $A^\alpha$ will be easier to integrate than $A$ itself, and

- *the estimation of $\langle A^{1-\alpha} \rangle_{A^\alpha}$ will serve a stream of data points distributed following $A^\alpha$ which can be used to create a u.p.d. for the estimation of $\langle A^\alpha \rangle$.*

Advantage:

- you find the peaks of the integrand immediately!

Example in phase space integration:

- integrate antenna-pole structures

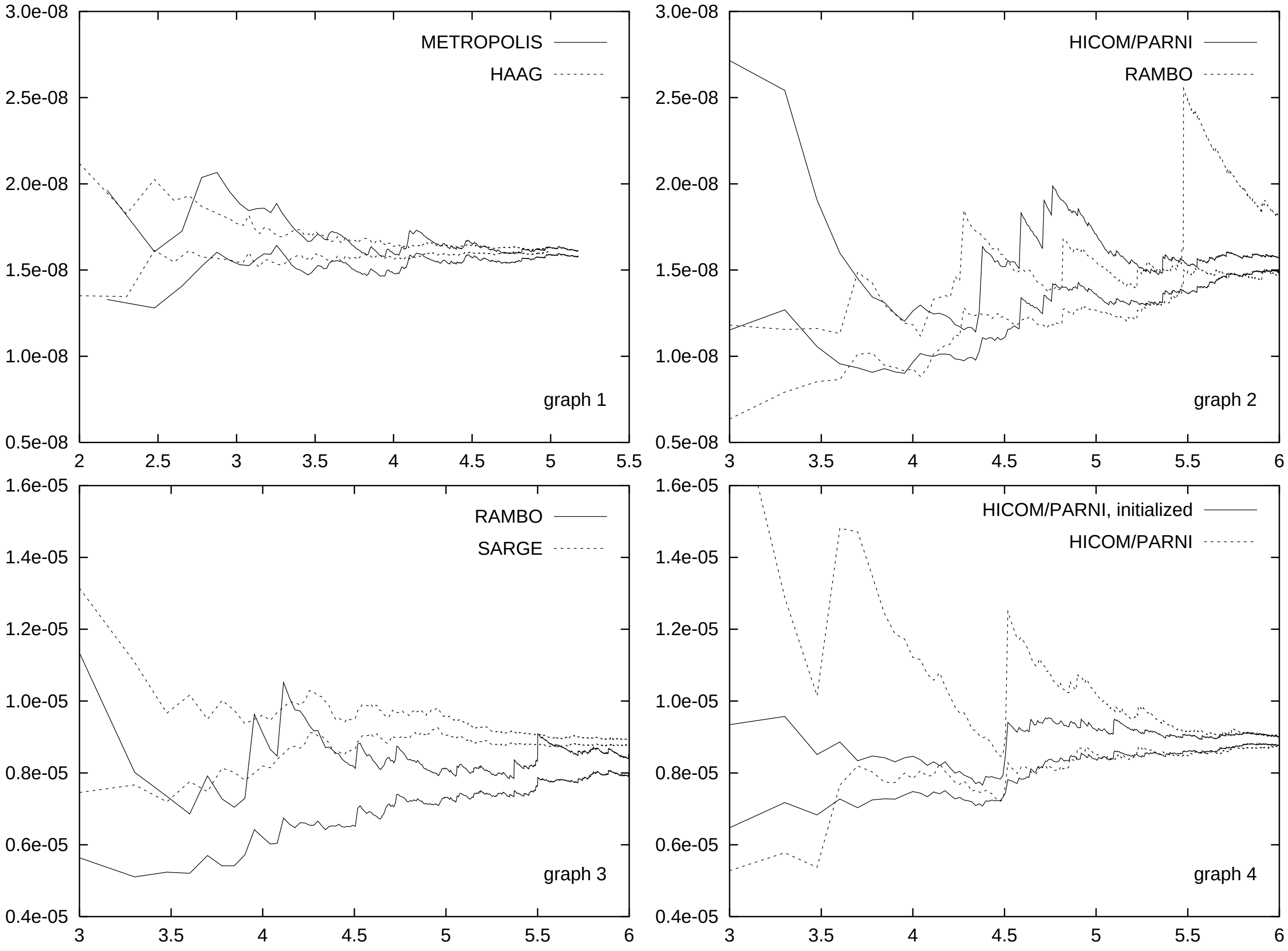
$$A = \sum_{\text{perm.}} \frac{(2\pi)^{4-3n}}{(p_0 \cdot p_1)(p_1 \cdot p_2)(p_2 \cdot p_3)\cdots(p_n \cdot p_{n+1})(p_{n+1} \cdot p_0)}$$

- calculation of $\langle A^{1-\alpha} \rangle_{A^\alpha}$ and initialization of `PARNI`:

$$\begin{tikzcd}
\boxed{\text{Metropolis with } A^\alpha} \arrow[r, "p"] \arrow[d, "p"] & \boxed{\mathtt{HICOM}^{-1}} \arrow[r, "\omega"] & \boxed{\mathtt{PARNI}} \\
\boxed{\text{integrand } A^{1-\alpha}} \arrow[r] & \boxed{\text{sum} \leftarrow \text{sum} + A^{1-\alpha}(p)} &
\end{tikzcd}$$

- integration of $\langle A^\alpha \rangle$:

$$\begin{tikzcd}
 & \boxed{\mathtt{PARNI}} \arrow[r, "\omega"] \arrow[dr, "w_P"] & \boxed{\mathtt{HICOM}} \arrow[r, "p"] \arrow[d, "w_H"] & \boxed{\text{integrand } A^\alpha} \arrow[dl, "A^\alpha(p)"] \\
w_P w_H A^\alpha(p) \arrow[ur] & & \boxed{\text{sum} \leftarrow \text{sum} + w_P w_H A^\alpha(p)} \arrow[ll] &
\end{tikzcd}$$




The process of convergence during the Monte Carlo calculation of $A$ for $n = 6$ massless momenta with $E_{\text{cm}} = 1000$ GeV and $s_{\text{cut}} = 450$ GeV$^2$. Along the horizontal axis runs $^{10}\log(\#\,\text{events})$. The curves give the average *plus* the standard deviation and the average *minus* the standard deviation.

| integrand | | algorithm | $I$ $10^{-8}$ | $\sigma$ % | $N$ $10^3$ | $\sigma^2 T_{tot}/T_{int}$ $10^3$ | $\sigma^2 N_{acc}$ $10^3$ |
|---|---|---|---|---|---|---|---|
| $A$ | 1 | Metropolis | 1.594 | 1.04 | 150 | 0.0134 | 0.0163 |
| | | `HAAG` | 1.617 | 0.713 | 100 | 0.0122 | 0.00139 |
| | 2 | `PARNI` | 1.528 | 2.58 | 1000 | 0.664 | 0.468 |
| | | `RAMBO` | 1.647 | 10.0 | 1000 | 8.03 | 7.44 |
| $(A)^{0.9}$ | 3 | `SARGE` | 885.6 | 0.866 | 1000 | 0.00424 | 0.00264 |
| | | `RAMBO` | 823.9 | 3.14 | 1000 | 0.765 | 0.732 |
| | 4 | `PARNI`, initialized | 889.4 | 1.32 | 1000 | 0.167 | 0.118 |
| | | `PARNI` | 886.8 | 1.69 | 1000 | 0.273 | 0.198 |

The final results corresponding with the graphs. $I$ is the integral, $\sigma$ the standard deviation, $N$ the number of generated events, $T_{tot}$ the total computation time, $T_{int}$ the time it takes to perform one evaluation of the integrand and $N_{acc}$ the number of accepted (non-zero weight) events.