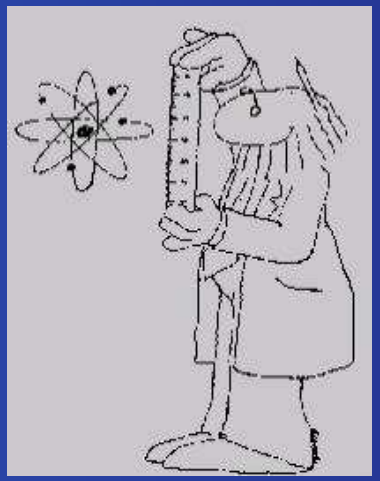
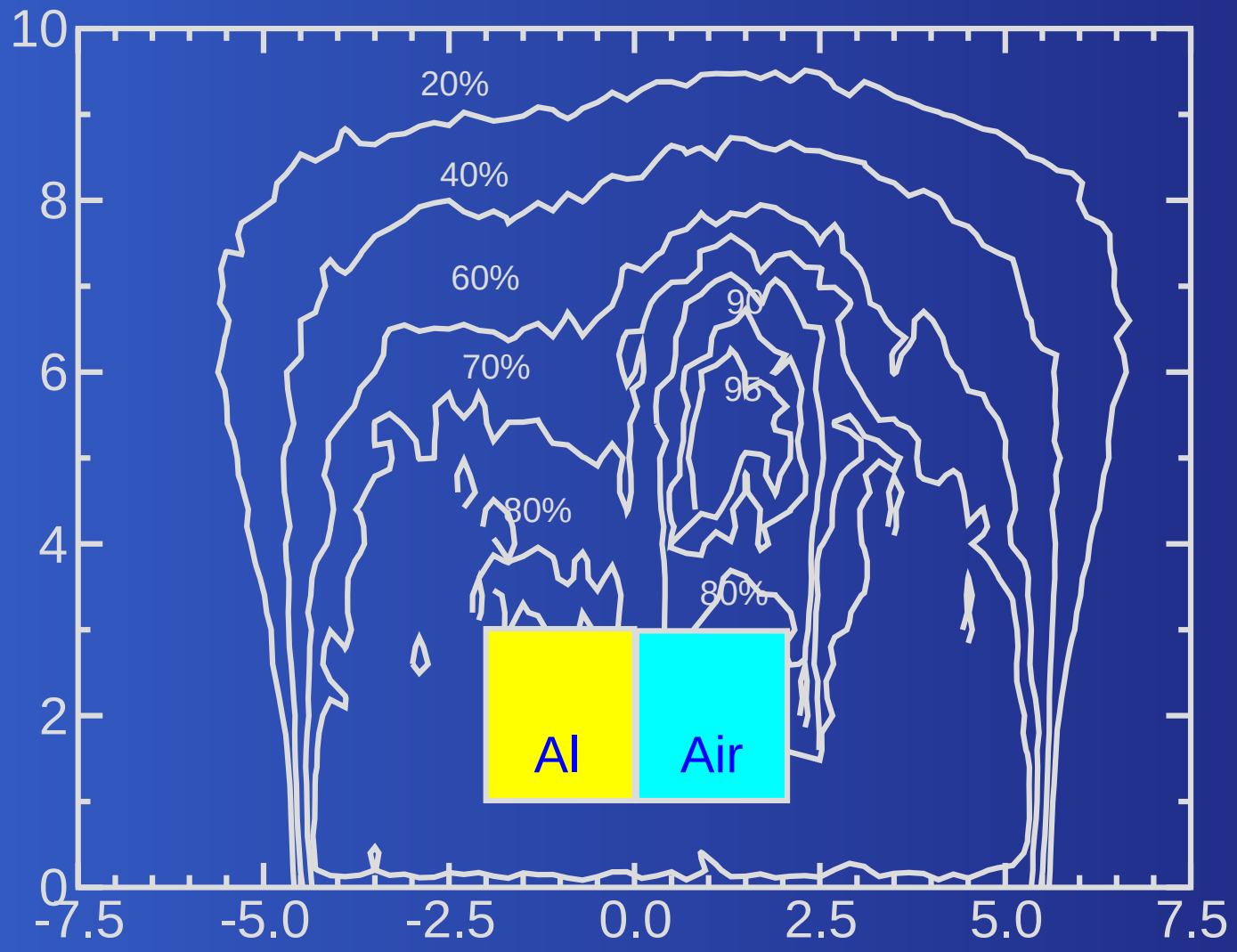
Monte Carlo treatment planning: interpretation of noisy dose distributions and review of denoising methods

Iwan Kawrakow

Ionizing Radiation Standards, NRC, Ottawa, Canada

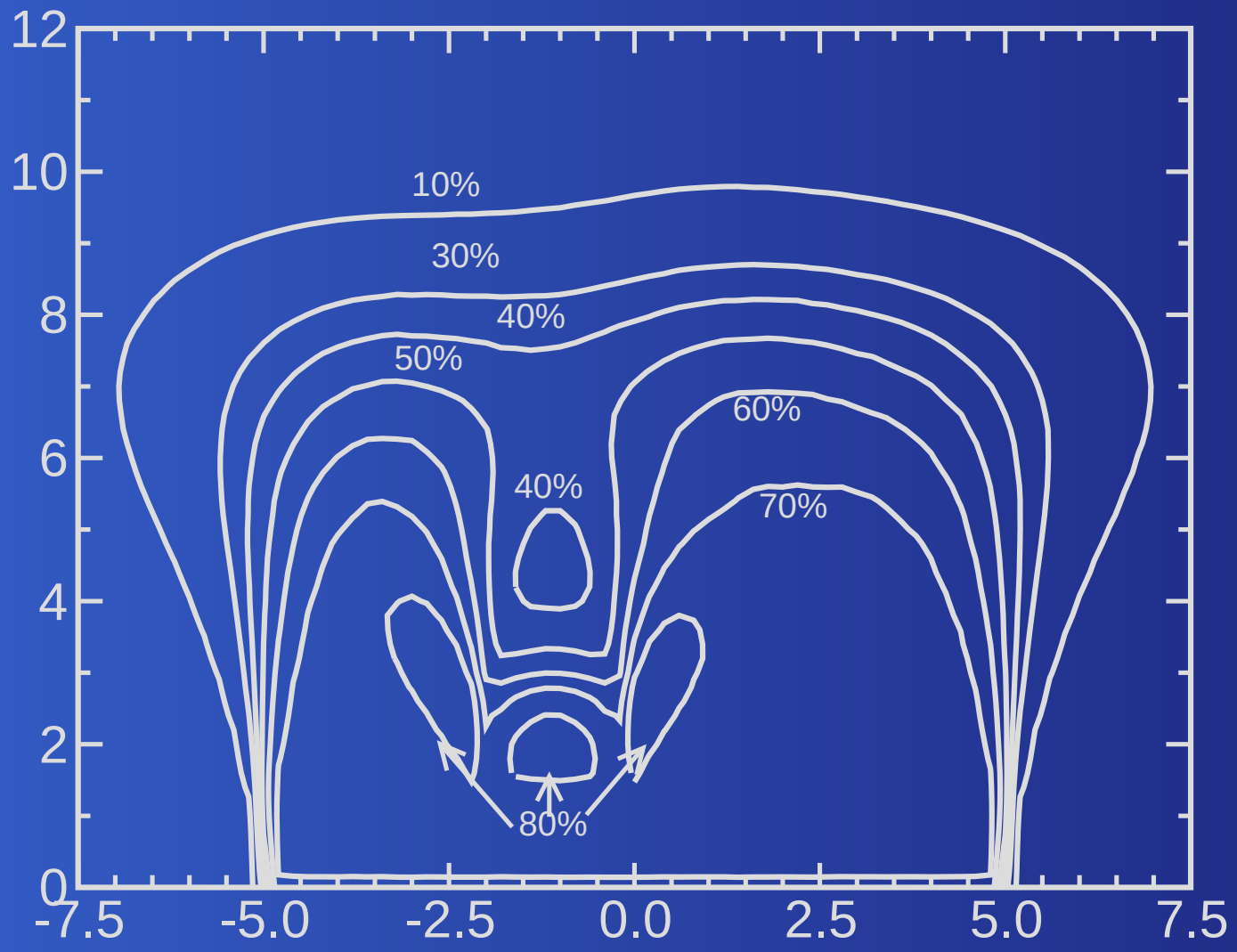




20 MeV e^-

5 mm voxels

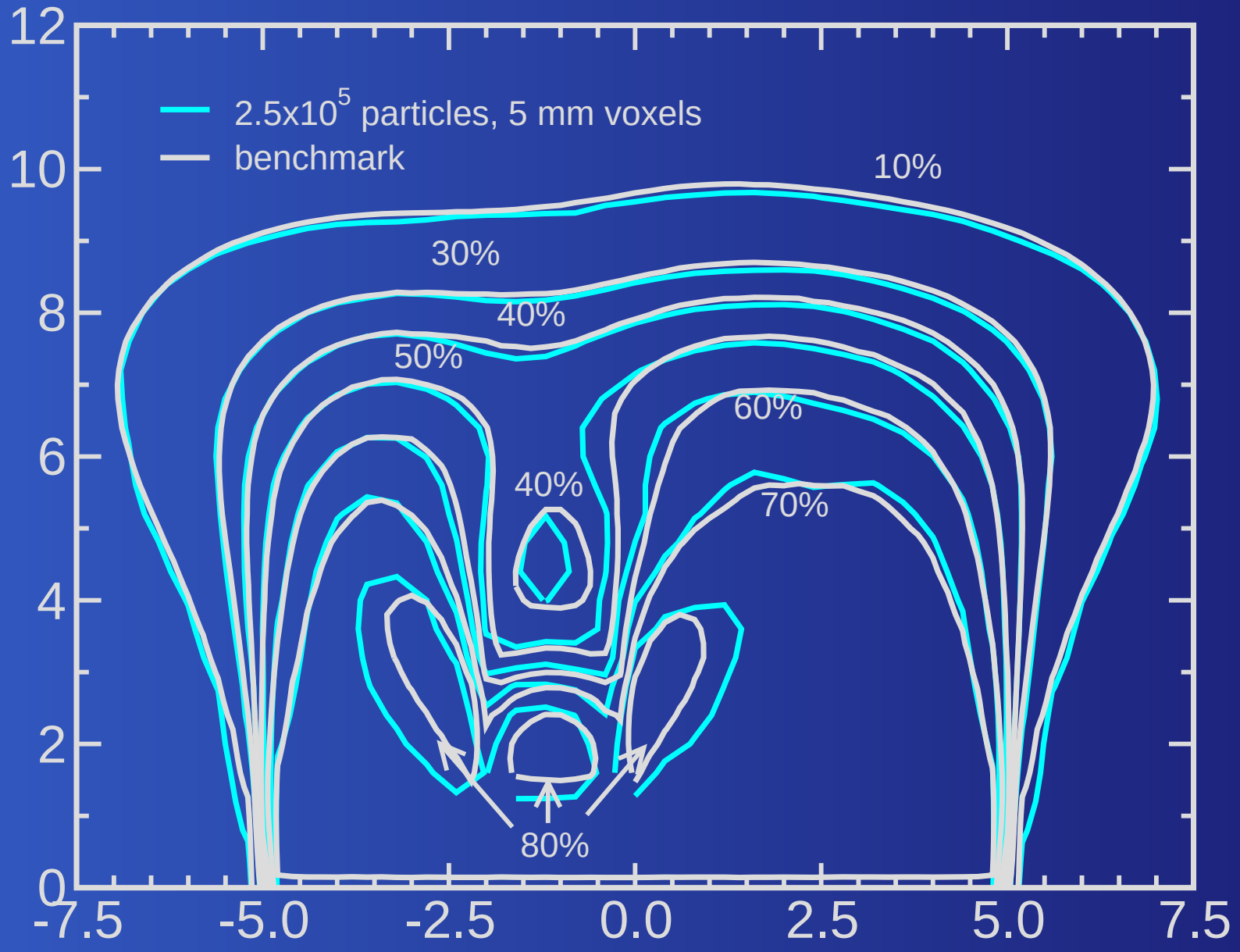
2 sec. of CPU time using VMC++ for a complete 3D dose calculation



20 MeV e^-

2.5 mm voxels

100 min. of CPU time using VMC++ for a complete 3D dose calculation


2.5x10^5 particles, 5 mm voxels
benchmark
10%
30%
40%
50%
60%
40%
70%
80%
12
10
8
6
4
2
0
-7.5
-5.0
-2.5
0.0
2.5
5.0
7.5

Contents

- Stochastic dose distributions
- Interpretation of noisy dose distributions
- Testing for systematic bias introduced by denoising
- Denoising integrated dose tallies
  - DVHs
  - Other integrated dose tallies
- Dose distribution denoising methods
  - Wavelet threshold denoising (WTD)
  - Locally-adaptive Savitzky-Goley (LASG) filter
  - Anisotropic diffusion (AD)
  - Iterative reduction of noise (IRON)
- Denoising examples
- Conclusions

Stochastic dose distributions

- The Monte Carlo (MC) technique is a stochastic integration method $\Rightarrow$ MC calculated quantities are subject to statistical uncertainties
- According to the Central Limit Theorem, for a large number N of simulated particle tracks, the statistical uncertainty of the dose distribution will approach zero as $N^{-1/2}$

$\Rightarrow$ One must simulate an infinite number of histories for a zero uncertainty

$\Rightarrow$ Any MC-calculated dose distribution will be a noisy representation of the true dose distribution

- This is a unique feature and requires special consideration

Statistical uncertainties

The statistical uncertainty $\sigma(D)$ on a dose value D is proportional to $D^{1/2}$ (Sempau and Bielajew 2000, Keall *et al* 2000)

$\Rightarrow$ Uncertainties are not the same in all voxels

$\Rightarrow$ Doses near $D_{\max}$ have the smallest *relative* uncertainties but the largest *absolute* uncertainties

$\Rightarrow$ Low doses have larger *relative* uncertainties and smaller *absolute* uncertainties

$\Rightarrow$ One must precisely specify the convention adopted to specify the uncertainty of a MC dose distribution with a single number

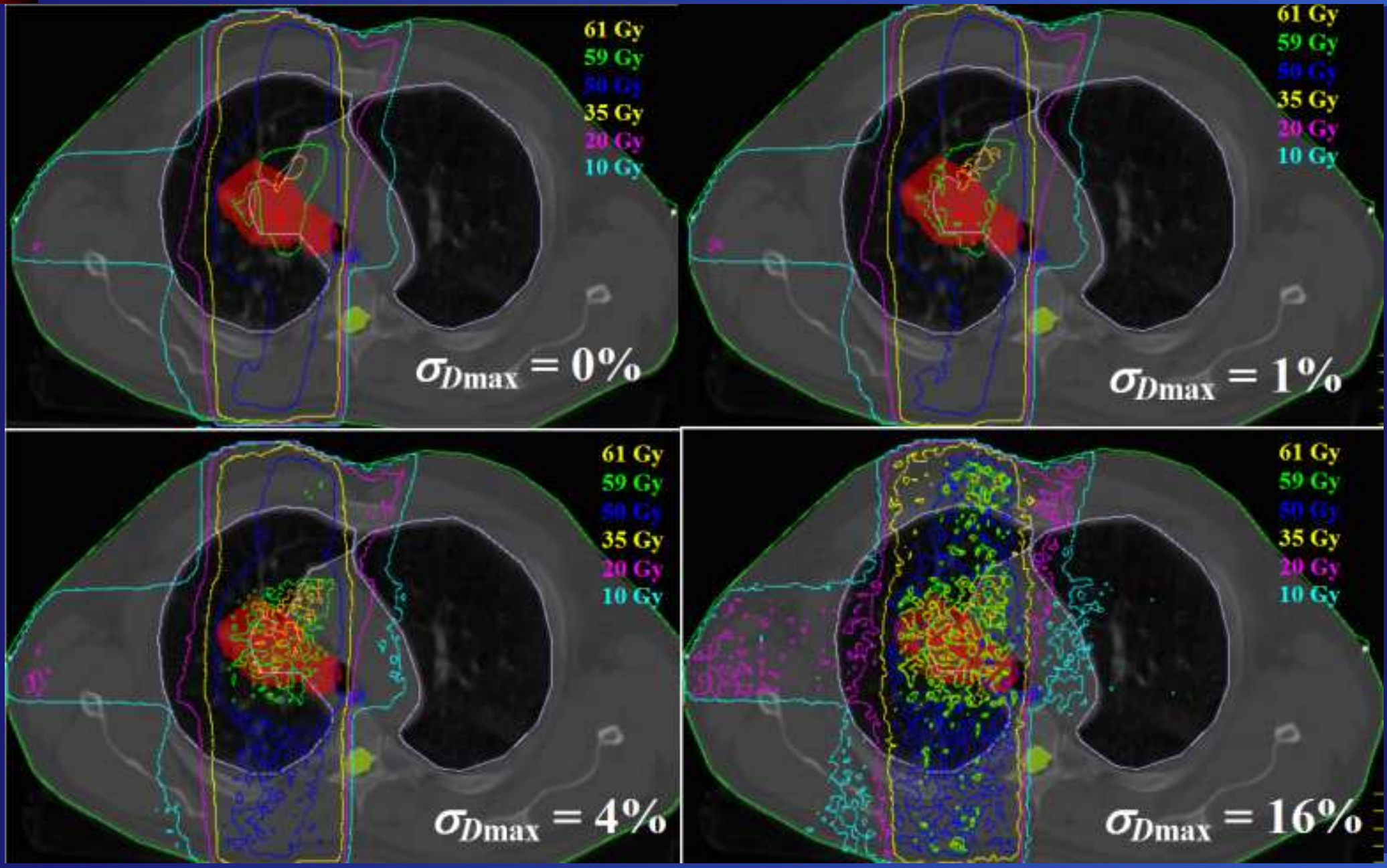
- x% of $D_{\max}$?
- x% relative uncertainty?
- Uncertainty on a single dose value or some average uncertainty?

Interpretation of noisy dose distributions

The statistical uncertainties present in MC-computed dose distributions affect all aspects of the treatment plan evaluation process:

- Isodose representations
- DVHs
- Maximum and minimum dose in a volume
- Dose metrics such as TCP, NTCP, EUD
- Cost functions used for treatment plan optimization

Isodose representations



Isodose representations

- There is always jitter in isodose representations of MC computed distributions, even for $\sigma \sim 1\%$
- The general consensus is the $\sigma \sim 2\%$ is sufficient for treatment planning purposes (but this is, of course, subjective)
- Dose delivery accuracy is limited to a few percent $\Rightarrow$ the MC computed dose distribution is a more realistic representation of the actually delivered dose
- The jitter in the MC isodose representation can be used to open a dialog with the planning team about realistic expectations on the dose delivery accuracy

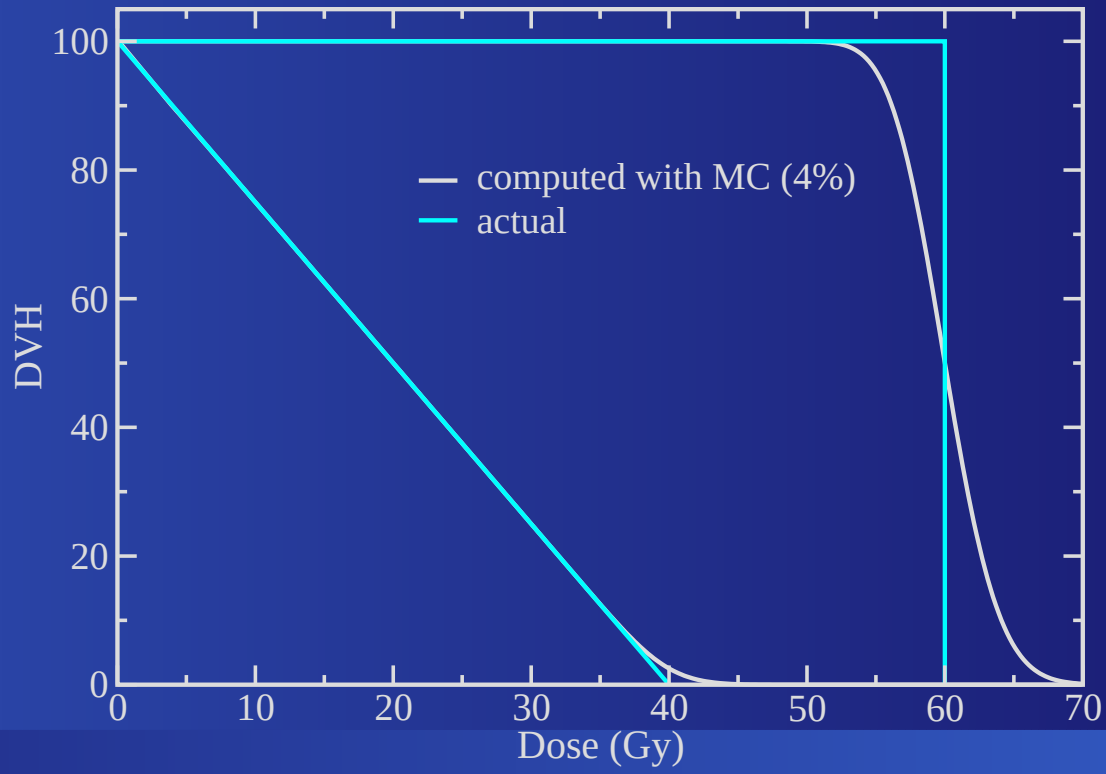
DVHs

The actually observed DVH, $\tilde{p}(D)$, is the convolution of the true DVH $p(D)$ with the statistical noise:

$$\tilde{p}(D) = \int \mathrm{d}D' p(D') \frac{\exp\left[-(D'-D)^2/2\sigma^2(D')\right]}{\sqrt{2\pi}\sigma(D')}$$

Due to
their typical shapes, the target DVH is normally more affected than the DVH of organs at risk

One can use the above equation to remove (or at least reduce) the noise from the DVHs



Integrated dose tallies

such as TCP, NTCP, EUD, etc., are also affected by the statistical noise.

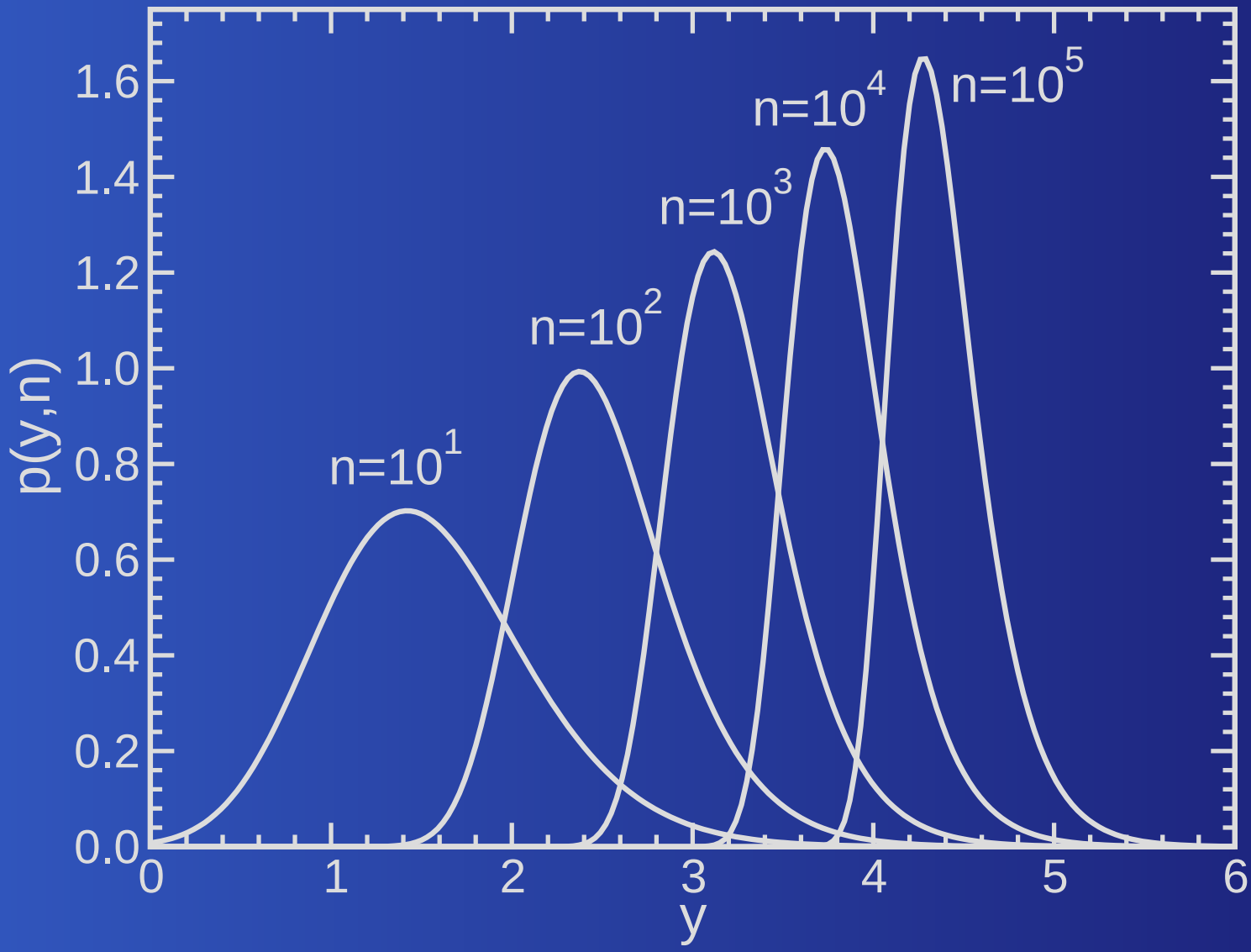
- The effect is typically less compared to isolines or target DVHs
- It has been observed empirically that the effect is negligible for $\sigma \lesssim 2\%$
- It has been shown theoretically that *any* tally (or cost function) will differ from its true value in a *systematic* way
- Fortunately, the systematic error is proportional to σ^2 and therefore decreases faster towards zero (as $1/N$) than the dose uncertainties
- The theoretical result can be used to remove the systematic uncertainty from the integrated dose tally

Maximum/Minimum dose

- The maximum or minimum dose in a region of interest (ROI) is the dose metric most affected by statistical noise
- Consider a ROI with n voxels with a uniform dose distribution $\bar{D}$. A MC-computed dose distribution will fluctuate around $\bar{D}$.
- The probability $p(y,n)$ that the maximum or minimum dose in the ROI found in the MC distribution differs by y standard deviations from $\bar{D}$ is given by

$$p(y,N) = \frac{\exp(-y^2/2)}{\sqrt{2\pi}} n \left(\frac{1+\mathrm{Erf}(y/\sqrt{2})}{2} \right)^{n-1}$$

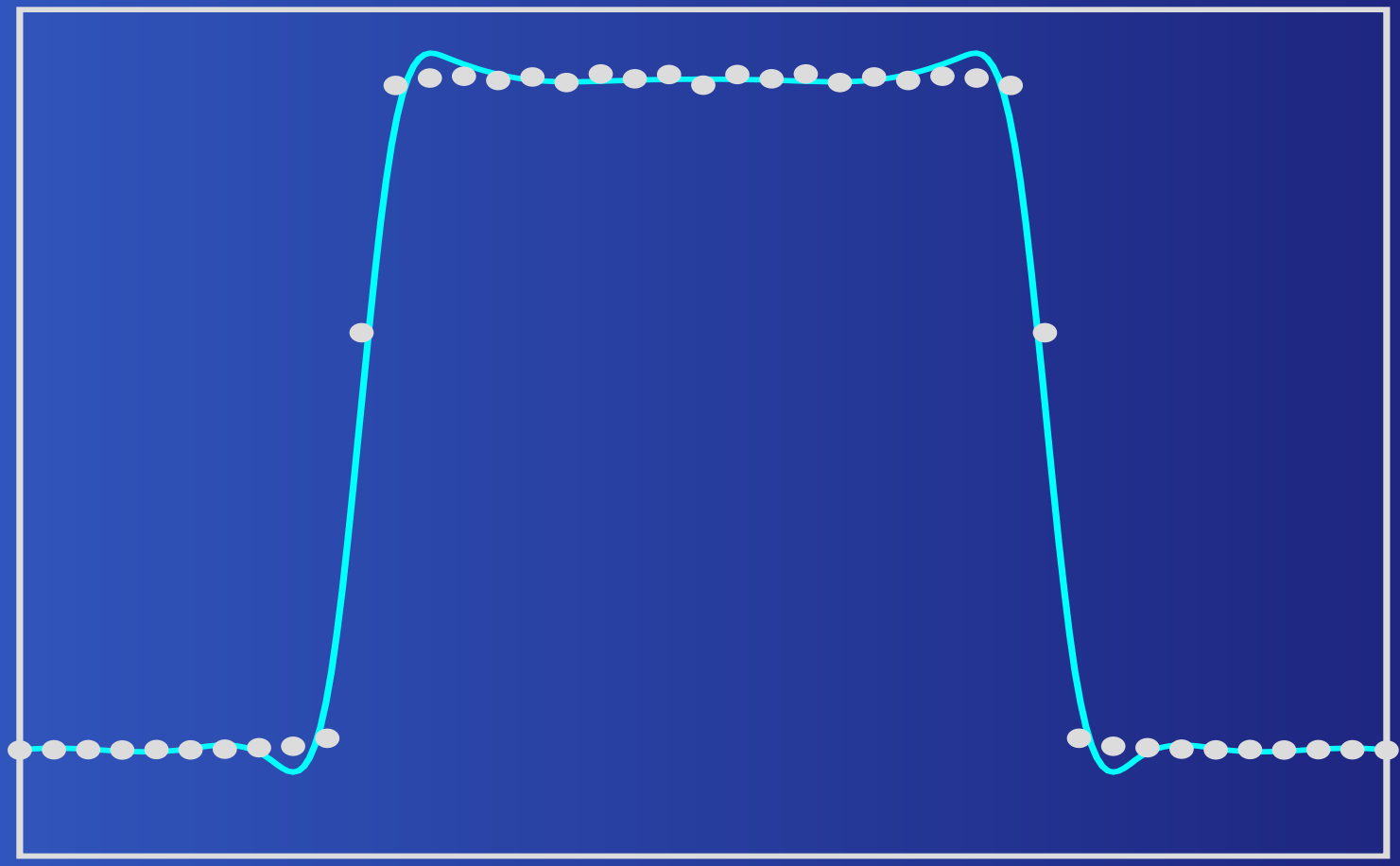
$p(y,n)$



⇒ One should *never* use the maximum or minimum dose in a ROI for dose prescription purposes

Approaches for reducing statistical noise

- Simply simulate more particles. This is not always practical: one needs 4 times more CPU time to reduce the uncertainty by a factor of 2
- Approaches for denoising integrated dose tallies
  - DVHs: Alternatives methods proposed in 2000 by Sempau & Bielajew and by Jiang, Pawlicki & Ma
  - Cost functions: Kawrakow (2004)
- Approaches for denoising dose distributions
  - First attempt by Deasy in 2000
  - More sophisticated methods developed during 2002–2003
  - Comprehensive comparison between algorithms by El Naqa *at al* (2005)
- Denoising = random uncertainty $\rightarrow$ systematic uncertainty ?



- Careless application of denoising leads to artifacts
- Quantitative tests for evaluating the performance of denoising algorithms are required

Denoising performance evaluation

Compare a denoised dose distribution to a benchmark distribution computed with the same MC algorithm to a negligible statistical uncertainty using the following 5 criteria

- Visual inspection of isodose lines
- Difference area between the resulting DVHs
- Mean square difference
- A x%/y mm test
- The maximum difference to the benchmark

Denoising DVHs

$$\tilde{p}(D) = \int \mathrm{d}D' p(D') \frac{\exp\left[-(D'-D)^2/2\sigma^2(D')\right]}{\sqrt{2\pi}\sigma(D')}$$

The above is a Fredholm equation of the first kind for the unknown function $p(D)$. Written in discrete form for the I bins of the DVH:

$$\tilde{P} = G \cdot P$$

$$G_{ij} = \frac{1}{\Delta D} \int_{\text{bin } i} \mathrm{d}D \int_{\text{bin } j} \mathrm{d}D' \frac{\exp\left[-(D-D')^2/2\sigma^2(D)\right]}{\sqrt{2\pi}\sigma(D)}$$

Easy solution:

$$P = G^{-1} \cdot \tilde{P}$$

Unfortunately, inverting G is numerically unstable $\Rightarrow$ other methods are needed

Denoising DVHs: Sempau & Bielajew method

Obtain an approximate solution P'

$$P' = Q^{-1} \cdot G^T \cdot \tilde{P}$$

where the matrix Q,

$$Q = G^T \cdot G + aD_2^T \cdot D_2$$

is well behaved (invertible). In the above D_2 is the discrete second order derivative operator and X^T is the transpose of X. Obviously

$$P' \to P \quad \text{as} \quad a \to 0$$

Strategy: use minimum a that permits Q to be reliably inverted

Denoising DVHs: Jiang *et al* method

Strategy: solve the Fredholm equation iteratively ($p^{(j)}(D)$ is the estimate of the true DVH $p(D)$ in the j'th iteration)

1. Calculate

$$\tilde{p}^{(j)}(D) = \int \mathrm{d}D' p^{(j)}(D') \frac{\exp(-(D-D')^2/(2\sigma^2(D))}{\sqrt{2\pi}\sigma(D)}$$

2. Terminate the iteration if $\chi^2_j \leq \varepsilon$ or $|\chi^2_j - \chi^2_{j-1}| < \varepsilon$, where

$$\chi^2_j = \sum_{i=1}^{I} \left[\tilde{p}^{(j)}(D_i) - \tilde{p}(D_i)\right]^2 ,$$

3. Set $j = j+1$, compute $p^{(j)}(D)$, and go to step 1, where

$$p^{(j)}(D) = p^{(j-1)}(D) + \gamma(D)\left[\tilde{p}^{(j-1)}(D) - \tilde{p}(D)\right]$$

Reducing the uncertainty of a cost function

Consider a dose distribution $\{\bar{D}_i\}$, a MC calculation of $\bar{D}_i$ with doses $\{D_i\}$ and uncertainties $\{\sigma_i\}$, and a "cost function" $F(\{D_i\})$. The probability distribution $p(f)$ that the cost function F has a certain value f resulting from the MC doses $\{D_i\}$ is a Gaussian:

$$p(f) = \frac{1}{\sqrt{2\pi}\Delta f} \exp\left(-\frac{(f - F(\{\bar{D}_i\}) - b)^2}{2\Delta f^2}\right) ,$$

i.e. the expectation value of f is *systematically* different from $F(\{\bar{D}_i\})$ computed using the true doses $\{\bar{D}_i\}$. Fortunately, the systematic difference b is known and can be removed thus reducing the uncertainty on F:

$$b = -\frac{1}{2}\sum_i \frac{\partial^2 F}{\partial D_i^2}\sigma_i^2 + O(\sigma_i^4)$$

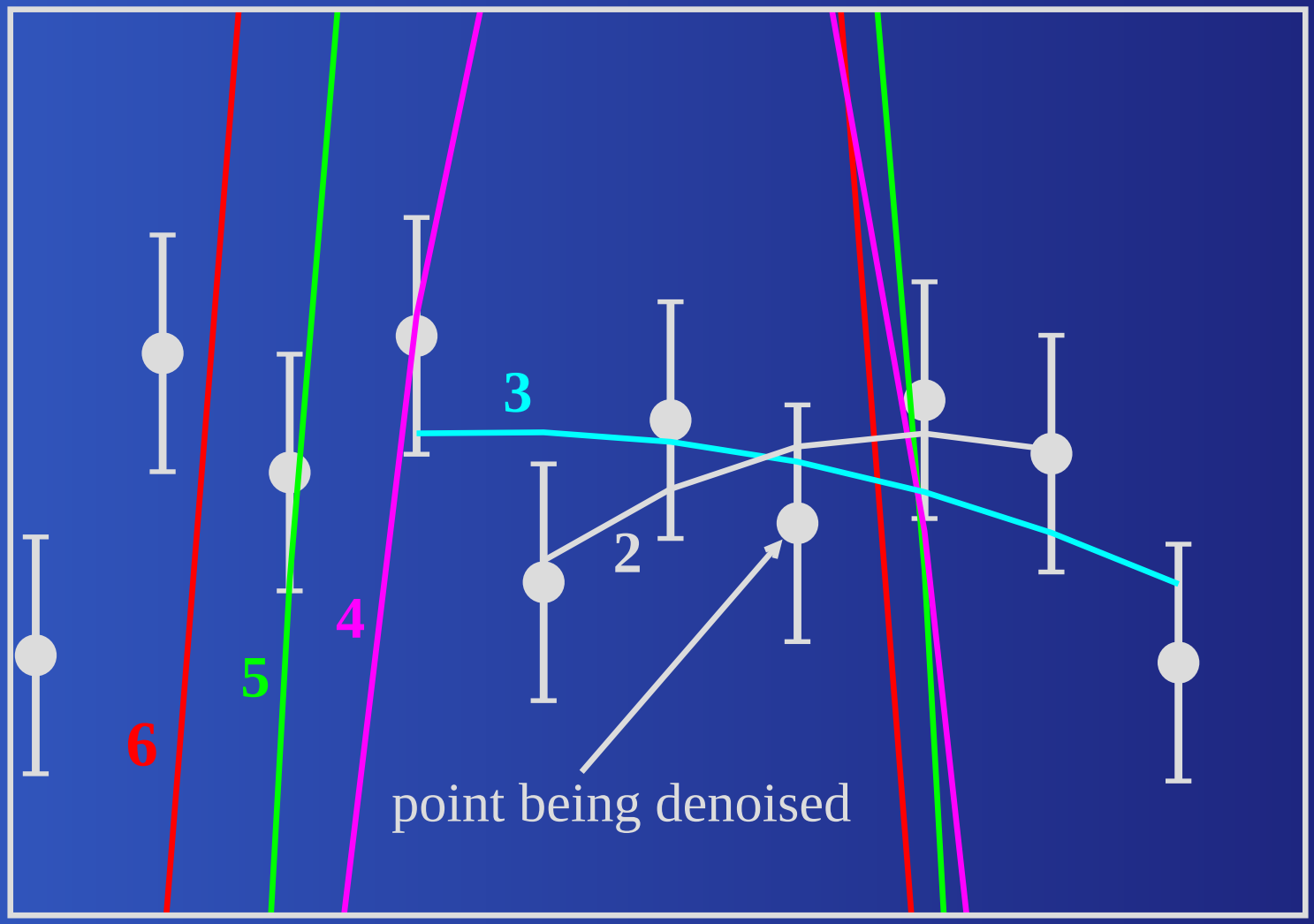
Global denoising: WTD

- Proposed by Deasy, Wickerhauser and Picard (2002)
- Method uses the computationally efficient 9-7 bi-orthogonal wavelet basis (also employed in the JPEG-2000 standard)
- Doses are transformed into wavelet space
- Wavelet coefficients below a certain threshold are set to zero (hence the term “threshold denoising”)
- The wavelets are transformed back to doses achieving a significant reduction in noise without smoothing out real dose features

Global denoising: LASG

- Developed by Kawrakow (2002)
- Uses a 3D form of a second order Savitzky-Goley filter
- The coefficients of a Savitzky-Goley filter result from the assumption that the doses within the smoothing window can be represented with a second order polynomial
- The maximum acceptable window size is determined for each point by subjecting the hypothesis to a χ^2-test
- Smoothing may be rejected altogether, or applied only in 1D or 2D
- To reduce the computational effort only symmetric windows are considered

Global denoising: LASG



Global denoising: AD

- Developed by the U Wisconsin group (Miao *at al* 2003)
- The noisy dose distribution $D(\vec{x},t)$ evolves with time according to a diffusion equation

$$\frac{\partial D(\vec{x},t)}{\partial t} = \text{div}\left[c(\vec{x},t)\nabla D(\vec{x},t)\right]$$

- The diffusivity $c(\vec{x},t)$ depends on the dose distribution via

$$c(\vec{x},t) = \left[1 + \left(\frac{|\nabla D(\vec{x},t)|}{K(\vec{x})}\right)^2\right]^{-1} \qquad K(\vec{x}) = 1.75\sigma(\vec{x})$$

  thus preventing diffusion across beam edges while allowing for "equilibration" in regions with relatively uniform dose
- The optimum number of time steps is determined from a series of numerical experiments

Global denoising: IRON

- Developed by Fippel and Nüsslin (2003) in terms of an optimization problem
- The cost function is

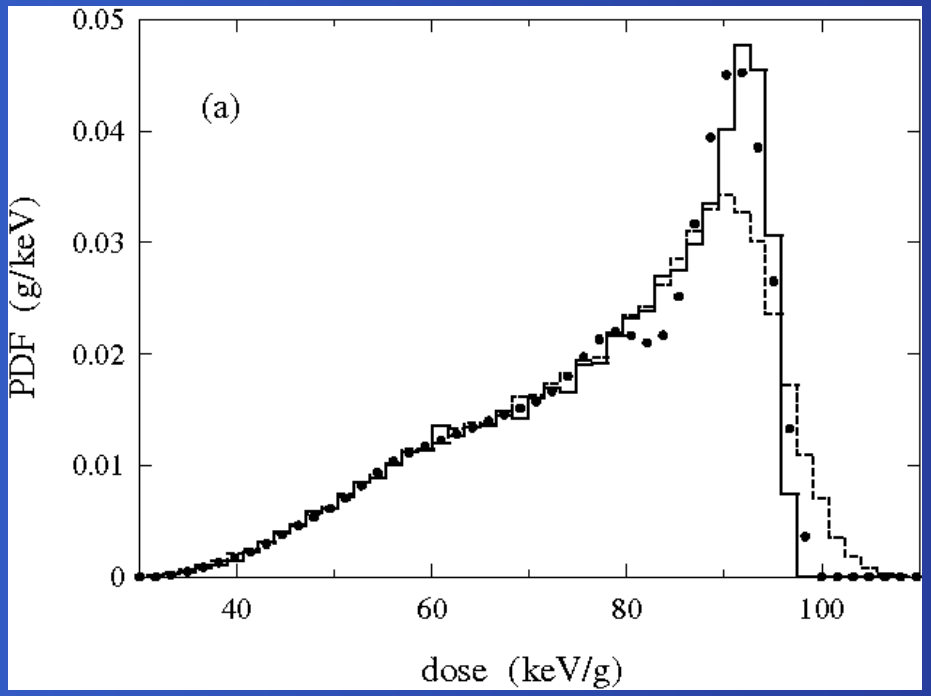
$$f = (1-\alpha)\sum_{ijk}\kappa_{ijk} + \alpha\sum_{ijk}\left(\frac{\tilde{D}_{ijk}-D_{ijk}}{\sigma_{ijk}}\right)^2$$

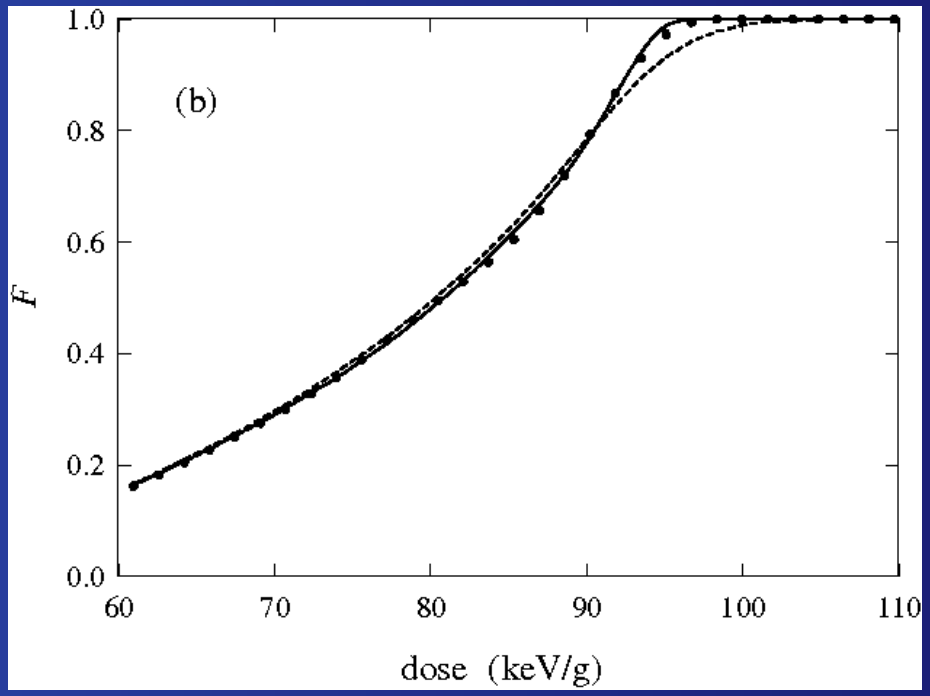
with $\alpha \sim 10^{-2}$ a free parameter and κ_{ijk} the curvature of the denoised dose distribution $\tilde{D}_{ijk}$ in voxel ijk:

$$\kappa_{ijk} = \frac{|2\tilde{D}_{ijk} - \tilde{D}_{i-1jk} - \tilde{D}_{i+1jk}|}{[1+1/4(D_{i-1jk}-D_{i+1jk})^2]^{3/2}} + j\text{-neighbors} + k\text{-neighbors}\ .$$

- A multi-dimensional conjugate-gradient method is used to optimize f, thus obtaining the denoised dose values $\tilde{D}_{ijk}$.

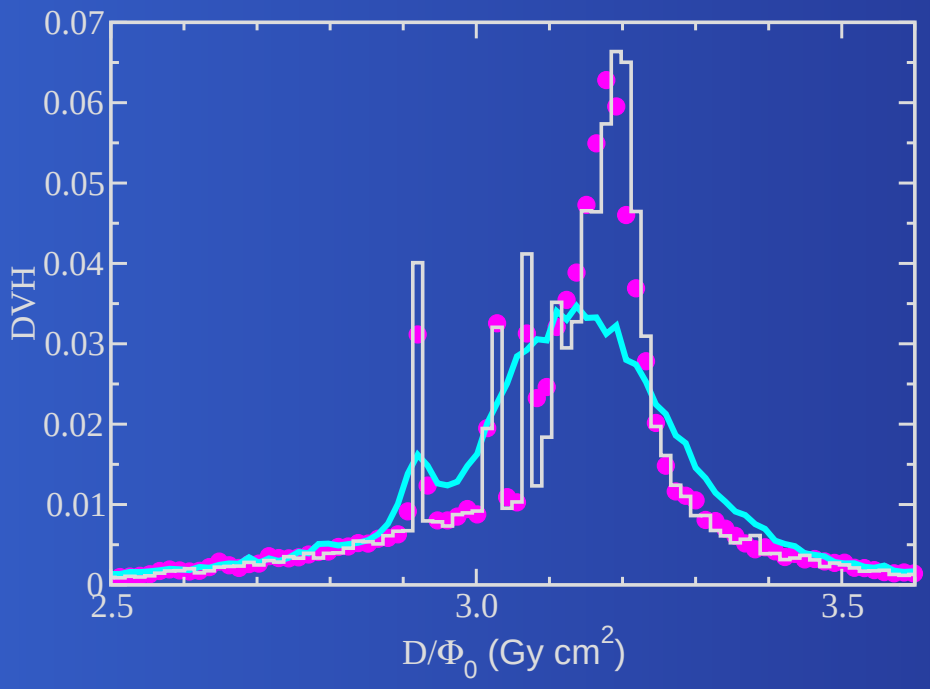
Denoising examples: DVH

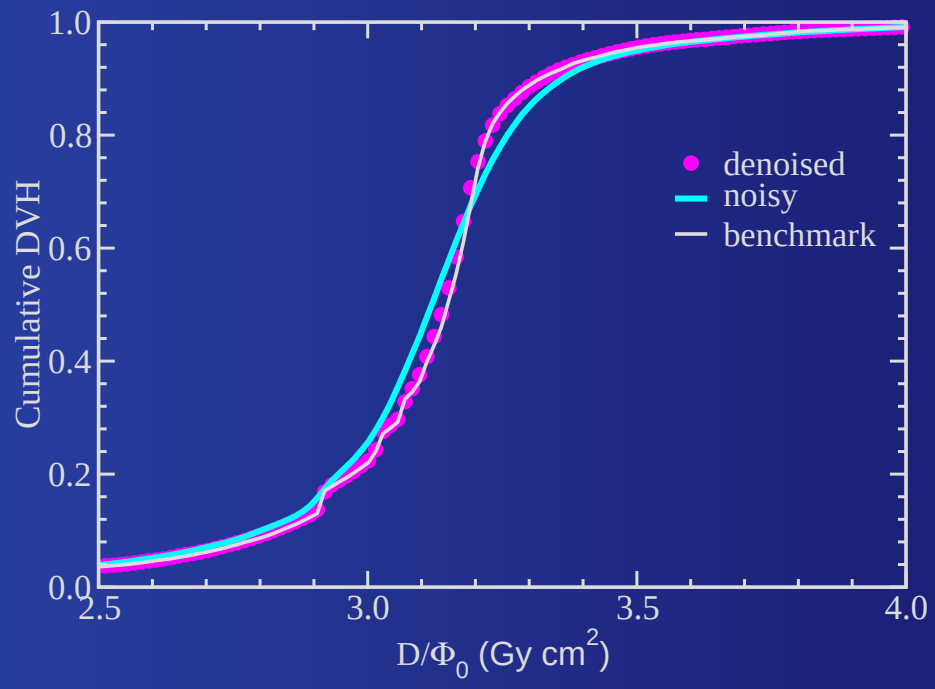




- DVH is for the high-dose region of a 10 MeV electron beam
- DVH denoising performed with the Sempau & Bielajew method
- Noisy distribution uses 100 times fewer particles

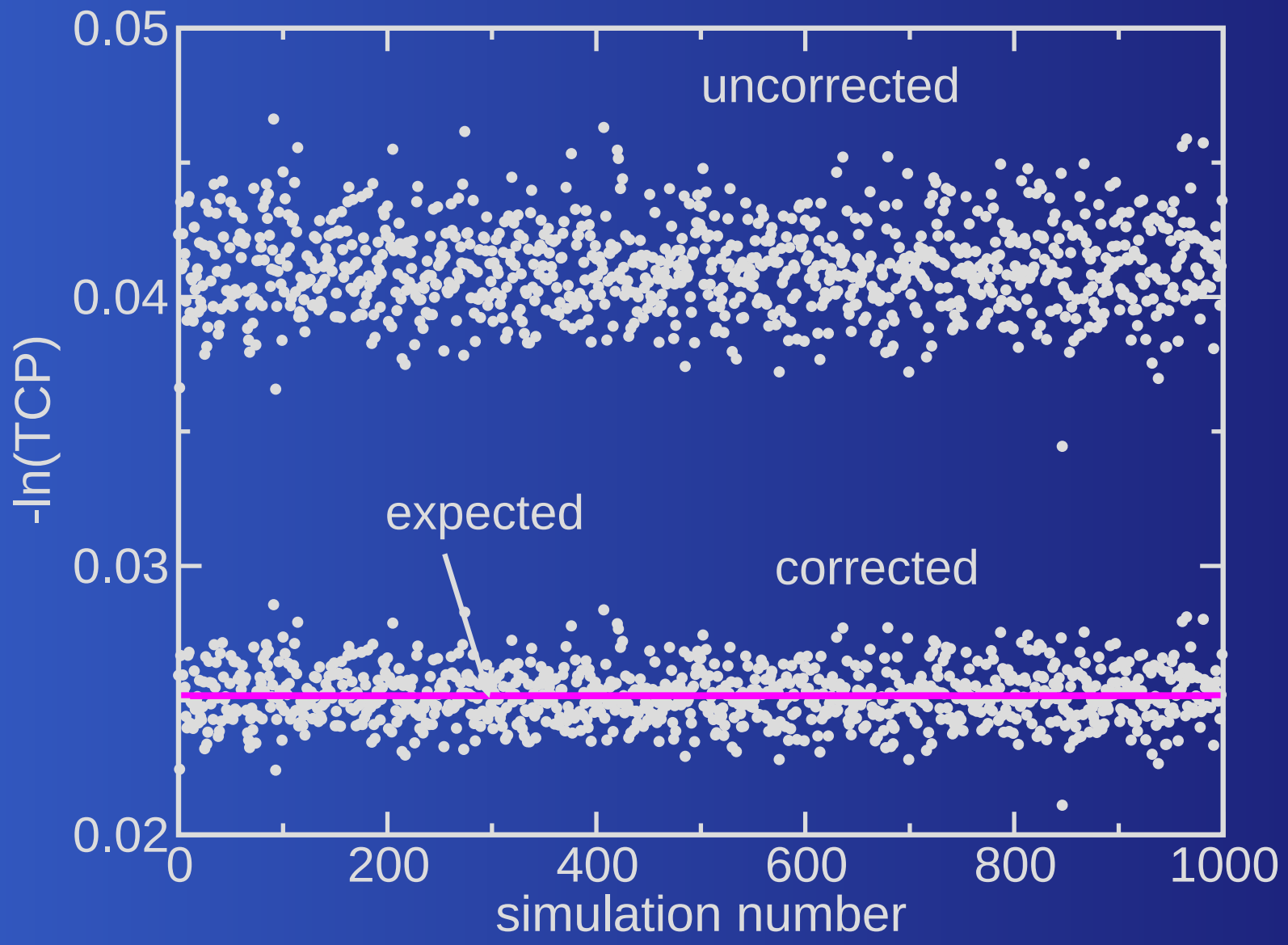
Denoising examples: DVH



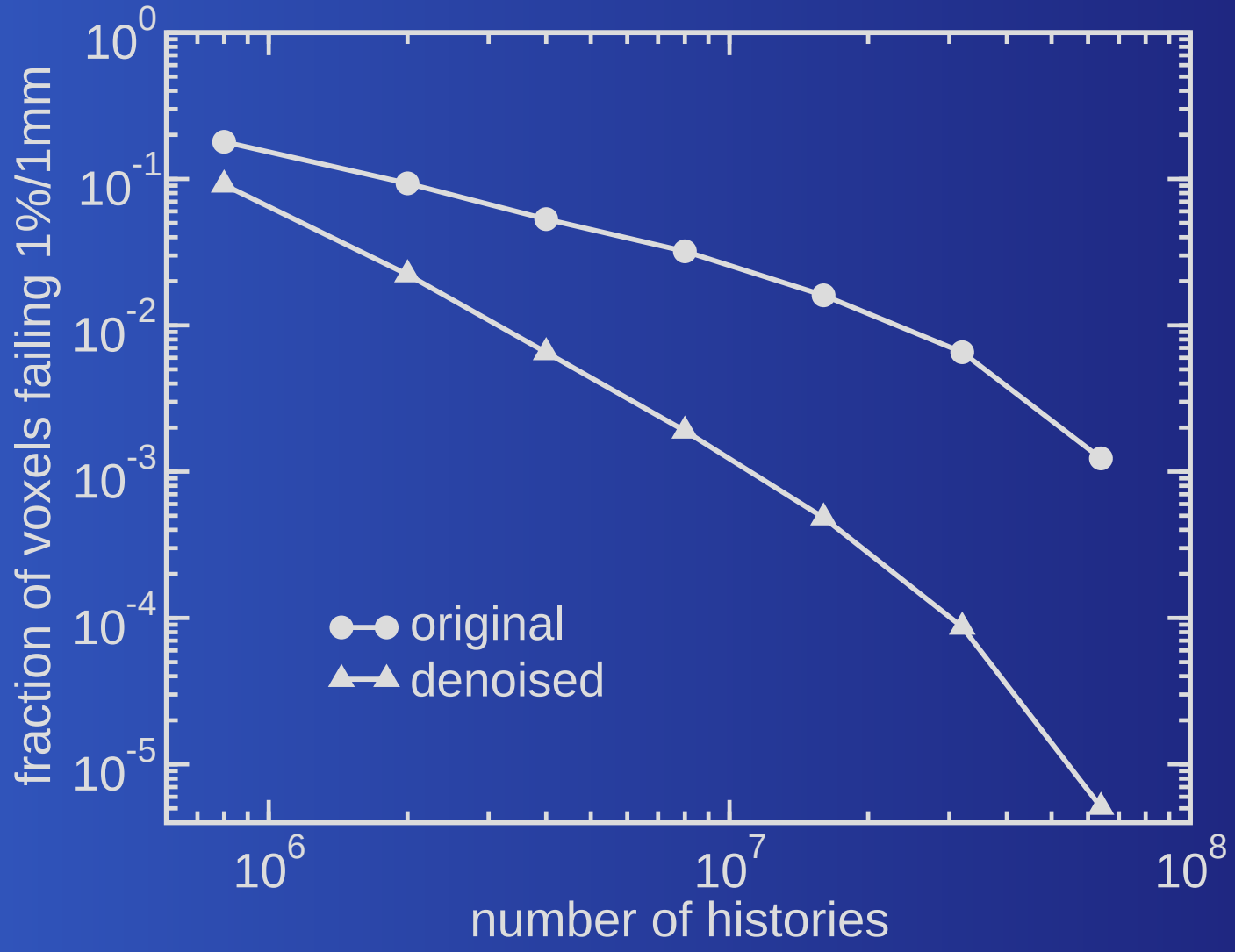


- DVH is for the high-dose region of a 20 MeV electron beam incident on a H_2O/Al/Air phantom
- DVH denoising performed with the LASG method
- Noisy distribution uses 600 times fewer particles

Denoising examples: TCP



Denoising examples: fraction failing 1%/1 mm



Denoising performed with the LASG algorithm

Denoising examples: fraction failing 2%/2 mm

Case	WTD	LASG	AD	IRON
HEL(2%)	4.92	44.29	5.64	8.38
HEL(3%)	4.18	18.40	4.95	5.97
HEL(5%)	3.00	7.16	3.71	3.94
IMRT(3%)	1.63	1.58	1.41	1.37
IMRT(7%)	1.46	1.49	1.43	1.36
LungA(1%)	5.67	5.10	3.00	7.29
LungA(2%)	3.52	4.63	4.25	4.49
LungA(6%)	1.43	2.30	1.46	2.05
LungB(3%)	2.35	2.84	2.00	2.16
LungB(7%)	1.73	1.81	1.65	1.62
Mean	2.647	4.576	2.567	3.119

Table 1: Factor by which the fraction of voxels failing a 2%/2 mm test is reduced

Denoising examples: MSD

Case	WTD	LASG	AD	IRON
HEL(2%)	4.17	12.42	5.04	6.20
HEL(3%)	5.16	14.01	6.49	7.54
HEL(5%)	6.31	16.00	9.25	9.00
IMRT(3%)	2.79	2.61	2.19	2.19
IMRT(7%)	4.49	4.33	4.16	3.63
LungA(1%)	2.53	2.78	2.27	2.70
LungA(2%)	3.61	3.96	3.60	3.75
LungA(6%)	4.34	7.51	4.77	6.82
LungB(3%)	3.51	3.81	2.78	3.03
LungB(7%)	4.97	6.13	4.72	4.97
Mean	4.05	6.00	4.12	4.51

Table 2: Factor by which the MSD is reduced

Conclusions

- The presence of statistical uncertainties in MC-computed dose distribution is a fundamental difference with traditional, deterministic dose calculation algorithms.
- The statistical uncertainties influence all aspects of the treatment planning and evaluation process.
- It is generally believed that statistical uncertainties 2% or better around the maximum dose are adequate for clinical use.
- Post-processing in form of DVH deconvolution, cost function uncertainty reduction, or *careful* global denoising, improves the accuracy of MC dose distributions
- Careless application of smoothing may decrease the accuracy (see *e.g.* [Ding *et al*, Phys.Med.Biol. 51 (2006) 2781-2799])
- Fast MC algorithms coupled with dose distribution post-processing can be very useful for the treatment plan optimization process.