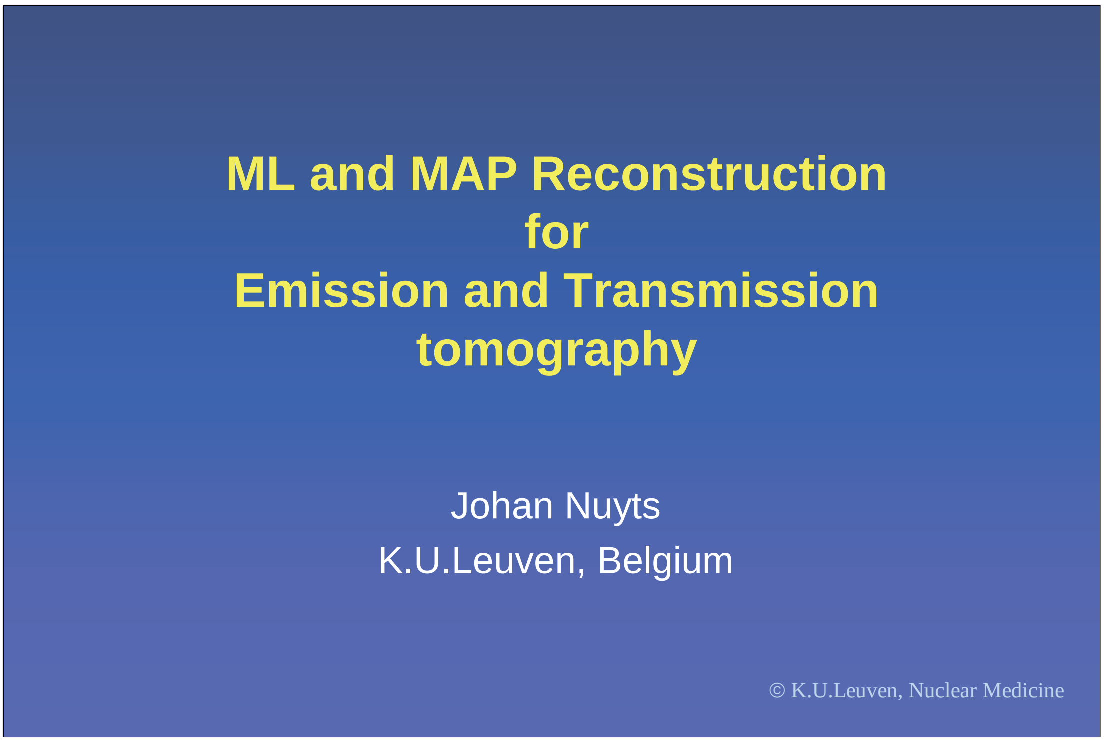
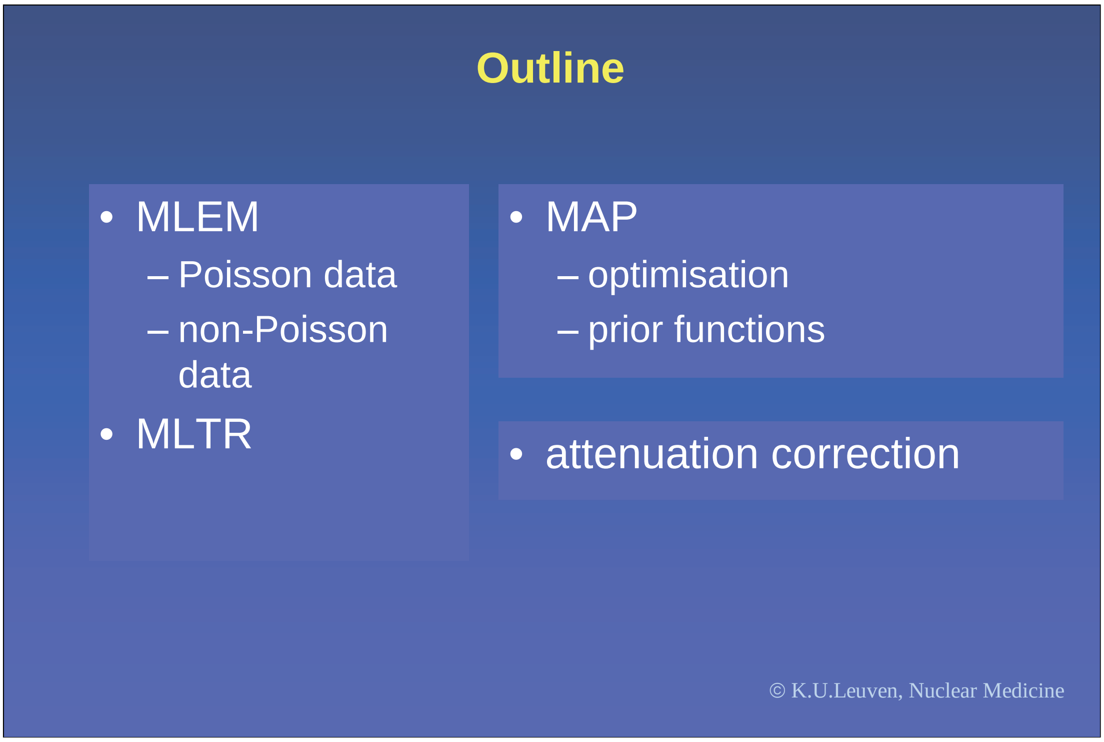


ML and MAP Reconstruction for Emission and Transmission tomography

Johan Nuyts
K.U.Leuven, Belgium



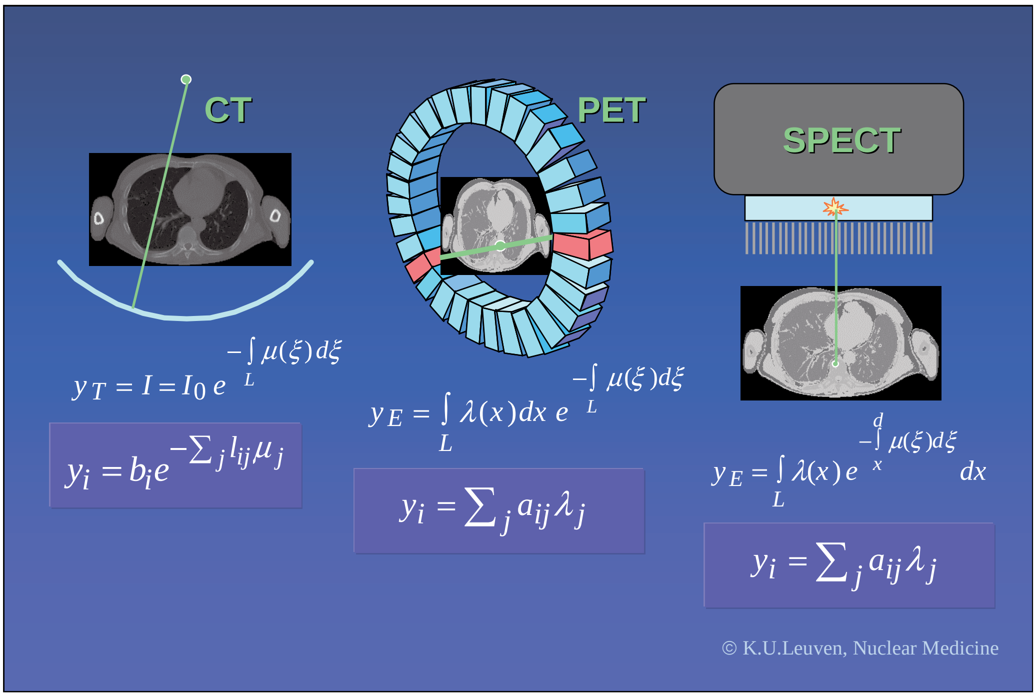
Outline

- MLEM
  - Poisson data
  - non-Poisson data
- MLTR

- MAP
  - optimisation
  - prior functions

- attenuation correction

MLEM stands for “Maximum Likelihood Expectation Maximisation”. In practice, it is often used for emission tomography, but rarily for transmission tomography. MLEM for transmission produces cumbersome equations, which are usually avoided by using other optimisation strategies. Many authors name their ML-algorithms for transmission tomography “MLTR”, as if the EM would stand for “emission”.



$$y_T = I = I_0\, e^{-\int_L \mu(\xi)\,d\xi}$$

$$y_i = b_i e^{-\sum_j l_{ij}\mu_j}$$

$$y_E = \int_L \lambda(x)\,dx\; e^{-\int_L \mu(\xi)\,d\xi}$$

$$y_i = \sum_j a_{ij}\lambda_j$$

$$y_E = \int_L \lambda(x)\, e^{-\int_x^d \mu(\xi)\,d\xi}\, dx$$

$$y_i = \sum_j a_{ij}\lambda_j$$

© K.U.Leuven, Nuclear Medicine

ML-algorithms assume that the measurement is discrete (which is correct) and that the distribution to be reconstructed is discrete (which is not correct). It is assumed that if the pixel size is sufficiently small, the digital image is a “sufficiently” accurate representation of the true continuous tracer distribution.

y_i : the measured count in detector bin i (transmission or emission)

b_i : the blank count in detector bin i

l_{ij} : (effective) intersection length of the projection line i with pixel j

μ_j : linear attenuation coefficient (dimension: 1/length)

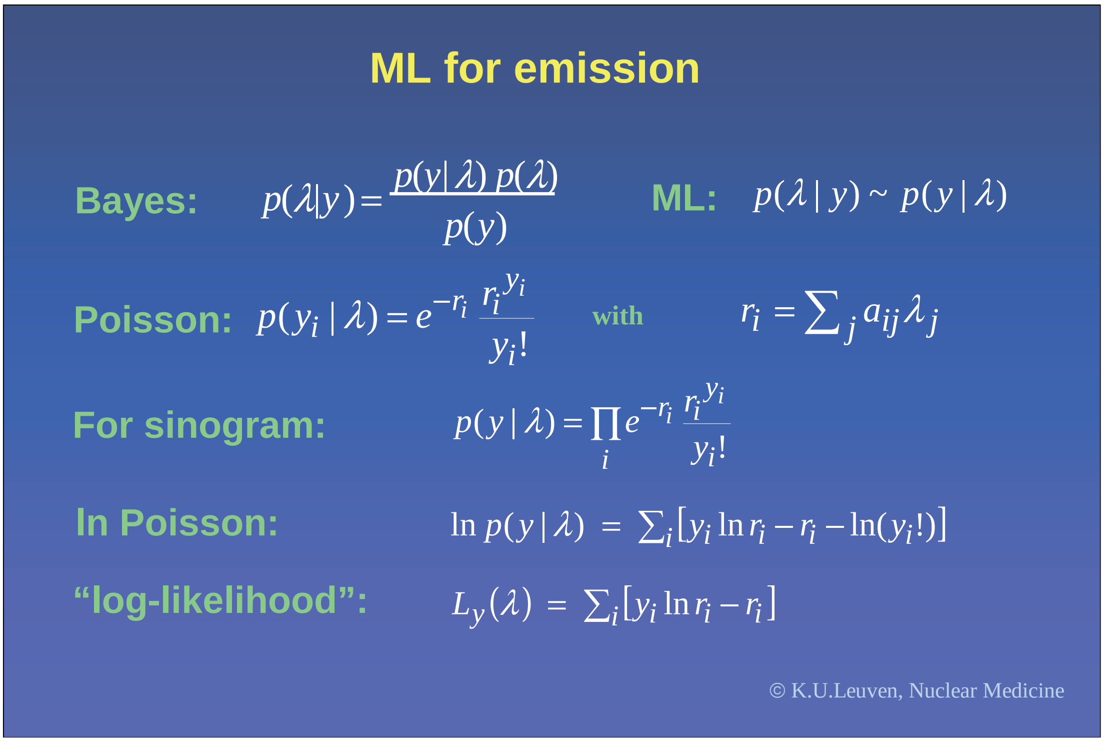
λ_j : activity in pixel j

a_{ij} : probability that photons emitted in j are detected in i

L: the projection line

I_0, I: the blank and transmission count

MLEM



ML for emission

Bayes: $p(\lambda|y) = \frac{p(y|\lambda)\,p(\lambda)}{p(y)}$ **ML:** $p(\lambda\,|\,y) \sim p(y\,|\,\lambda)$

Poisson: $p(y_i\,|\,\lambda) = e^{-r_i}\frac{r_i^{\,y_i}}{y_i!}$ **with** $r_i = \sum_j a_{ij}\lambda_j$

For sinogram: $p(y\,|\,\lambda) = \prod_i e^{-r_i}\frac{r_i^{\,y_i}}{y_i!}$

In Poisson: $\ln p(y\,|\,\lambda) \;=\; \sum_i [y_i \ln r_i - r_i - \ln(y_i!)]$

“log-likelihood”: $L_y(\lambda) \;=\; \sum_i [y_i \ln r_i - r_i]$

© K.U.Leuven, Nuclear Medicine

We would like to optimize p(λ|y) but we don’t know how to compute it (inverse problem).

However, Bayes rule allows it to compute as a function of p(y|λ) (forward problem), which we can “easily” compute if λ is known.

p(λ|y) is the posterior, which we want to maximize,

p(y|λ) is the likelihood,

p(λ) is the prior,

p(y) is a constant, since y is given.

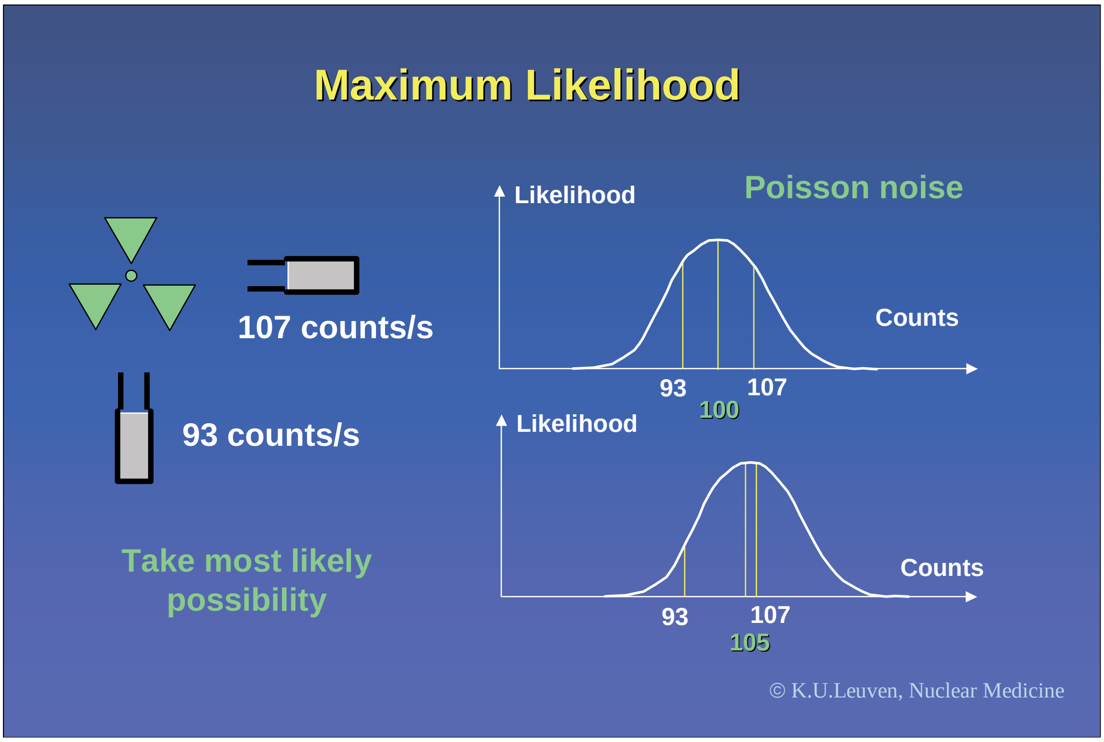
In ML (maximum likelihood), it is assumed that the prior is uniform (no prior knowledge). Maximizing the posterior is then equivalent to maximizing the likelihood.

In MAP (maximum a posteriori), the prior is not uniform, and the posterior is the product of the likelihood and the prior.

Since the logarithm is monotonically increasing, we may maximize the log of the posterior, which is easier.

MLEM: see [1]-[10].

The likelihood is not the only useful objective function. Examples of alternatives are [11]-[13].



Maximum Likelihood

107 counts/s

93 counts/s

Take most likely possibility

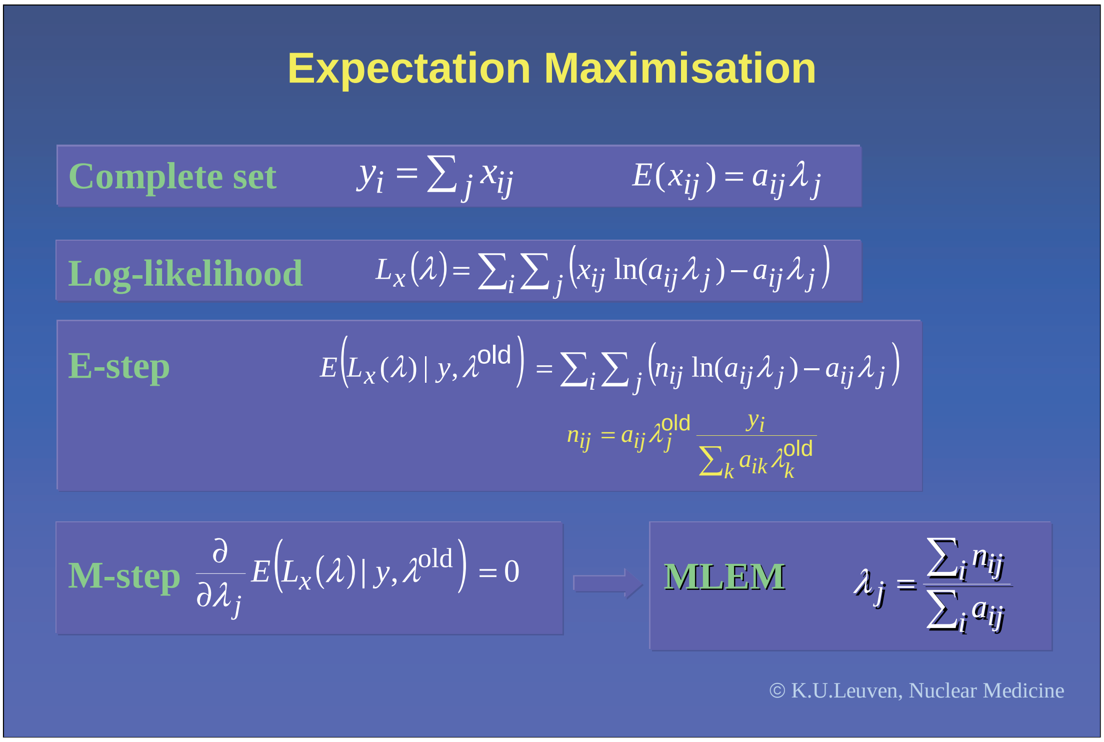
Poisson noise

© K.U.Leuven, Nuclear Medicine

The reconstruction problem does not have a unique solution.

In ML, the image that maximizes the likelihood is proposed as the solution. However, in this simple single pixel example, the probability that the ML solution (100) is correct, equals p(93|100) x p(107|100) = 0.00095. For an image with 10000 pixels, the corresponding probabilty is effectively zero, so we *know* that the ML-solution is wrong!

We hope that the ML-solution shares important features with the true solution, such that it is clinically useful. In practice, this is usually not the case and additional information is introduced to improve the image features. This information can be explicit (prior density $p(\lambda)$) or implicit (stopping iterations early, smoothing...). All these noise suppression techniques are often referred to as regularization.



Expectation Maximisation

Complete set $y_i = \sum_j x_{ij}$ $E(x_{ij}) = a_{ij}\lambda_j$

Log-likelihood $L_x(\lambda) = \sum_i \sum_j \left(x_{ij} \ln(a_{ij}\lambda_j) - a_{ij}\lambda_j \right)$

E-step $E\left(L_x(\lambda) \mid y, \lambda^{\text{old}}\right) = \sum_i \sum_j \left(n_{ij} \ln(a_{ij}\lambda_j) - a_{ij}\lambda_j \right)$

$$n_{ij} = a_{ij}\lambda_j^{\text{old}} \frac{y_i}{\sum_k a_{ik}\lambda_k^{\text{old}}}$$

M-step $\frac{\partial}{\partial \lambda_j} E\left(L_x(\lambda) \mid y, \lambda^{\text{old}}\right) = 0$

MLEM $\lambda_j = \frac{\sum_i n_{ij}}{\sum_i a_{ij}}$

© K.U.Leuven, Nuclear Medicine

For the EM-principle, see Dempster et al [2].

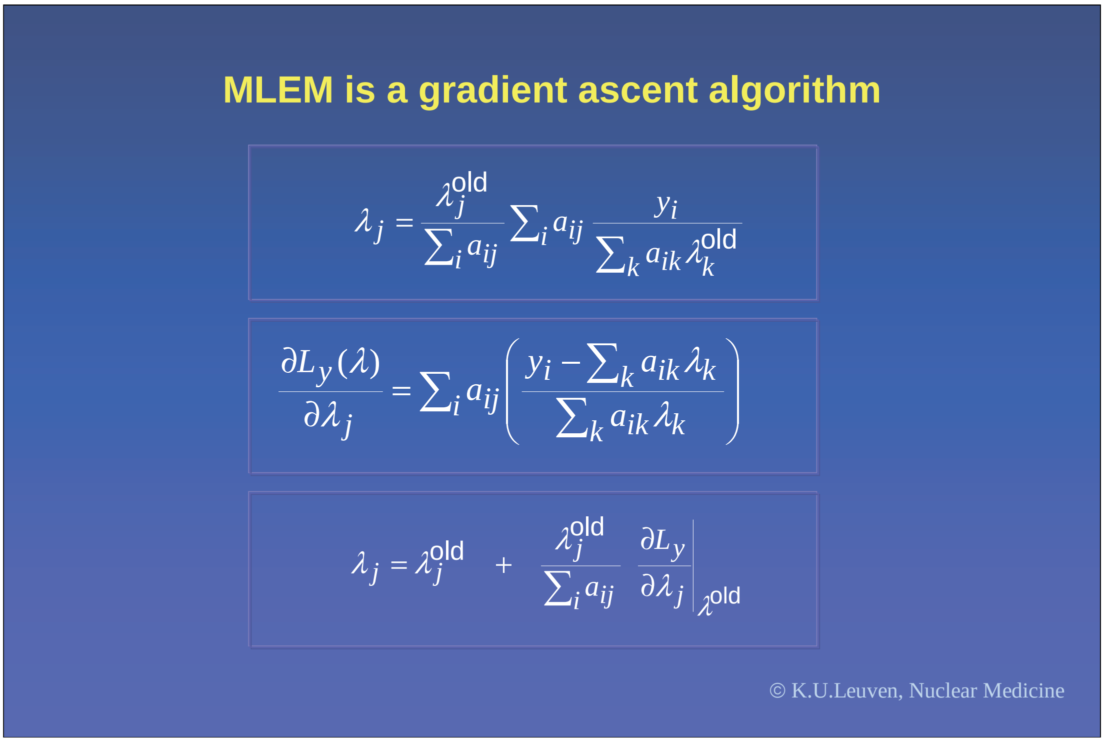
The EM-algorithm received much attenuation from the nuclear medicine community after the papers by Shepp and Vardi (emission tomography) [3] and Lange and Carson (emission and transmission tomography) [4].

The number of complete variables is huge (equals the number of reconstruction pixels times the number of sinogram pixels). So it is very nice to see that the complete variables disappear from the final MLEM equation!

A different choice of complete variables leads to a different MLEM algorithm [16].

A.R. De Pierro [10] found an alternative derivation of the MLEM algorithm, using surrogate functions. This derivation is useful for extension to maximum-a-posteriori reconstruction.

Many researchers have studied how MLEM can be accelerated [26]-[33]. A very successful method is OSEM (ordered subsets), proposed by Hudson and Larkin [28]. With OSEM, convergence in absence of noise is only guaranteed if the “subset balance” condition is met. RBI-methods (rescaled block iterative methods) eliminate the need for this condition, but convergence in presence of noise is still not guaranteed: the algorithm converges towards a limit cycle rather than towards a single solution [30],[31],[33]. Under-relaxation decreases the “radius” of the limit cycle by decreasing the step size per iteration (RAMLA [29]). One can also gradually decrease the number of subsets to obtain convergence similar to that of the original MLEM algorithm.



MLEM is a gradient ascent algorithm

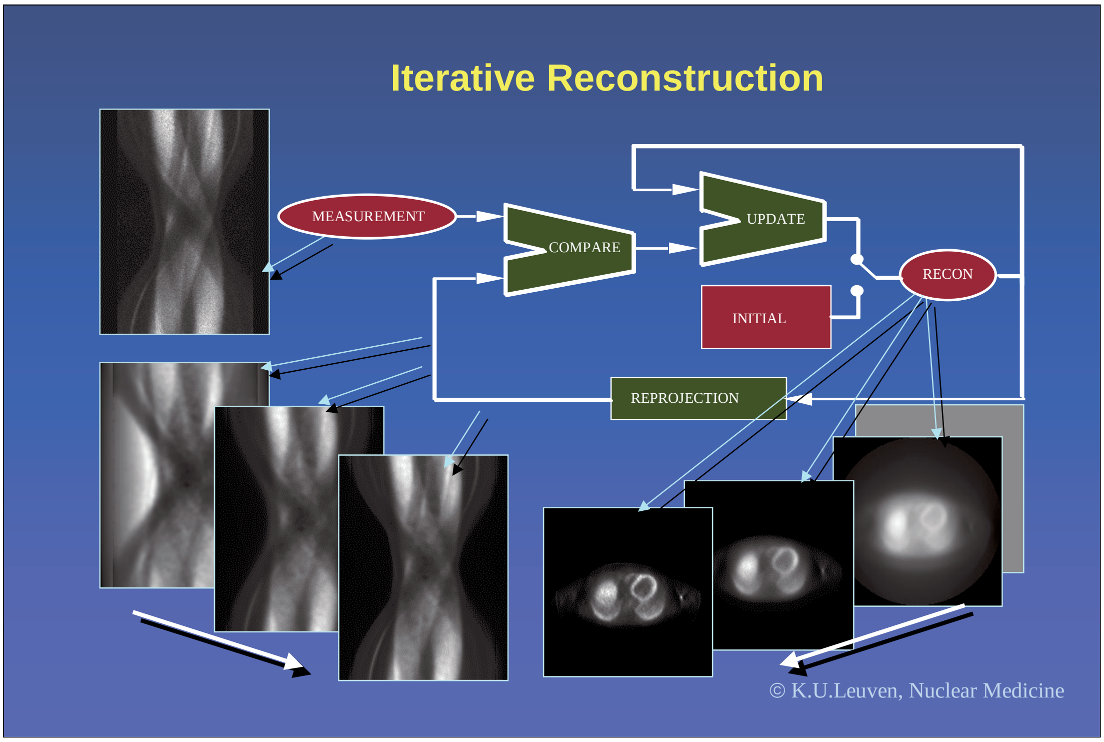
$$\lambda_j = \frac{\lambda_j^{\text{old}}}{\sum_i a_{ij}} \sum_i a_{ij} \frac{y_i}{\sum_k a_{ik} \lambda_k^{\text{old}}}$$

$$\frac{\partial L_y(\lambda)}{\partial \lambda_j} = \sum_i a_{ij} \left(\frac{y_i - \sum_k a_{ik} \lambda_k}{\sum_k a_{ik} \lambda_k} \right)$$

$$\lambda_j = \lambda_j^{\text{old}} + \frac{\lambda_j^{\text{old}}}{\sum_i a_{ij}} \left. \frac{\partial L_y}{\partial \lambda_j} \right|_{\lambda^{\text{old}}}$$

The MLEM algorithm for emission tomography turns out to be a gradient ascent algorithm (see e.g. [26]). The algorithm provides a step size that

- guarantees convergence
- ensures nonnegativity (step size vanishes before λ can become negative).



Iterative Reconstruction

MEASUREMENT

COMPARE

UPDATE

RECON

INITIAL

REPROJECTION

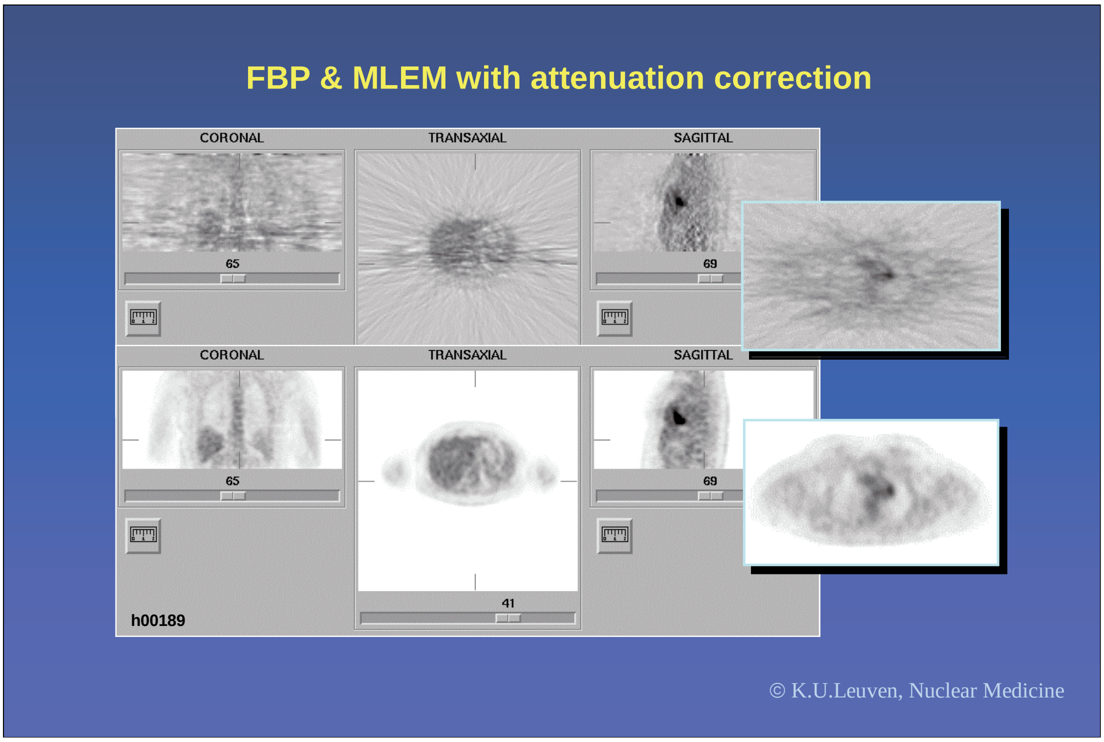
© K.U.Leuven, Nuclear Medicine

For intuitive reasoning, it may be handy to consider the MLEM algorithm (and indeed nearly all iterative algorithms) as a feedback system.

The feedback loop contains the “reprojection”. The reprojection is a simulation of the operation that must be inverted, i.e. the acquistion process. Effects not simulated here, will be ignored during reconstruction.

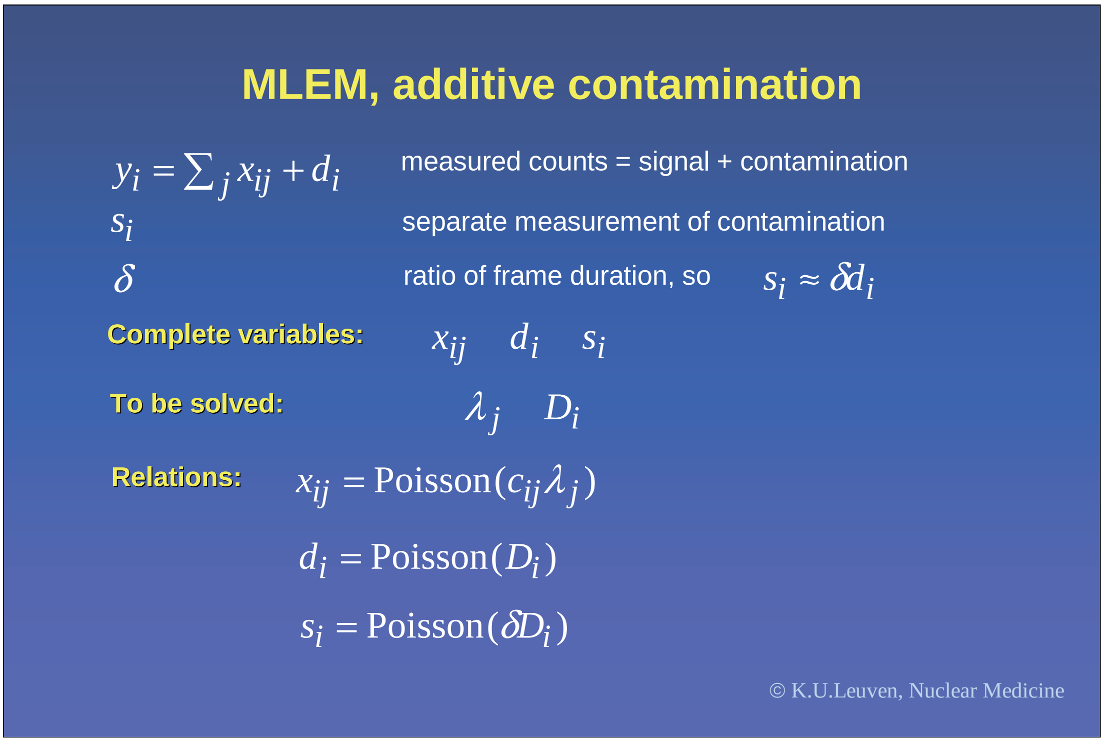
The forward loop contains an operation that updates the reconstruction based on deviations between the true and the simulated data, aiming at decreasing the deviations.

It is well known that the performance of a feedback system is very sensitive to its feedback loop, and less sensitive to its forward loop. So it is not suprising that the differences between various iterative algorithms are in de forward loop, rarely in the feedback loop. E.g., some authors propose to use a simplified backprojector (*back*projector is in the *forward* loop) to accelerate convergence [14],[15].



FBP & MLEM with attenuation correction

A PET study (2D PET, ^{18}FDG) reconstructed with filtered backprojection and MLEM.



MLEM, additive contamination

$y_i = \sum_j x_{ij} + d_i$	measured counts = signal + contamination
s_i	separate measurement of contamination
δ	ratio of frame duration, so $s_i \approx \delta d_i$

Complete variables: $x_{ij} \quad d_i \quad s_i$

To be solved: $\lambda_j \quad D_i$

Relations:

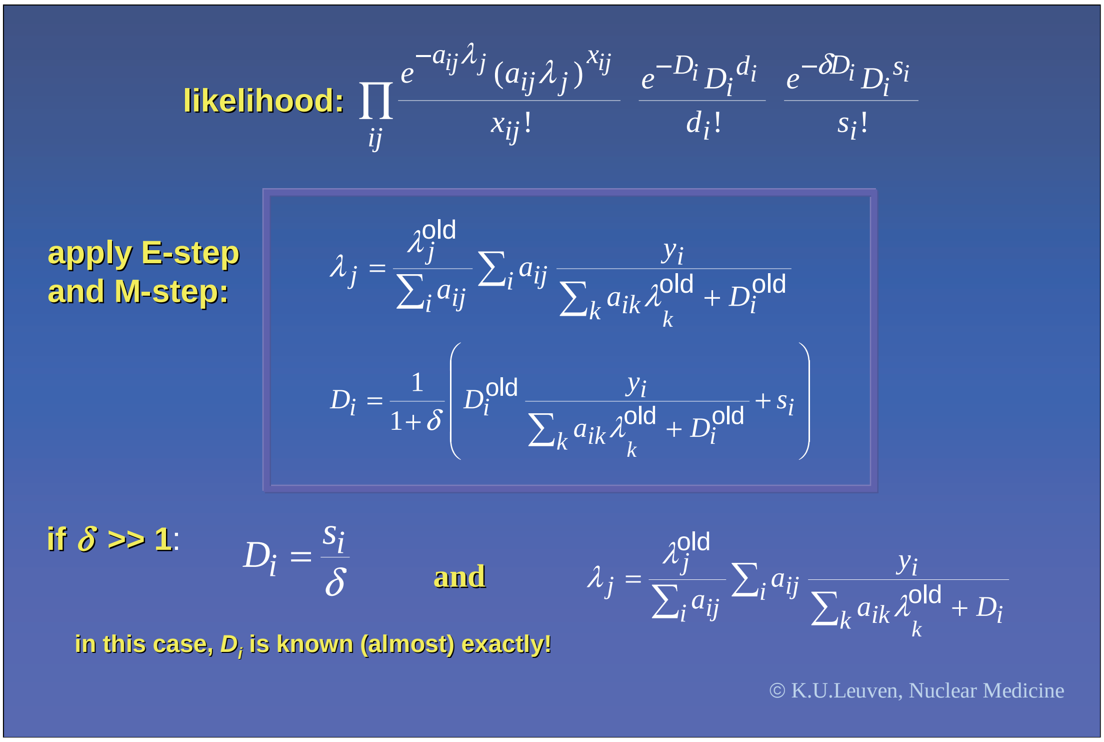
$$x_{ij} = \text{Poisson}(c_{ij}\lambda_j)$$

$$d_i = \text{Poisson}(D_i)$$

$$s_i = \text{Poisson}(\delta D_i)$$

δ represents possible differences in total effective counts (and therefore signal to noise ratio) between the actual contamination and the estimate of the contamination. If δ > 1 the estimate is less noisy than the true contamination. Reducing the noise of the estimate (e.g. smoothing of a scatter estimate, using singles to compute the randoms contribution in PET) increases δ.

MLEM can be applied to this problem by extending the set of complete variables with d_i (the true contamination) and s_i (the estimate). The unknown mean of the contamination D_i is treated as part of the solution (although we usually don’t care to know it).



likelihood: $$\prod_{ij} \frac{e^{-a_{ij}\lambda_j}(a_{ij}\lambda_j)^{x_{ij}}}{x_{ij}!}\;\frac{e^{-D_i} D_i^{d_i}}{d_i!}\;\frac{e^{-\delta D_i} D_i^{s_i}}{s_i!}$$

apply E-step and M-step:

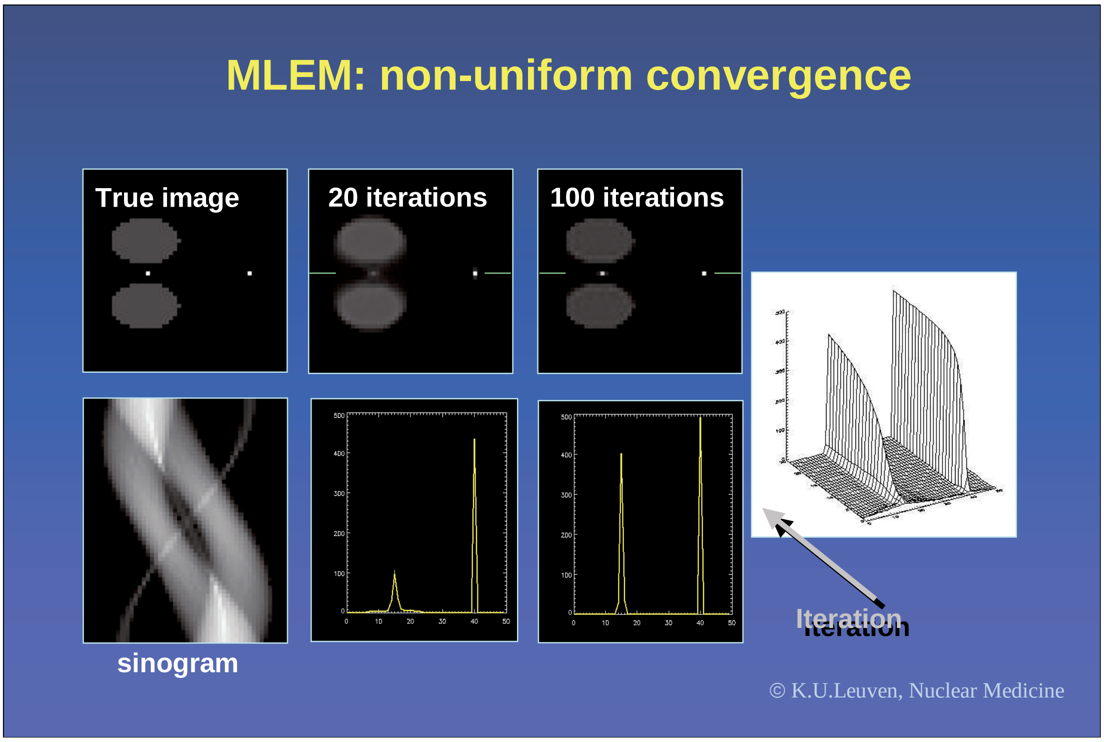
$$\boxed{\begin{aligned} \lambda_j &= \frac{\lambda_j^{\text{old}}}{\sum_i a_{ij}} \sum_i a_{ij} \frac{y_i}{\sum_k a_{ik}\lambda_k^{\text{old}} + D_i^{\text{old}}} \\ D_i &= \frac{1}{1+\delta}\left(D_i^{\text{old}} \frac{y_i}{\sum_k a_{ik}\lambda_k^{\text{old}} + D_i^{\text{old}}} + s_i \right) \end{aligned}}$$

if δ >> 1: $$D_i = \frac{s_i}{\delta}$$ **and** $$\lambda_j = \frac{\lambda_j^{\text{old}}}{\sum_i a_{ij}} \sum_i a_{ij} \frac{y_i}{\sum_k a_{ik}\lambda_k^{\text{old}} + D_i}$$

in this case, *D_i* is known (almost) exactly!

MLEM is now used to reconstruct both λ_j and D_i. Reconstruction of D_i is not very interesting and is often avoided by introducing approximations [35].

If s_i is noise-free then D_i is exactly known and a simple MLEM algorithm for λ_j is obtained: the contamination must simply be added in the denominator [16],[17].



MLEM: non-uniform convergence

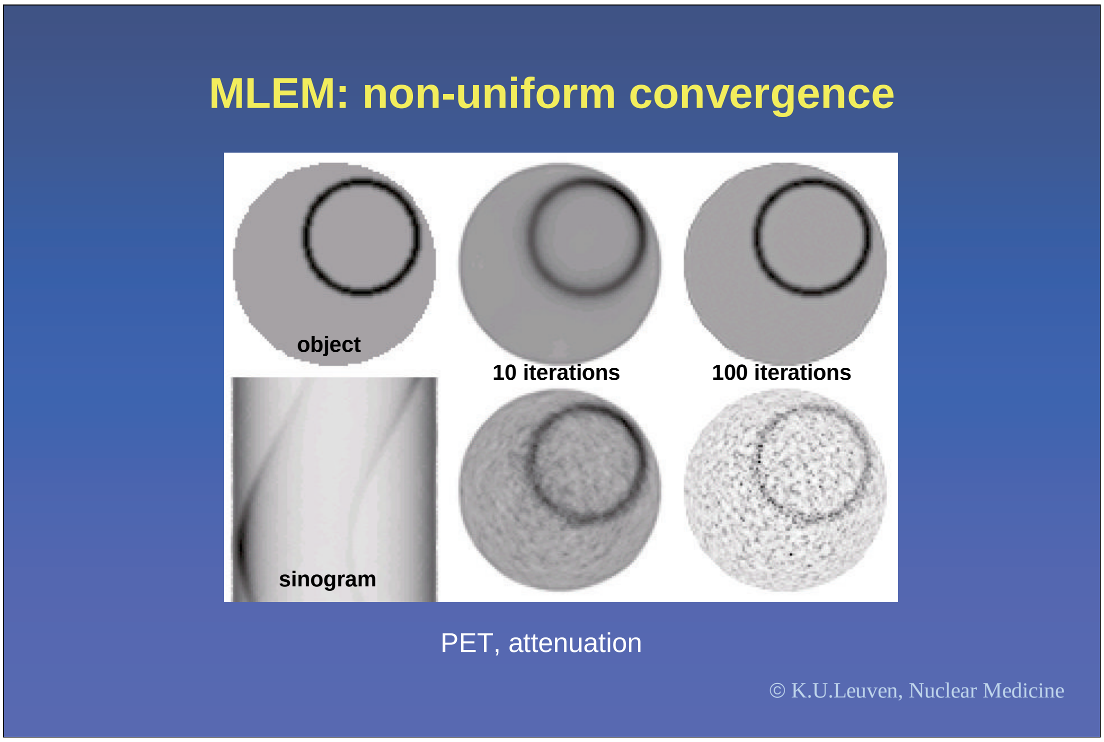
© K.U.Leuven, Nuclear Medicine

The simulation shows that the surrounding activity slows down convergence of the left point source (the true point value is 500). There was no attenuation or resolution blurring.

Moreover, the convergence is anisotropic: for the left point source, convergence of the x-resolution is worse than that of the y-resolution at low iterations.

So convergence is dependent on position, orientation and on the entire activity distribution within the field of view!

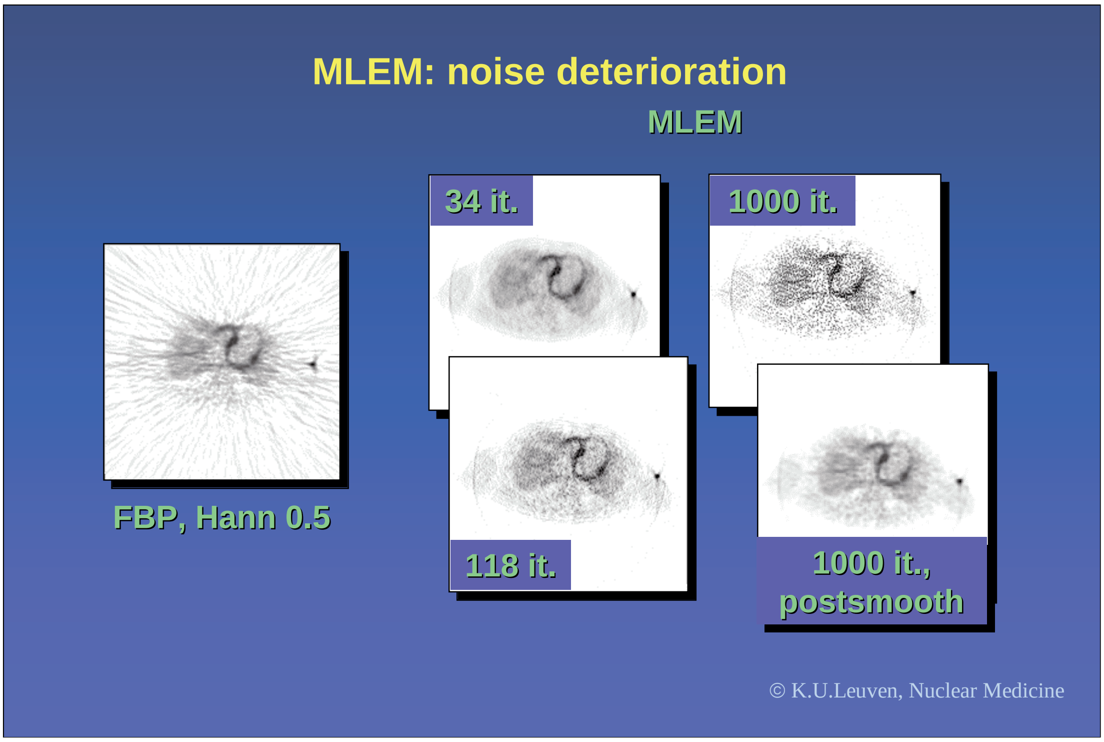
A simular problem shows up in penalized-likelihood reconstruction.



MLEM: non-uniform convergence

PET, attenuation

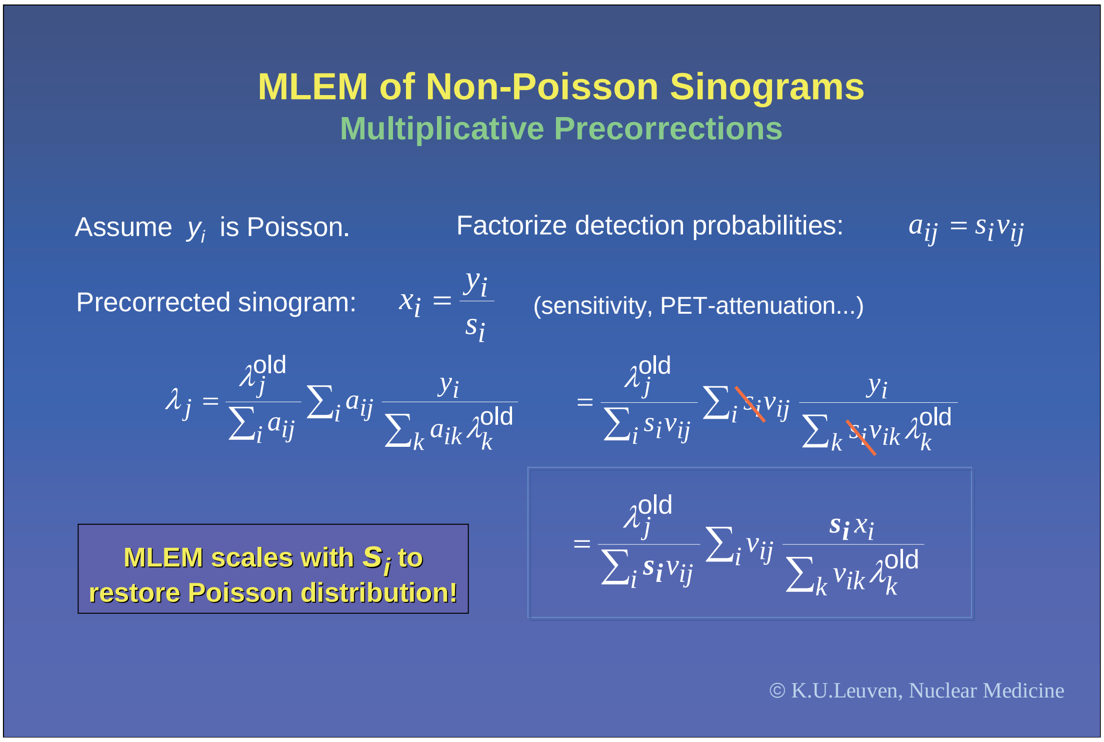
Non-uniform convergence in a simulated PET study, taking into account attenuation, and assuming ideal resolution. The central part is surrounded by more activity and suffers more from attenuation, and needs more iterations for accurate reconstruction.



MLEM: noise deterioration

© K.U.Leuven, Nuclear Medicine

The likelihood is maximum if the reprojection of the image equals the measurement. So the likelihood increases if the sinogram noise is incorporated in the reconstruction (overfitting). Postsmoothing (2D Gaussian) reveals that the noisy image at 1000 iterations contains valuable information. Convergence of MLEM has been studied by many researchers [18]-[25].



MLEM of Non-Poisson Sinograms

Multiplicative Precorrections

Assume y_i is Poisson.

Factorize detection probabilities: $a_{ij} = s_i v_{ij}$

Precorrected sinogram: $x_i = \frac{y_i}{s_i}$ (sensitivity, PET-attenuation...)

$$\lambda_j = \frac{\lambda_j^{\text{old}}}{\sum_i a_{ij}} \sum_i a_{ij} \frac{y_i}{\sum_k a_{ik} \lambda_k^{\text{old}}} = \frac{\lambda_j^{\text{old}}}{\sum_i s_i v_{ij}} \sum_i \cancel{s_i} v_{ij} \frac{y_i}{\sum_k \cancel{s_i} v_{ik} \lambda_k^{\text{old}}}$$

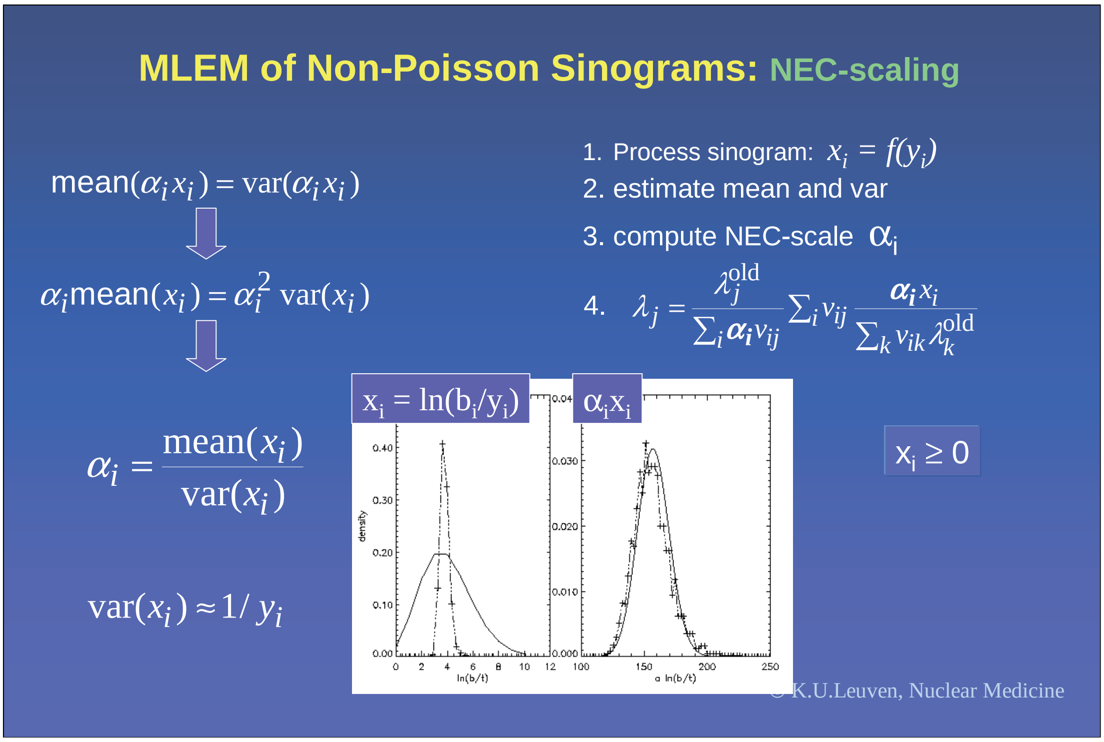
$$\boxed{= \frac{\lambda_j^{\text{old}}}{\sum_i \mathbf{s_i} v_{ij}} \sum_i v_{ij} \frac{\mathbf{s_i} x_i}{\sum_k v_{ik} \lambda_k^{\text{old}}}}$$

MLEM scales with S_i to restore Poisson distribution!

© K.U.Leuven, Nuclear Medicine

After precorrection, the sinogram values are no longer Poisson distributed.

Here, we rewrite the MLEM algorithm as a reconstruction algorithm for a precorrected sinogram (multiplicative precorrection). This shows that MLEM scales the data to "restore" the Poisson characteristics.



MLEM of Non-Poisson Sinograms: NEC-scaling

$$\text{mean}(\alpha_i x_i) = \text{var}(\alpha_i x_i)$$

$$\Downarrow$$

$$\alpha_i \text{mean}(x_i) = \alpha_i^2 \, \text{var}(x_i)$$

$$\Downarrow$$

$$\alpha_i = \frac{\text{mean}(x_i)}{\text{var}(x_i)}$$

$$\text{var}(x_i) \approx 1/y_i$$

1. Process sinogram: $x_i = f(y_i)$
2. estimate mean and var
3. compute NEC-scale α_i
4. $\lambda_j = \frac{\lambda_j^{\text{old}}}{\sum_i \boldsymbol{\alpha_i} v_{ij}} \sum_i v_{ij} \frac{\boldsymbol{\alpha_i} x_i}{\sum_k v_{ik} \lambda_k^{\text{old}}}$

$x_i = \ln(b_i/y_i)$

$\alpha_i x_i$

$x_i \geq 0$

© K.U.Leuven, Nuclear Medicine

Any random variable can be rescaled such that mean and variance correspond to that of a Poisson distribution. The scaled distribution coincides with the Poisson distribution up to the second moment [82]. By definition, this scaled value is the noise equivalent count (NEC) [34].

Here the NEC-scaling is illustrated on the log-converted transmission scan. The raw transmission scan is Poisson distributed, but the log-converted distribution is very different. After scaling with $1/y_i$ the distribution is very similar to Poisson.

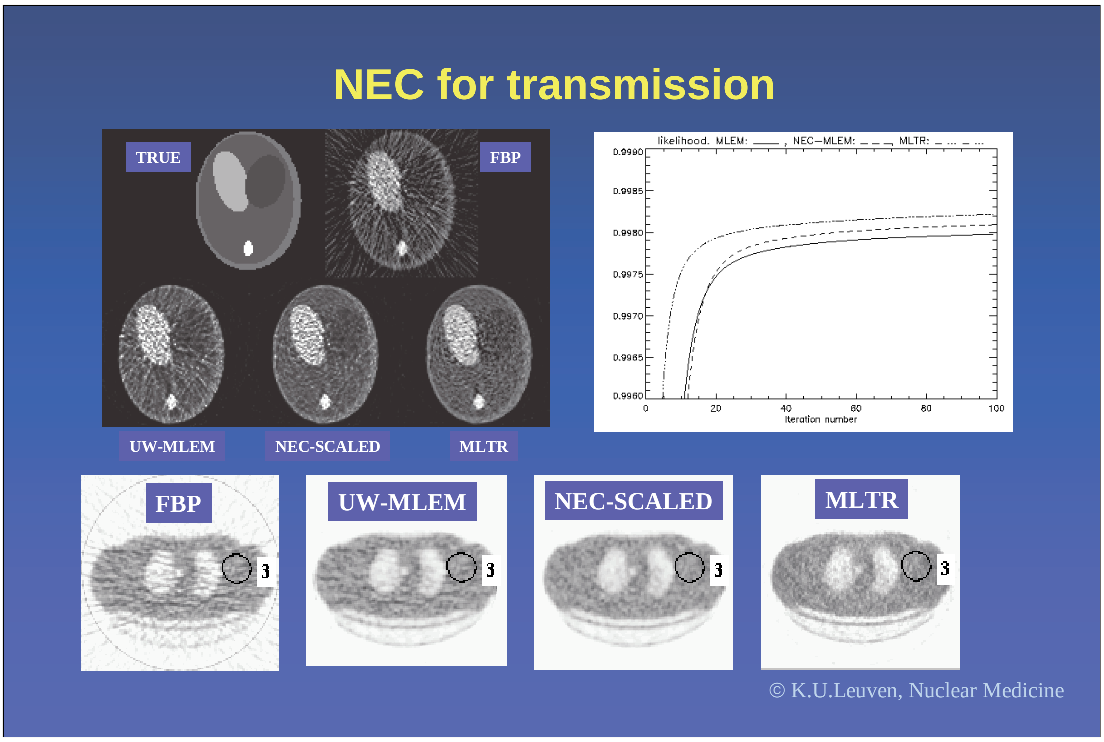
Log conversion of transmission scans: $x_i = \ln \frac{b_i}{y_i}$ (y_i = trans, b_i = blank)

$$\text{var}(x_i) \approx \left(\frac{\partial x_i}{\partial y_i}\right)^2 \text{var}(y_i) = \frac{1}{\text{mean}(y_i)}$$

We estimate the scale factor as:

$$\vec{\alpha} = \vec{g} \otimes \left(\vec{x} \cdot \vec{y}\right)$$

(dot product denotes pixel by pixel multiplication
and g represents a smoothing filter)

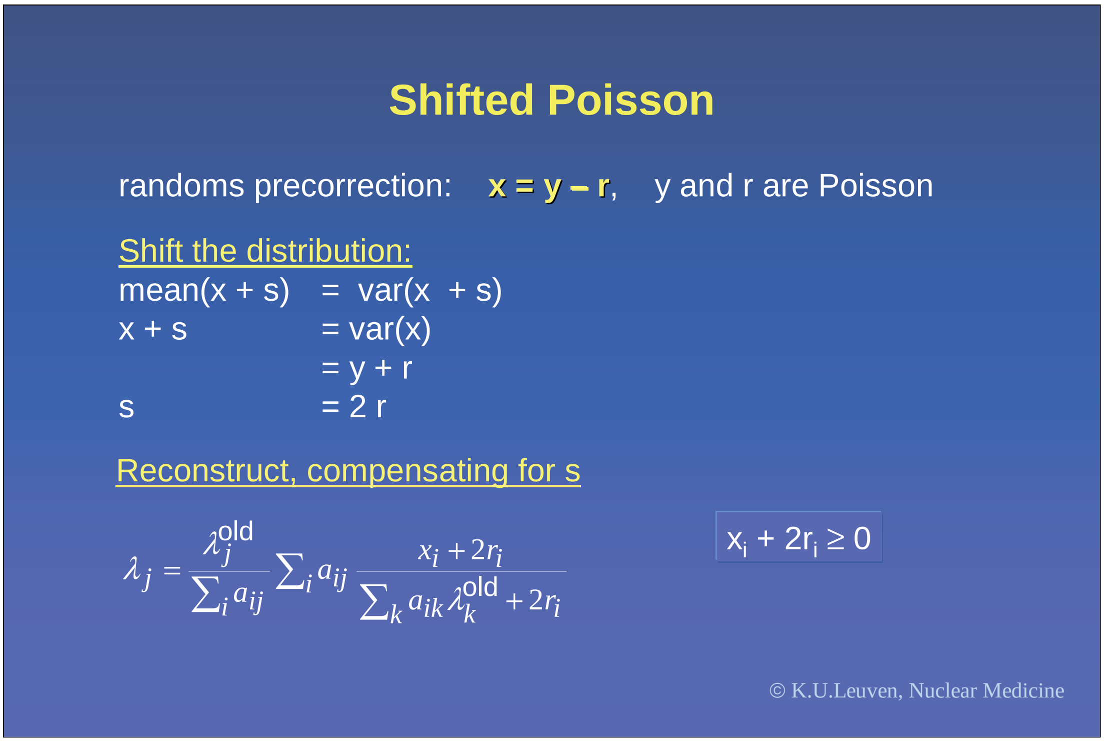


NEC for transmission

MLTR is a dedicated ML algorithm for transmission tomography, discribed later.

NEC-scaled MLEM outperforms unweighted MLEM reconstruction of the log-converted data ln(b/y): the likelihood is higher and the streak artifacts are eliminated.

MLTR achieves still higher likelihood than NEC-scaled MLEM.



Shifted Poisson

randoms precorrection: **$x = y - r$**, y and r are Poisson

Shift the distribution:

mean(x + s) = var(x + s)

x + s = var(x)

= y + r

s = 2 r

Reconstruct, compensating for s

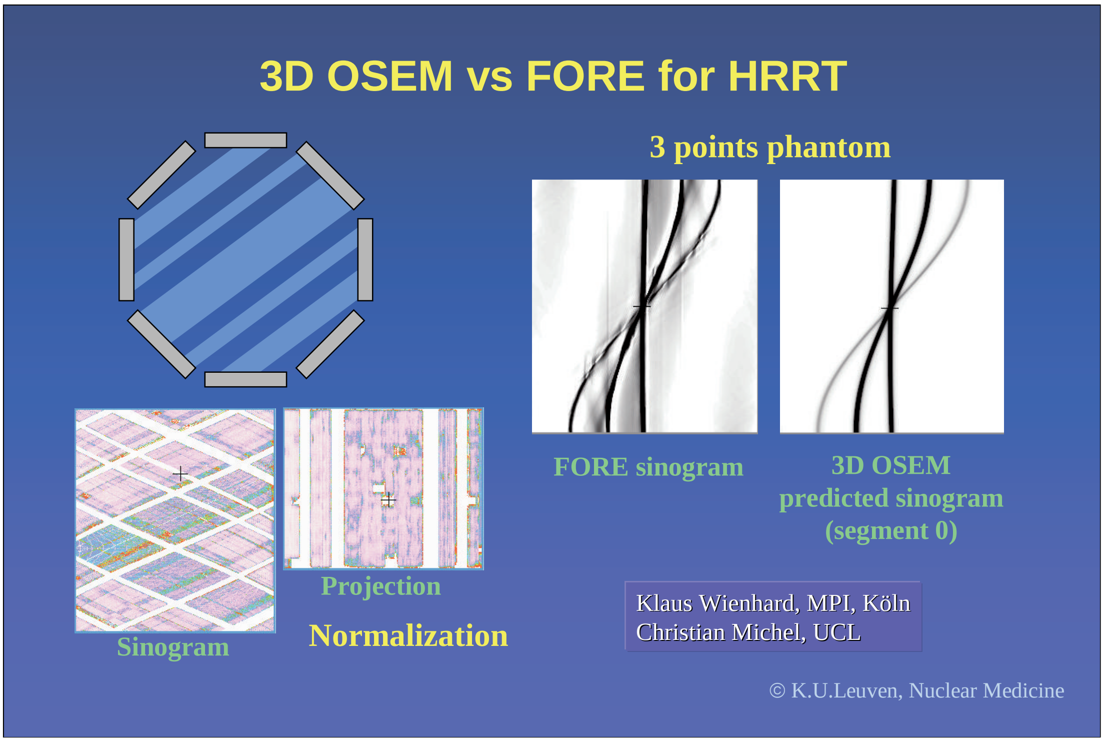
$$\lambda_j = \frac{\lambda_j^{\text{old}}}{\sum_i a_{ij}} \sum_i a_{ij} \frac{x_i + 2r_i}{\sum_k a_{ik}\lambda_k^{\text{old}} + 2r_i}$$

$x_i + 2r_i \geq 0$

With NEC-scaling, a nonnegativity requirement holds for the corrected sinogram. After randoms subtraction in PET, the corrected sinogram contains negatives, which have to be set to zero with NEC-scaling. This might introduce positive bias in low count regions (the background), leading to a compensating negative bias in other regions.

For this purpose, it is better to shift the distribution than to scale it. This yields the shifted Poisson approximation, proposed by Yavuz et al [35]. MLEM now requires that the shifted distribution is nonnegative. This requirement is less stringent: one still has to zero a few negatives but the effect on the bias is probably negligible.

MLEM converges a bit slower with the shifted distribution.



3D OSEM vs FORE for HRRT

3 points phantom

FORE sinogram

3D OSEM predicted sinogram (segment 0)

Sinogram

Projection

Normalization

Klaus Wienhard, MPI, Köln
Christian Michel, UCL

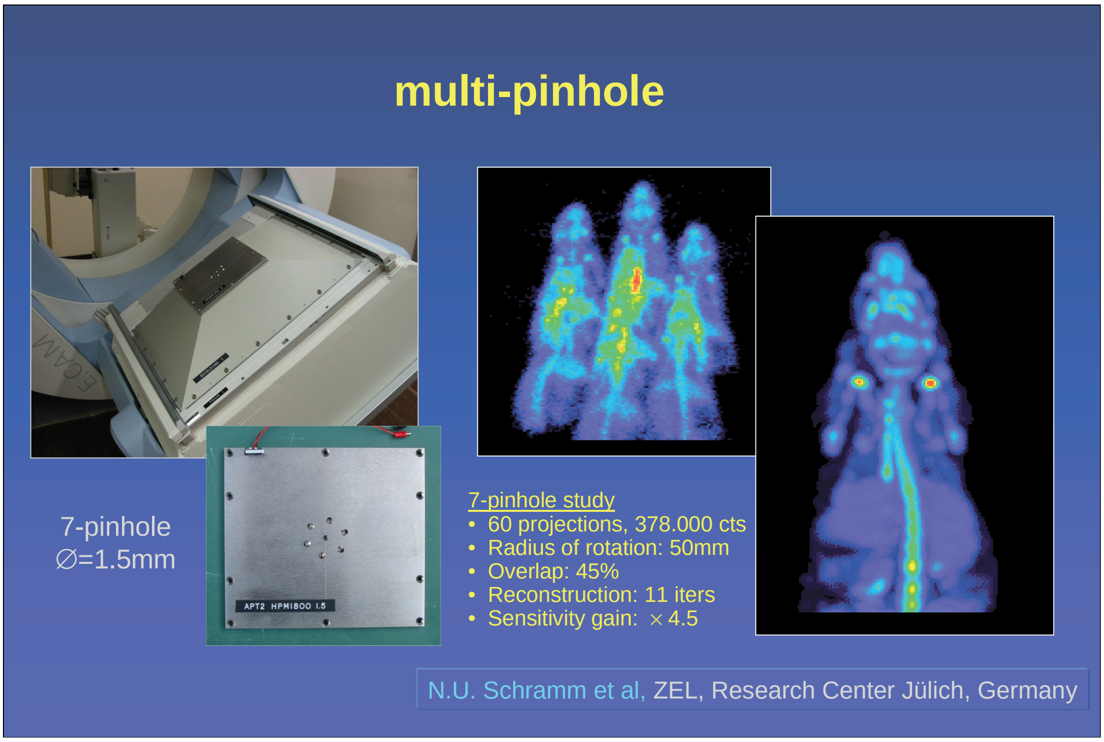
© K.U.Leuven, Nuclear Medicine

The HRRT (CTI-Siemens) is an octagonal LSO camera. There are gaps between the detector panels, which pose problems during reconstruction. The gaps appear as oblique lines of zero count in the sinogram, and as vertical lines in the projection. (The top level drawing explains why there are four vertical lines in the projection for this angle.) In addition, some detectors have defects, resulting in more missing data.

Fourier rebinning [39] was not designed to deal with the gaps, the 2D FORE sinogram shows artifacts.

OSEM reconstruction can be easily adapted to ignore the blind spots in the sinograms. The same 2D sinogram was now produced by reprojection of the 3D OSEM reconstruction. The artifacts have been eliminated (the missing data have been accurately predicted from the available data), indicating that OSEM produced an accurate reconstruction image.

Work done by C. Michel, then at UCL, Louvain-la-Neuve, Belgium, now at CTI, Knoxville, USA, and K Wienhard, MPI, Köln, Germany.



multi-pinhole

7-pinhole
∅=1.5mm

7-pinhole study
- 60 projections, 378.000 cts
- Radius of rotation: 50mm
- Overlap: 45%
- Reconstruction: 11 iters
- Sensitivity gain: × 4.5

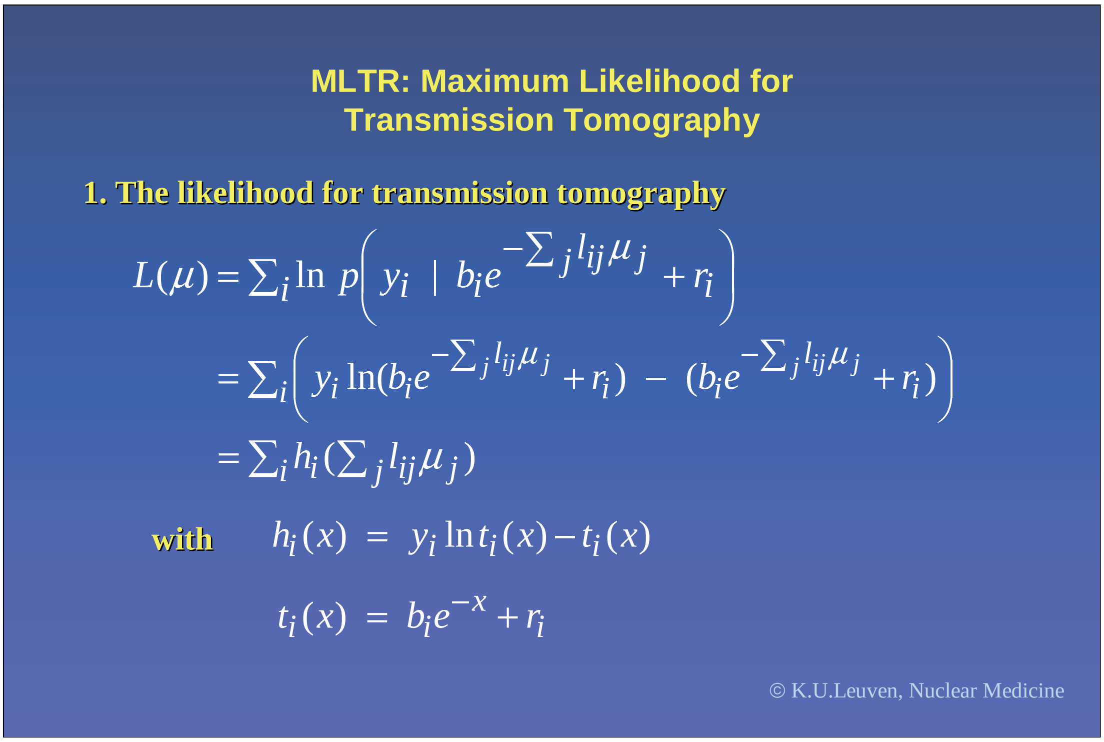
N.U. Schramm et al, ZEL, Research Center Jülich, Germany

To increase the sensitivity of pinhole SPECT, multi-pinhole configurations are being explored. Here is an example of mouse bone scan with a 7-pinhole design. The projections partially overlap. Extension of MLEM to this configuration is straightforward, one ``only‘‘ needs to adapt the system matrix (i.e. implementing a projector/backprojector pair). The sensitivity is estimated to be 4.5 times that of a single pinhole (to my knowledge, we currently don‘t have an expression to compute the effective sensitivity gain exactly).

NU Schramm [1], T Schurrat [2], M Béhé [2], M Schipper [2], U Engeland [3], G Ebel [3], TM Behr [2].

[1] ZEL, Research Center Jülich, [2] Dept. Nuclear Medicine, University of Marburg, [3] Scivis GmbH, Göttingen, GERMANY

MLTR



MLTR: Maximum Likelihood for Transmission Tomography

1. The likelihood for transmission tomography

$$L(\mu) = \sum_i \ln p\left(y_i \mid b_i e^{-\sum_j l_{ij}\mu_j} + r_i \right)$$

$$= \sum_i \left(y_i \ln(b_i e^{-\sum_j l_{ij}\mu_j} + r_i) - (b_i e^{-\sum_j l_{ij}\mu_j} + r_i) \right)$$

$$= \sum_i h_i\left(\sum_j l_{ij}\mu_j\right)$$

with $h_i(x) = y_i \ln t_i(x) - t_i(x)$

$t_i(x) = b_i e^{-x} + r_i$

© K.U.Leuven, Nuclear Medicine

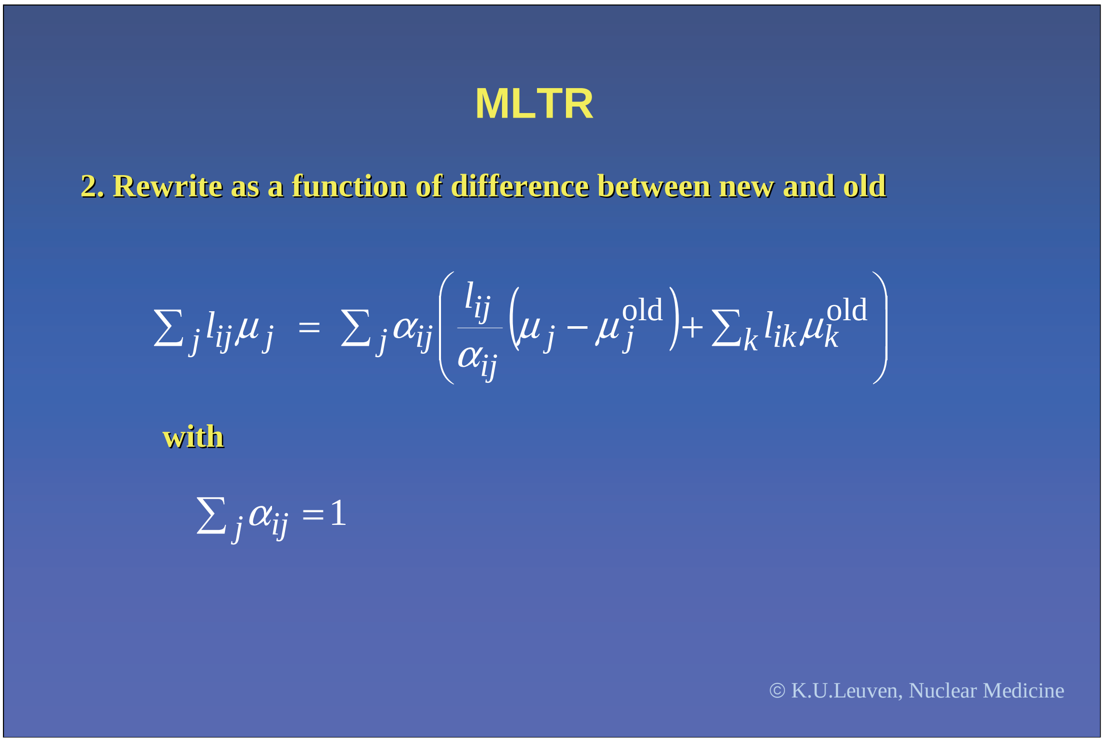
The EM -strategy can be applied to transmission tomography. However, in constrast to the emission case, the complete variables do not disappear from the equations, resulting in rather cumbersome expressions [4].

Consequently, various other strategies have been applied. [40]-[45].

Here we consider the strategy proposed by Fessler et al [44], which is an extension of an idea from De Pierro [10], developed for emission tomography.

We derive a simple expression for a simultaneous update strategy. The resulting algorithm shares many features with the MLEM algorithm for emission tomography: based on alternated projection and backprojection, higher frequencies have slower convergence, the number of iterations required is of the same order.

Fessler et al extend this to a grouped coordinate ascent method [44] by updating only groups of pixels, rather than the whole image. This strategy results in faster convergence and very different behavior: it makes the high frequencies converge faster. But implementation is not trivial.



MLTR

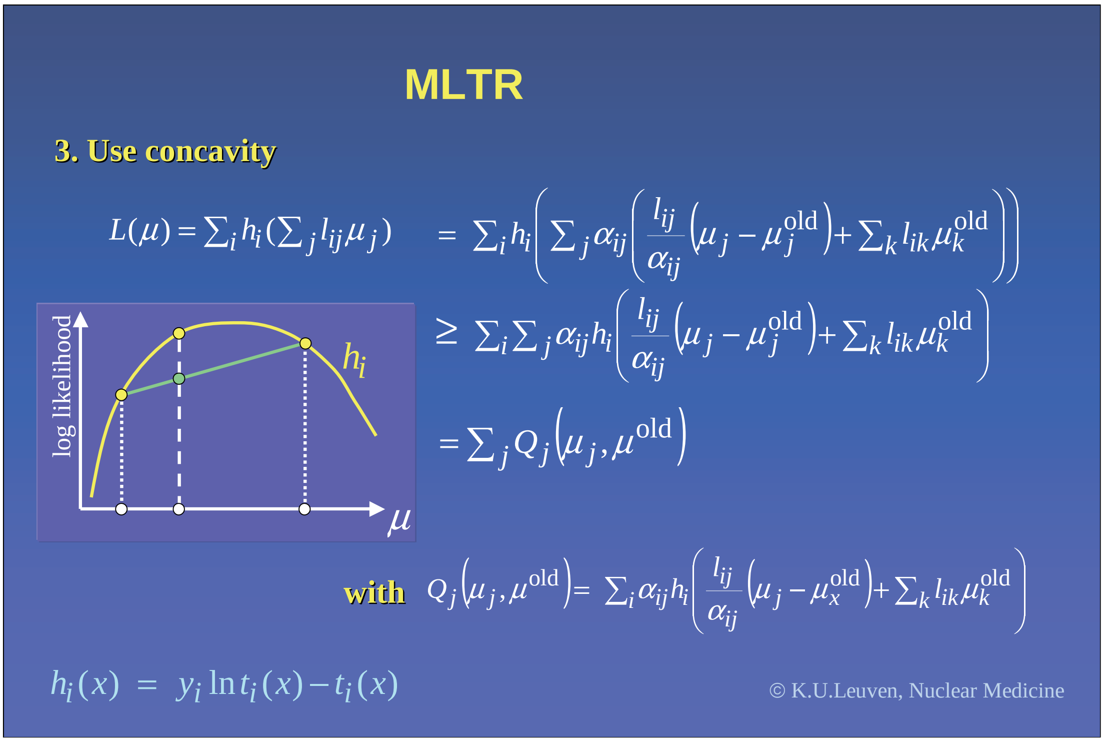
2. Rewrite as a function of difference between new and old

$$\sum_j l_{ij}\mu_j = \sum_j \alpha_{ij}\left(\frac{l_{ij}}{\alpha_{ij}}\left(\mu_j - \mu_j^{\text{old}}\right) + \sum_k l_{ik}\mu_k^{\text{old}}\right)$$

with

$$\sum_j \alpha_{ij} = 1$$

We are free to chose the α_{ij}, but they have to be positive and the sum over j should equal 1. Different choices lead to different algorithms.



MLTR

3. Use concavity

$$L(\mu) = \sum_i h_i\left(\sum_j l_{ij}\mu_j\right) \quad = \sum_i h_i\left(\sum_j \alpha_{ij}\left(\frac{l_{ij}}{\alpha_{ij}}\left(\mu_j - \mu_j^{\mathrm{old}}\right) + \sum_k l_{ik}\mu_k^{\mathrm{old}}\right)\right)$$

$$\geq \sum_i \sum_j \alpha_{ij} h_i\left(\frac{l_{ij}}{\alpha_{ij}}\left(\mu_j - \mu_j^{\mathrm{old}}\right) + \sum_k l_{ik}\mu_k^{\mathrm{old}}\right)$$

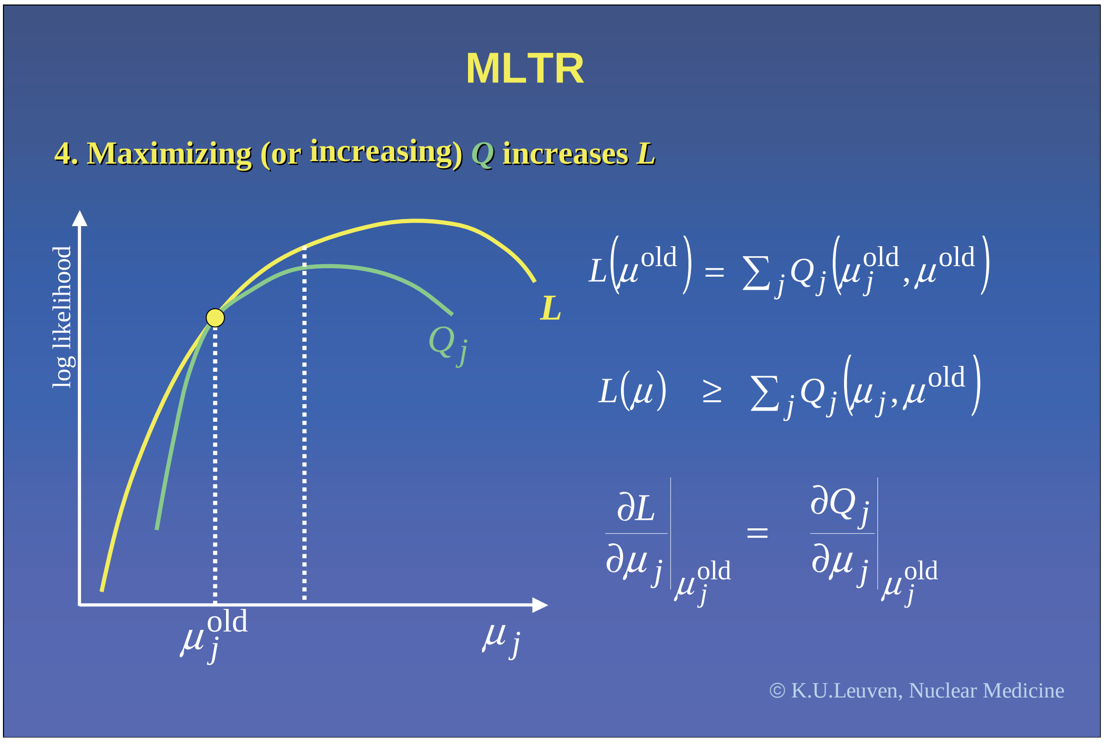
$$= \sum_j Q_j\left(\mu_j, \mu^{\mathrm{old}}\right)$$

with $Q_j\left(\mu_j, \mu^{\mathrm{old}}\right) = \sum_i \alpha_{ij} h_i\left(\frac{l_{ij}}{\alpha_{ij}}\left(\mu_j - \mu_x^{\mathrm{old}}\right) + \sum_k l_{ik}\mu_k^{\mathrm{old}}\right)$

$$h_i(x) = y_i \ln t_i(x) - t_i(x)$$

A function is concave if the matrix of second derivatives is negative definite. The log likelihood for transmission tomography is concave.

If *f* is concave, then $f(ax + by) \geq af(x) + bf(y)$.



MLTR

4. Maximizing (or increasing) *Q* increases *L*

$$L\left(\mu^{\text{old}}\right) = \sum_j Q_j\left(\mu_j^{\text{old}}, \mu^{\text{old}}\right)$$

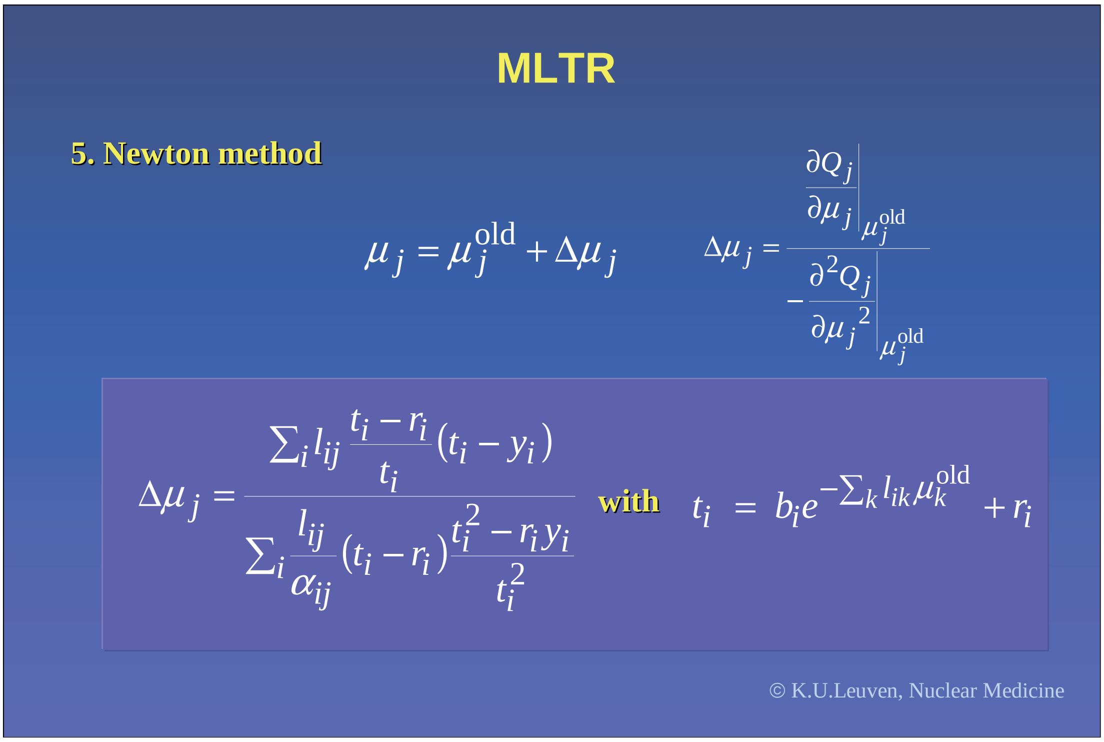
$$L(\mu) \geq \sum_j Q_j\left(\mu_j, \mu^{\text{old}}\right)$$

$$\left.\frac{\partial L}{\partial \mu_j}\right|_{\mu_j^{\text{old}}} = \left.\frac{\partial Q_j}{\partial \mu_j}\right|_{\mu_j^{\text{old}}}$$

© K.U.Leuven, Nuclear Medicine

We could try to maximize *L* directly. However, *L* is a complex function of the reconstruction values, the second derivatives are all non-zero. Consequently, we cannot apply simple one-dimensional optimization techniques.

By exploiting the concavity of *L*, we have derived a new function that is a very good approximation of *L* near the current reconstruction λ^{old}, and that can be written as a linear combination of one-dimensional functions Q_j. Maximizing this new function is much easier, and doing so is guaranteed to yield an increase in the likelihood *L*.



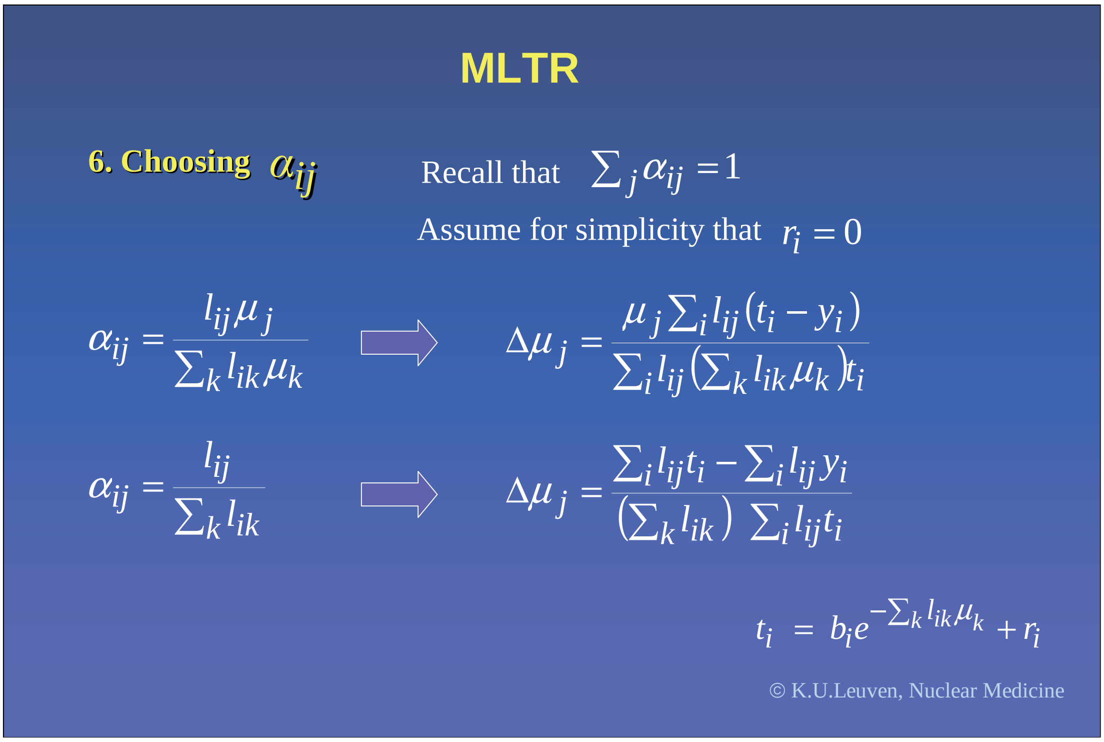
MLTR

5. Newton method

$$\mu_j = \mu_j^{\text{old}} + \Delta\mu_j \qquad \Delta\mu_j = \frac{\left.\dfrac{\partial Q_j}{\partial \mu_j}\right|_{\mu_j^{\text{old}}}}{-\left.\dfrac{\partial^2 Q_j}{\partial {\mu_j}^2}\right|_{\mu_j^{\text{old}}}}$$

$$\Delta\mu_j = \frac{\sum_i l_{ij} \dfrac{t_i - r_i}{t_i}(t_i - y_i)}{\sum_i \dfrac{l_{ij}}{\alpha_{ij}}(t_i - r_i)\dfrac{t_i^2 - r_i y_i}{t_i^2}} \quad \textbf{with} \quad t_i = b_i e^{-\sum_k l_{ik}\mu_k^{\text{old}}} + r_i$$

If Q_j were a quadratic function then Newton’s method would maximize Q_j in one step. It is not a quadratic function, so Newton’s method is not exact, and even not guaranteed to be effective. In practice, however, it works all the time. If you want to be really sure, you can verify by computing the likelihood.



MLTR

6. Choosing α_{ij}

Recall that $\sum_j \alpha_{ij} = 1$

Assume for simplicity that $r_i = 0$

$$\alpha_{ij} = \frac{l_{ij}\mu_j}{\sum_k l_{ik}\mu_k} \quad \Rightarrow \quad \Delta\mu_j = \frac{\mu_j \sum_i l_{ij}(t_i - y_i)}{\sum_i l_{ij}\left(\sum_k l_{ik}\mu_k\right)t_i}$$

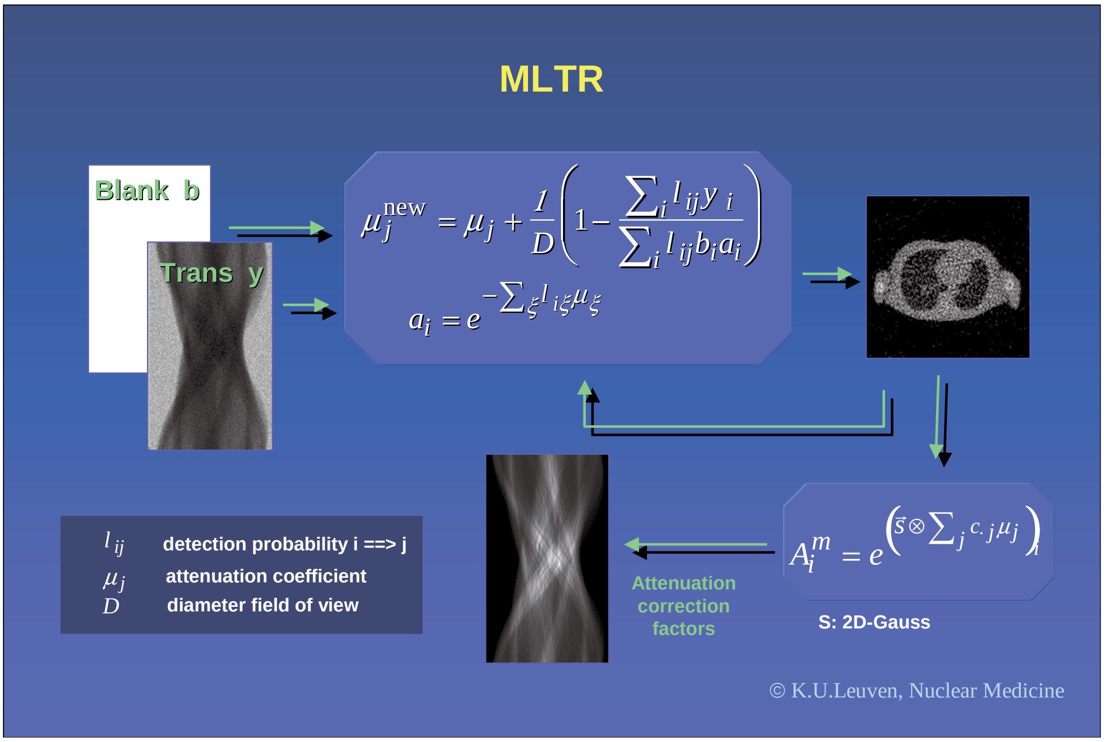
$$\alpha_{ij} = \frac{l_{ij}}{\sum_k l_{ik}} \quad \Rightarrow \quad \Delta\mu_j = \frac{\sum_i l_{ij}t_i - \sum_i l_{ij}y_i}{\left(\sum_k l_{ik}\right)\sum_i l_{ij}t_i}$$

$$t_i = b_i e^{-\sum_k l_{ik}\mu_k} + r_i$$

The first algorithm is the convex algorithm of Lange and Fessler [52], which was obtained earlier using a different (but similar) approach.

The second algorithm has been derived earlier in an intuitive way [45] and has received a better derivation thanks to Fessler et al. It requires one projection (to compute t_i) and one backprojection per iteration (the backprojection $\Sigma_i l_{ij} y_i$ of the measurement can be computed in advance). Note that these algorithms do not need the raw sinogram, the backprojection of the raw data is sufficient.

Lange’s convex algorithm needs one projection and two backprojections per iteration. However, in the few simulations we did, the iterations seemed a bit more effective, so both algorithms require similar processing times.



MLTR

Blank b

Trans y

$$\mu_j^{\text{new}} = \mu_j + \frac{1}{D}\left(1 - \frac{\sum_i l_{ij} y_i}{\sum_i l_{ij} b_i a_i}\right)$$

$$a_i = e^{-\sum_\xi l_{i\xi}\mu_\xi}$$

$$A_i^m = e^{\left(\vec{s} \otimes \sum_j c_{\cdot j}\mu_j\right)_i}$$

l_{ij}	**detection probability i ==> j**
μ_j	**attenuation coefficient**
D	**diameter field of view**

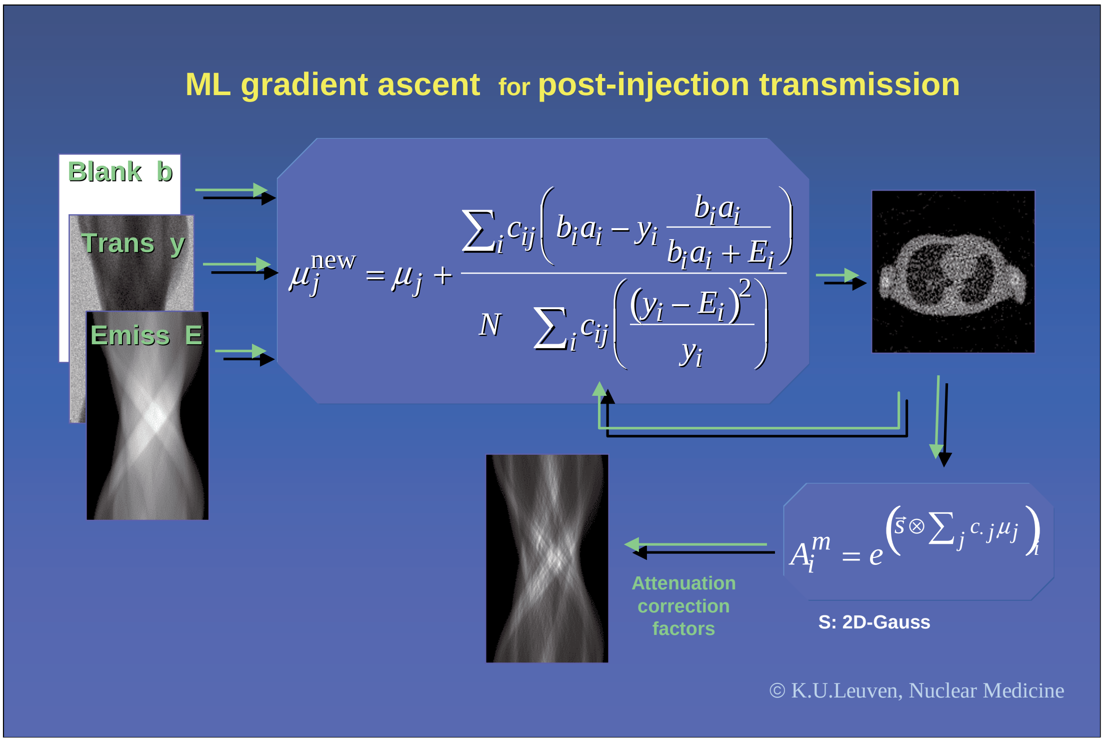
Attenuation correction factors

S: 2D-Gauss

© K.U.Leuven, Nuclear Medicine

Attenuation correction in PET can be done directly, by multiplying with the blank scan and dividing by the transmission scan.

Reprojection of MLTR reconstruction may improve the quality of the attenuation factor, by removing noise from the sinogram. Similar as in MLEM, only the inconsistent noise is removed. Consistent noise is included in the reconstruction image. The noise amplitude increases with increasing iteration number.



ML gradient ascent for post-injection transmission

$$\mu_j^{\text{new}} = \mu_j + \frac{\sum_i c_{ij}\left(b_i a_i - y_i \dfrac{b_i a_i}{b_i a_i + E_i}\right)}{N \quad \sum_i c_{ij}\left(\dfrac{(y_i - E_i)^2}{y_i}\right)}$$

$$A_i^m = e^{\left(\vec{s} \otimes \sum_j c_{\cdot j}\mu_j\right)_i}$$

S: 2D-Gauss

© K.U.Leuven, Nuclear Medicine

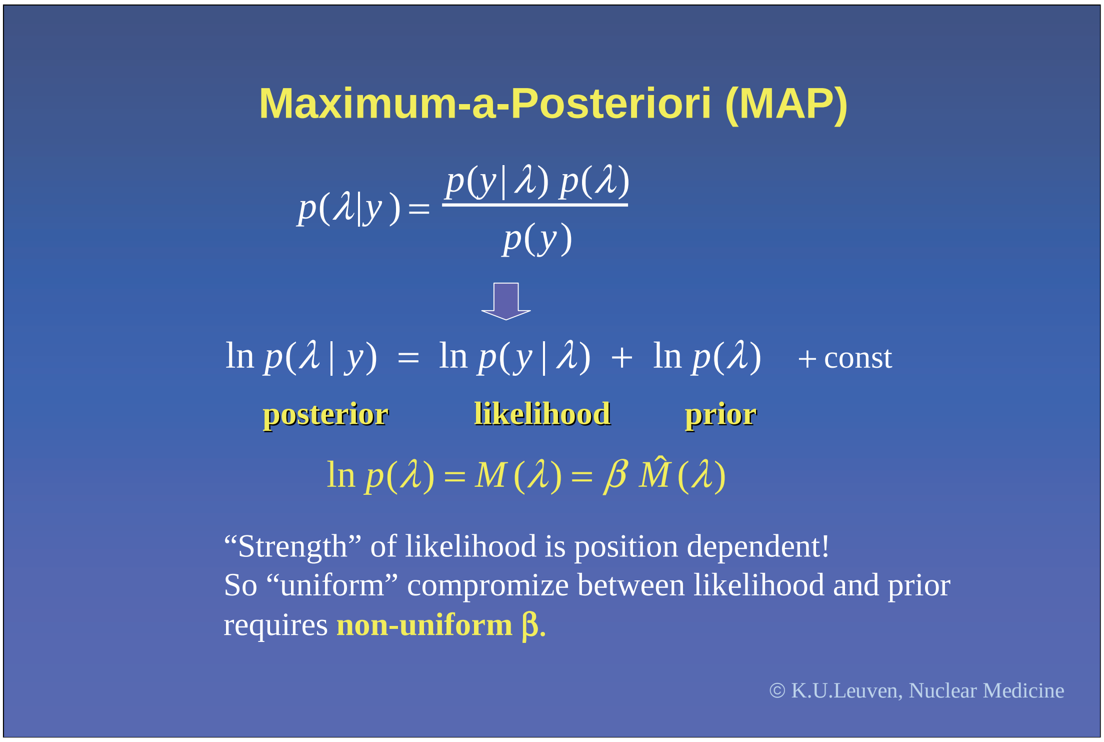
b_i: blank scan

y_i: transmission scan

E_i: estimate of contamination, assumed to be noise free

In windowed PET post-injection transmission scanning [84], the noise in the contamination is higher than that in the corresponding emission scan, so the assumption that E_i is noise-free is acceptable for this case.

MAP



Maximum-a-Posteriori (MAP)

$$p(\lambda|y) = \frac{p(y|\lambda)\; p(\lambda)}{p(y)}$$

$$\Downarrow$$

$$\ln p(\lambda \mid y) \;=\; \ln p(y \mid \lambda) \;+\; \ln p(\lambda) \quad + \text{const}$$

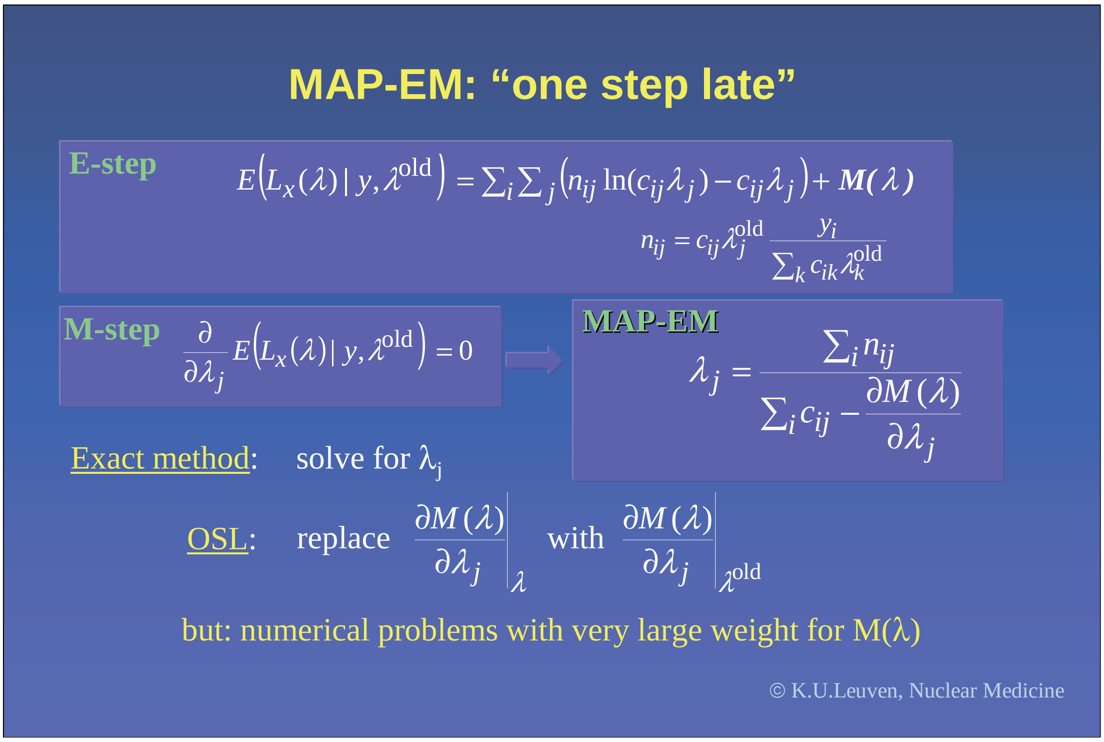
posterior **likelihood** **prior**

$$\ln p(\lambda) = M(\lambda) = \beta\, \hat{M}(\lambda)$$

“Strength” of likelihood is position dependent!
So “uniform” compromize between likelihood and prior
requires **non-uniform β.**

MAP reconstruction: [46]-[57].

Some authors propose to use a position dependent β, to obtain more uniform image characteristics [58]-[60].



MAP-EM: "one step late"

E-step $E\left(L_x(\lambda)\mid y,\lambda^{\text{old}}\right) = \sum_i \sum_j \left(n_{ij}\ln(c_{ij}\lambda_j) - c_{ij}\lambda_j\right) + \boldsymbol{M(\lambda)}$

$$n_{ij} = c_{ij}\lambda_j^{\text{old}} \frac{y_i}{\sum_k c_{ik}\lambda_k^{\text{old}}}$$

M-step $\frac{\partial}{\partial \lambda_j} E\left(L_x(\lambda)\mid y,\lambda^{\text{old}}\right) = 0$ ⇒

MAP-EM

$$\lambda_j = \frac{\sum_i n_{ij}}{\sum_i c_{ij} - \frac{\partial M(\lambda)}{\partial \lambda_j}}$$

Exact method: solve for λ_j

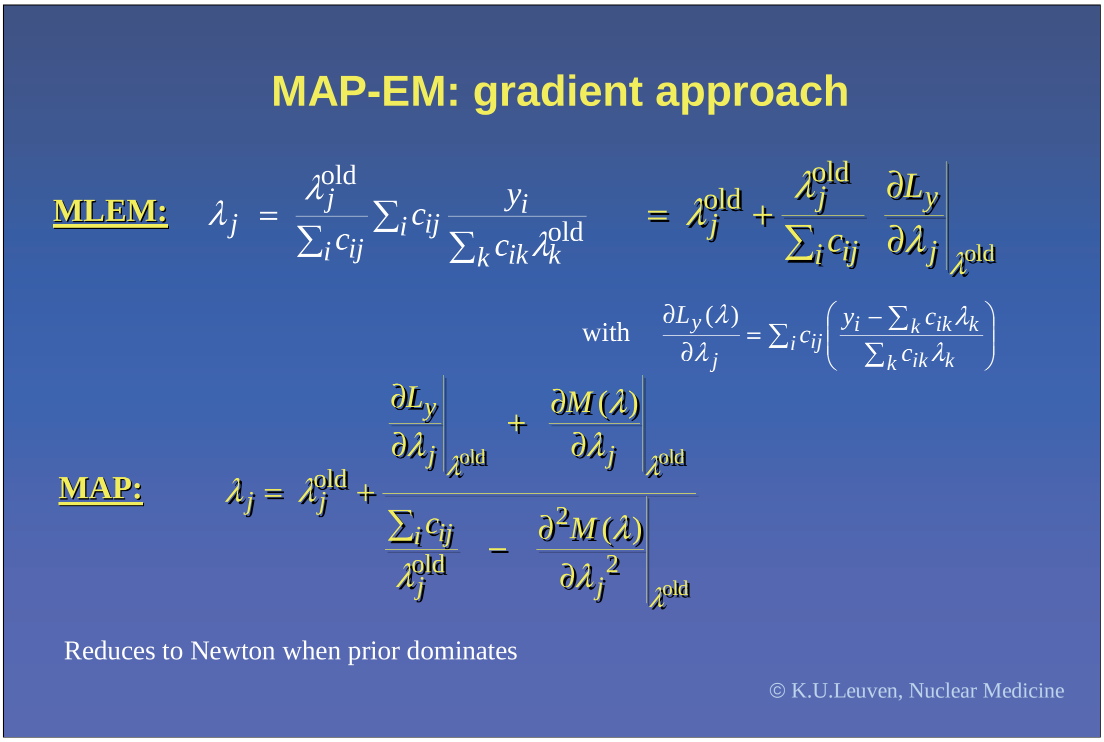
OSL: replace $\left.\frac{\partial M(\lambda)}{\partial \lambda_j}\right|_{\lambda}$ with $\left.\frac{\partial M(\lambda)}{\partial \lambda_j}\right|_{\lambda^{\text{old}}}$

but: numerical problems with very large weight for $M(\lambda)$

© K.U.Leuven, Nuclear Medicine

OSL (one step late) approach: [49].

OSL works for priors with reasonable strength. In extreme cases (usually not clinically useful) it may have numerical problems, because the denominator becomes zero or negative.



MAP-EM: gradient approach

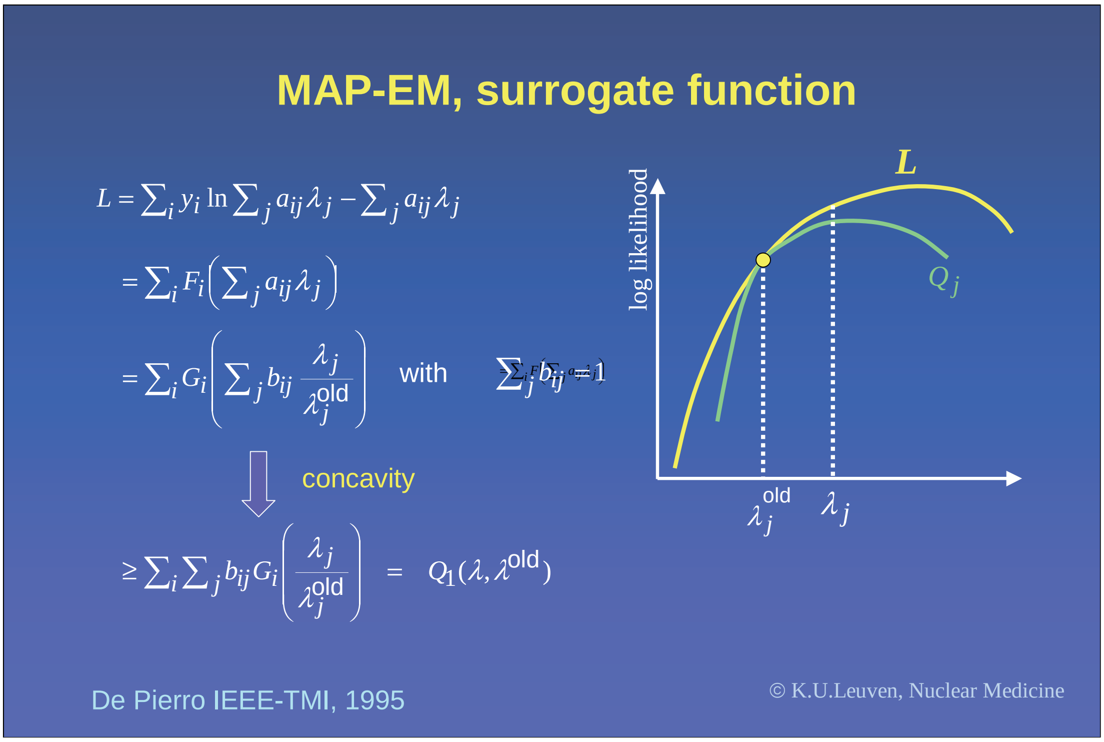
MLEM: $$\lambda_j = \frac{\lambda_j^{\text{old}}}{\sum_i c_{ij}} \sum_i c_{ij} \frac{y_i}{\sum_k c_{ik}\lambda_k^{\text{old}}} = \lambda_j^{\text{old}} + \frac{\lambda_j^{\text{old}}}{\sum_i c_{ij}} \left.\frac{\partial L_y}{\partial \lambda_j}\right|_{\lambda^{\text{old}}}$$

with $$\frac{\partial L_y(\lambda)}{\partial \lambda_j} = \sum_i c_{ij} \left(\frac{y_i - \sum_k c_{ik}\lambda_k}{\sum_k c_{ik}\lambda_k} \right)$$

MAP: $$\lambda_j = \lambda_j^{\text{old}} + \frac{\left.\frac{\partial L_y}{\partial \lambda_j}\right|_{\lambda^{\text{old}}} + \left.\frac{\partial M(\lambda)}{\partial \lambda_j}\right|_{\lambda^{\text{old}}}}{\frac{\sum_i c_{ij}}{\lambda_j^{\text{old}}} - \left.\frac{\partial^2 M(\lambda)}{\partial {\lambda_j}^2}\right|_{\lambda^{\text{old}}}}$$

Reduces to Newton when prior dominates

For those extreme cases, we use a more conservative (and slower) gradient ascent version [83]. It reduces to MLEM if the likelihood dominates, and to an approximate Newton's method if the prior dominates. With new priors, we check convergence of the prior alone. If that converges, the combinations converges as well. We have no proof. We never saw it fail, but it is rather slow.



MAP-EM, surrogate function

$$L = \sum_i y_i \ln \sum_j a_{ij}\lambda_j - \sum_j a_{ij}\lambda_j$$

$$= \sum_i F_i\left(\sum_j a_{ij}\lambda_j\right)$$

$$= \sum_i G_i\left(\sum_j b_{ij}\frac{\lambda_j}{\lambda_j^{\text{old}}}\right) \quad \text{with} \quad \sum_j b_{ij} = 1$$

concavity

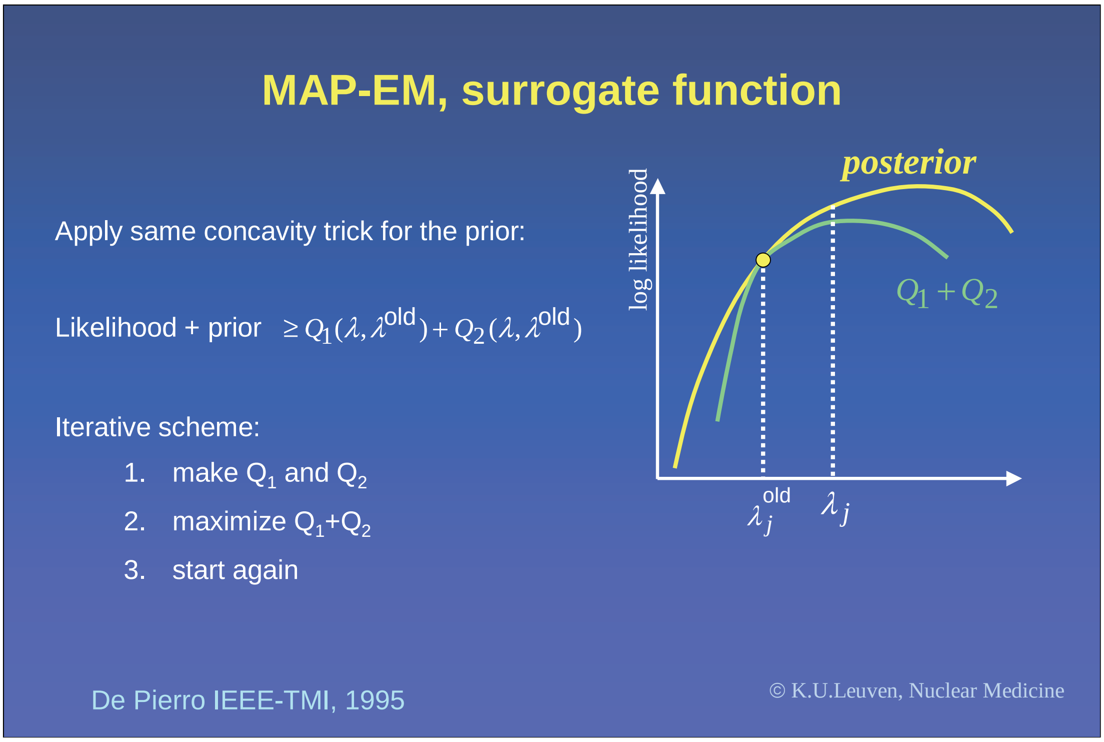
$$\geq \sum_i \sum_j b_{ij} G_i\left(\frac{\lambda_j}{\lambda_j^{\text{old}}}\right) = Q_1(\lambda, \lambda^{\text{old}})$$

De Pierro IEEE-TMI, 1995

© K.U.Leuven, Nuclear Medicine

De Pierro derived a technique with convergence proof in [10]. The method introduces a surrogate function which coincides with the likelihood up to the first derivative. It does so by rewriting the likelihood as a function of the ratio between the new and the current reconstruction in every pixel. This is different from the surrogate typically used in transmission, which is a function of the difference between the new and the old reconstruction image.

For the prior, a similar surrogate function is derived. The sum of both surrogate functions is used to maximise the sum of likelihood and prior. With methods like these, MAP should converge faster than MLEM per iteration.



MAP-EM, surrogate function

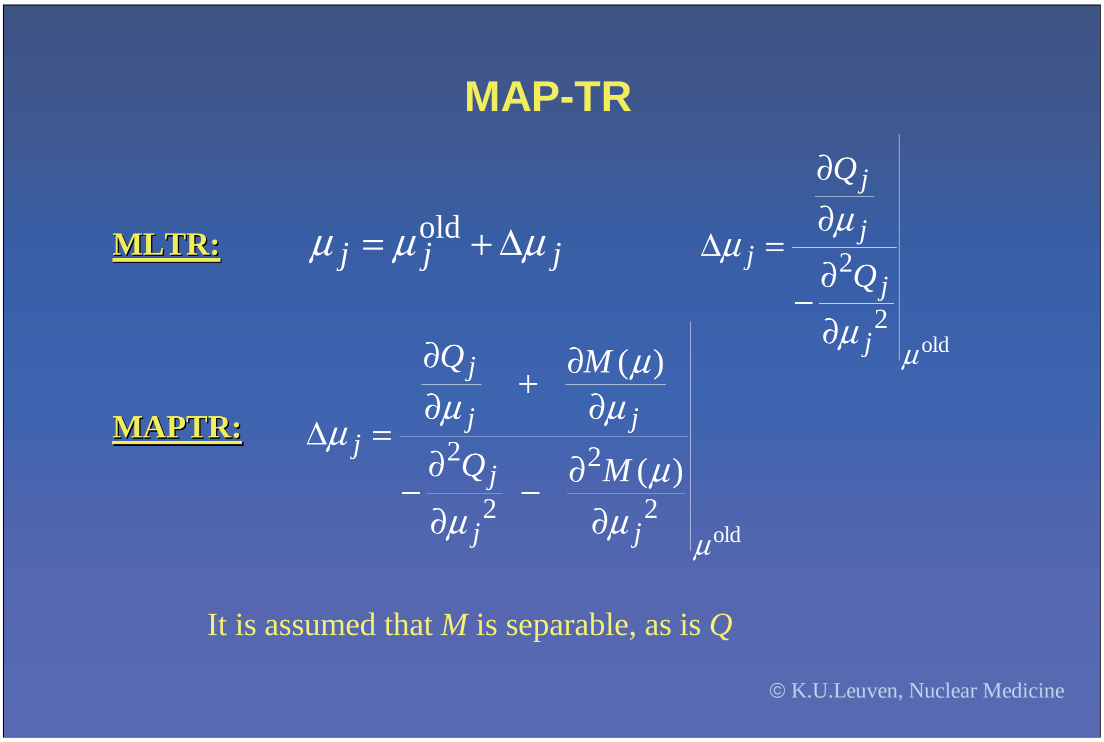
Apply same concavity trick for the prior:

Likelihood + prior $\geq Q_1(\lambda, \lambda^{\text{old}}) + Q_2(\lambda, \lambda^{\text{old}})$

Iterative scheme:

1. make Q_1 and Q_2
2. maximize Q_1+Q_2
3. start again

De Pierro IEEE-TMI, 1995



MAP-TR

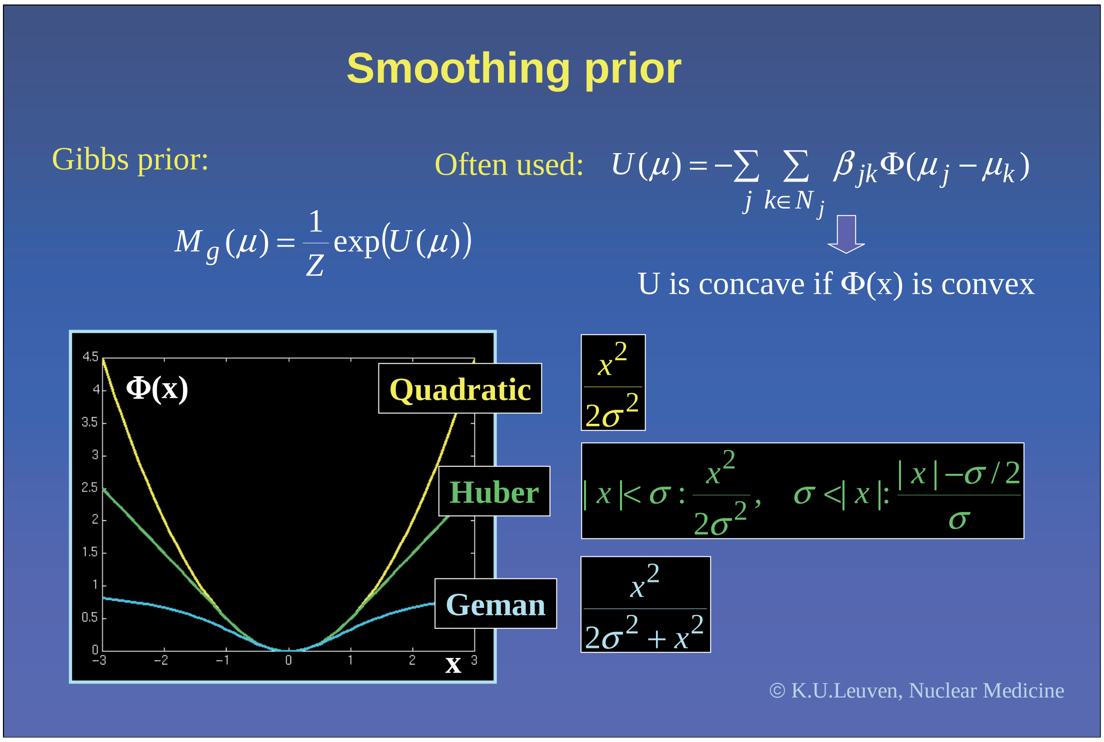
MLTR: $\mu_j = \mu_j^{\text{old}} + \Delta\mu_j$ $\quad$ $\Delta\mu_j = \left.\dfrac{\dfrac{\partial Q_j}{\partial \mu_j}}{-\dfrac{\partial^2 Q_j}{\partial {\mu_j}^2}}\right|_{\mu^{\text{old}}}$

MAPTR: $\Delta\mu_j = \left.\dfrac{\dfrac{\partial Q_j}{\partial \mu_j} + \dfrac{\partial M(\mu)}{\partial \mu_j}}{-\dfrac{\partial^2 Q_j}{\partial {\mu_j}^2} - \dfrac{\partial^2 M(\mu)}{\partial {\mu_j}^2}}\right|_{\mu^{\text{old}}}$

It is assumed that M is separable, as is Q

Q_j is the surrogate function.

If the prior is also separable, the method still involves only one-dimensional optimization.



Smoothing prior

Gibbs prior:

$$M_g(\mu) = \frac{1}{Z}\exp\left(U(\mu)\right)$$

Often used: $U(\mu) = -\sum_j \sum_{k\in N_j} \beta_{jk}\Phi(\mu_j - \mu_k)$

⇩

U is concave if Φ(x) is convex

Quadratic: $\frac{x^2}{2\sigma^2}$

Huber: $|x|<\sigma: \frac{x^2}{2\sigma^2}, \quad \sigma<|x|: \frac{|x|-\sigma/2}{\sigma}$

Geman: $\frac{x^2}{2\sigma^2+x^2}$

© K.U.Leuven, Nuclear Medicine

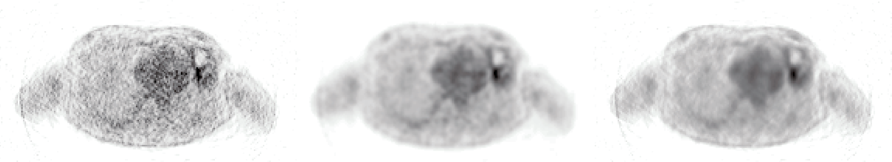
N_j is the neighborhood of pixel j. Usually the closest four or eight neighbors are used. (In our experience, 4 is fine and twice as fast as 8 neighbors.)

Z is a normalization factor, which can be ignored during optimization.

Since the likelihood is concave, combination with a concave prior produces a concave posterior.

The Huber and Geman priors preserve (or at least penalize less) differences that are large compared to σ. Tuning s is relatively straightforward in transmission tomography (very predictable absolute voxel intensities), but is difficult in emission tomography (unpredictable voxel intensities).

U is not concave for the Geman prior, it introduces local maxima. Optimizing a posterior with local maxima requires huge processing times (simulated annealing). However, for some applications (e.g. transmission tomography) it is possible to write programs fast enough for clinical use, that find a local maximum corresponding to an excellent reconstruction.



MLEM

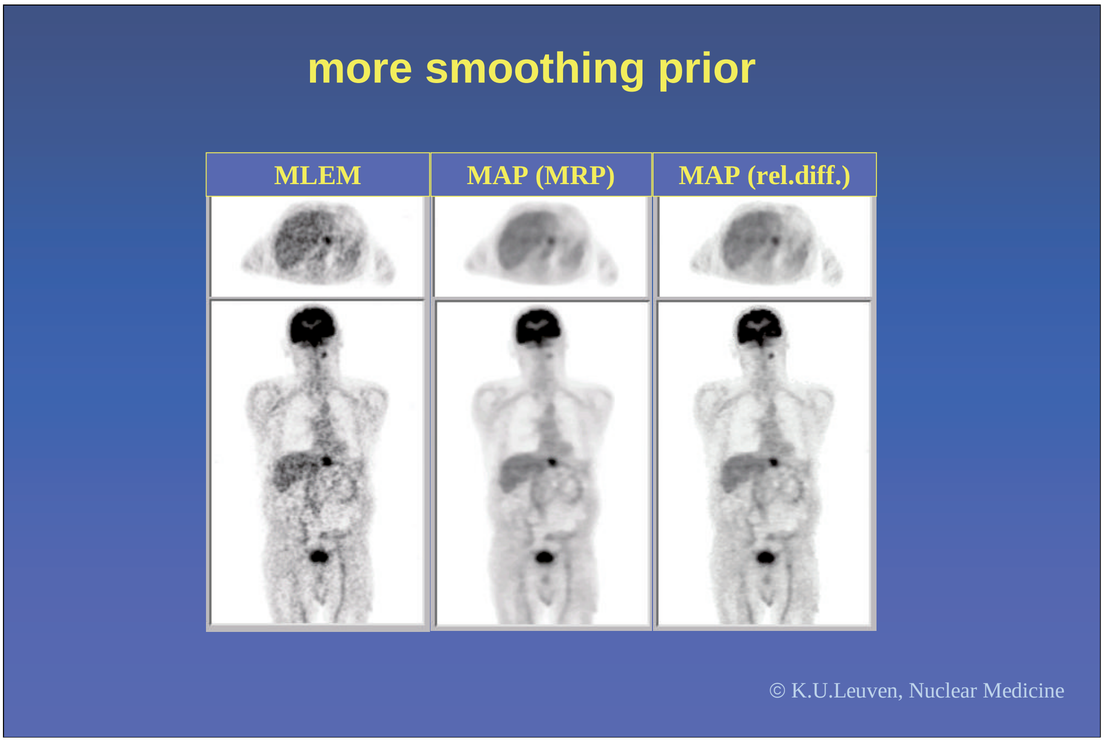
MLEM postsmoothed

MAP (Huber prior)

PET emission image

© K.U.Leuven, Nuclear Medicine

90 iterations were applied. Post-smoothing was done with a 2D Gaussian with FWHM = 3 pixels. Note that the prior produces better noise suppression in the mediastinum, and better resolution near the skin; it is not equivalent to a simple post-smoothing filter.



more smoothing prior

MLEM

MAP (MRP)

MAP (rel.diff.)

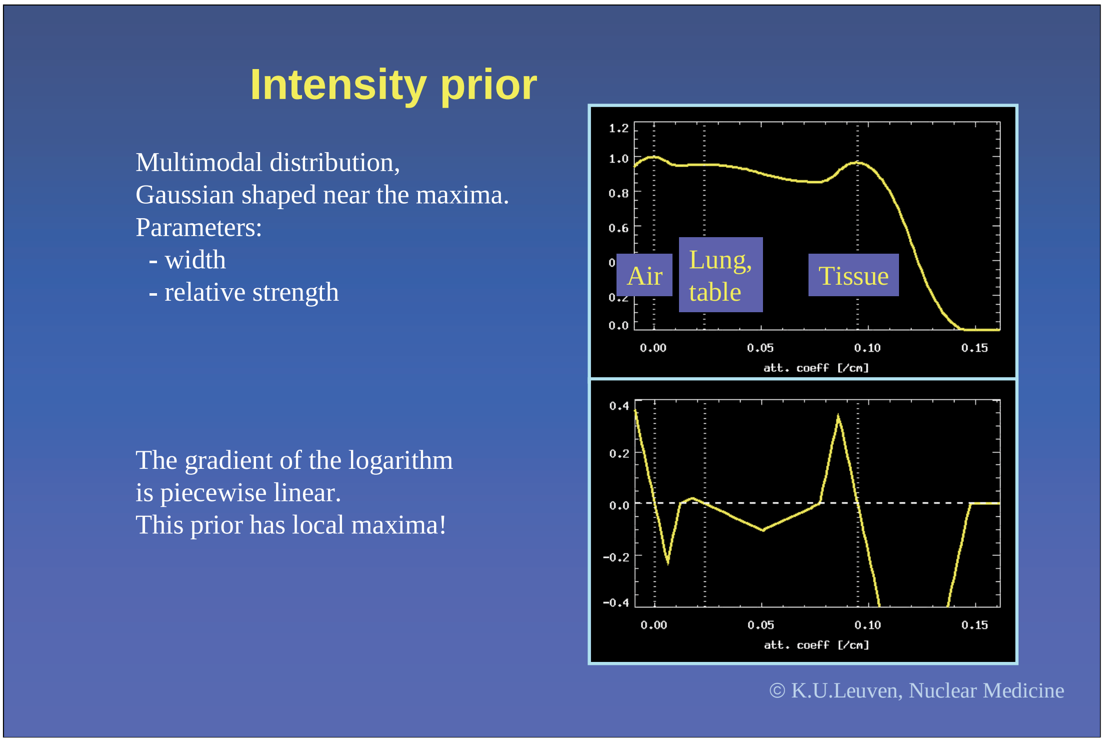
Many other smoothing priors have been proposed. One that has been used by several groups is the median root prior, developed by Alenius et al. [54]. A modified version with better theoretical derivation was proposed by Hsiao et al [85].

The median root prior penalizes the relative difference between the image and a median filtered version of that image. As a result, it penalizes hot spots that are small compared to the size of the filter, but it preserves larger spots and edges.

The relative difference prior [86] penalizes the relative difference between neighboring pixels. It has tolerance for large relative differences.

The reconstructions show that MRP can combine very strong noise suppression with excellent edge preservation. A limitation is the suppression of hot spots. The relative difference prior preserves the hot spot better, but smooths more over edges than MRP.

MRP and the relative difference prior both penalize relative differences rather than absolute ones. For that reason, tuning them for emission tomography is easier than with Huber or Geman, where the voxel intensities (and their differences) are unpredictable.

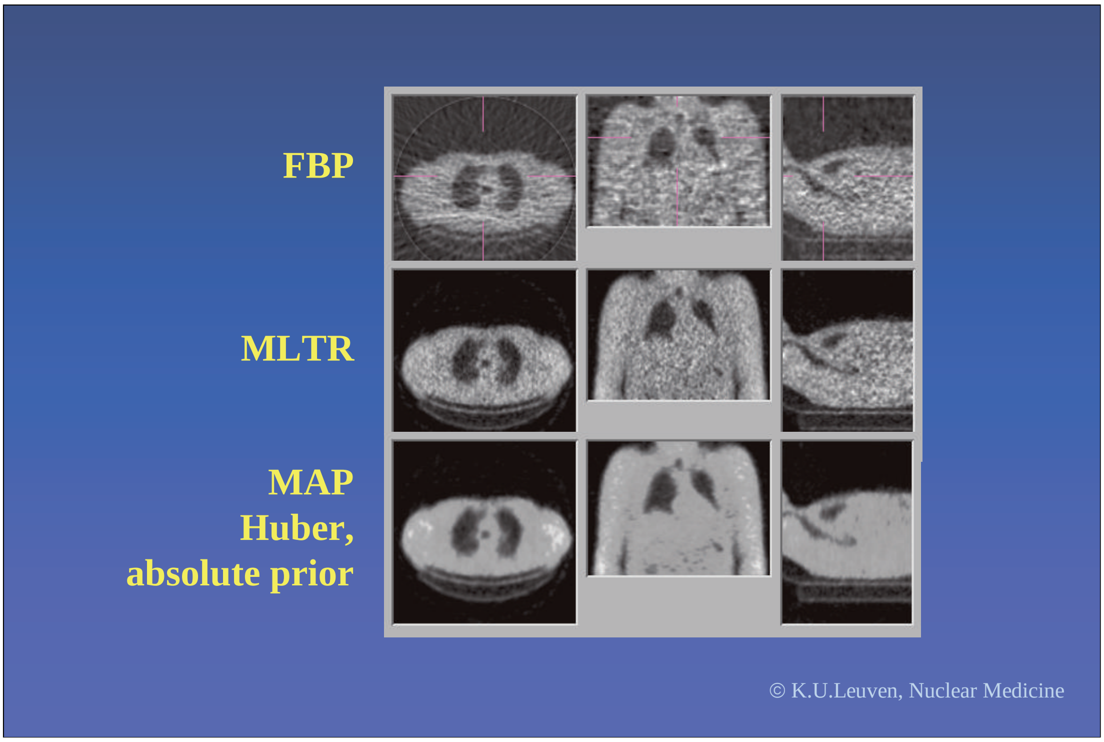


Intensity prior

Multimodal distribution,
Gaussian shaped near the maxima.
Parameters:
- width
- relative strength

The gradient of the logarithm
is piecewise linear.
This prior has local maxima!

This prior is useful to produce a “segmented” attenuation image with a MAP reconstruction algorithm [55].

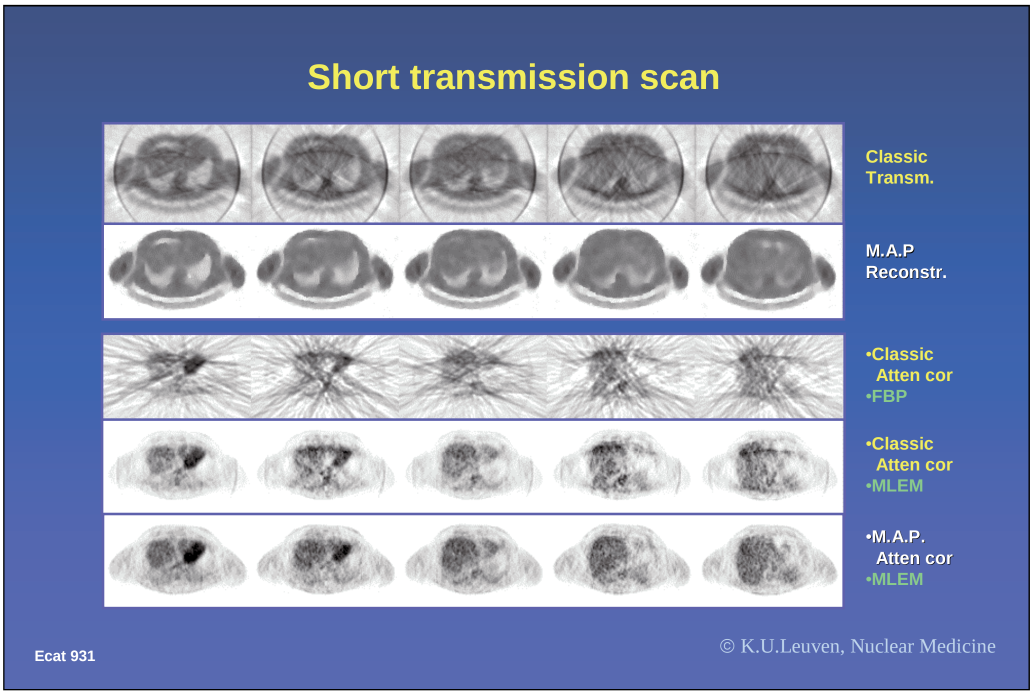


FBP

MLTR

MAP Huber, absolute prior

Example of multi-bed PET transmission study.



Short transmission scan

Classic
Transm.

M.A.P
Reconstr.

- Classic
  Atten cor
- FBP

- Classic
  Atten cor
- MLEM

- M.A.P.
  Atten cor
- MLEM

Ecat 931

PET study:

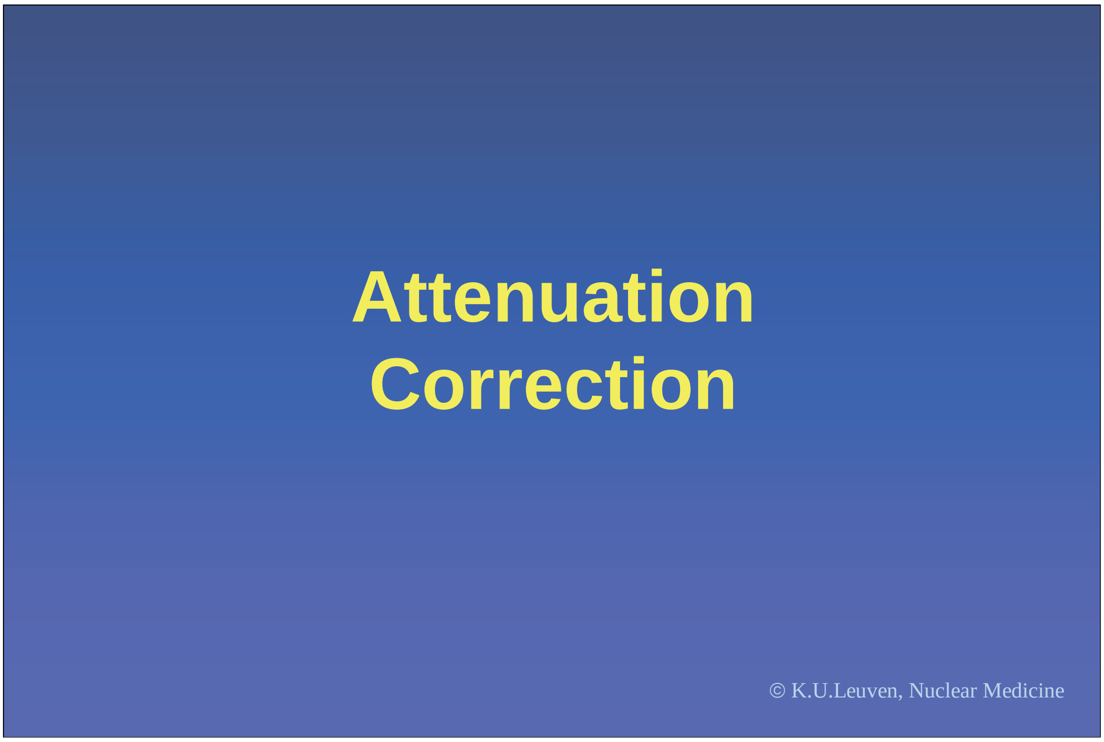
Top row: FBP-reconstruction of ln(b/y)

Row 2: MAP-reconstruction of (b,y)

Row 3: FBP-reconstruction of tracer distribution with classical attenuation correction (= smoothed blank/smoothed transmission)

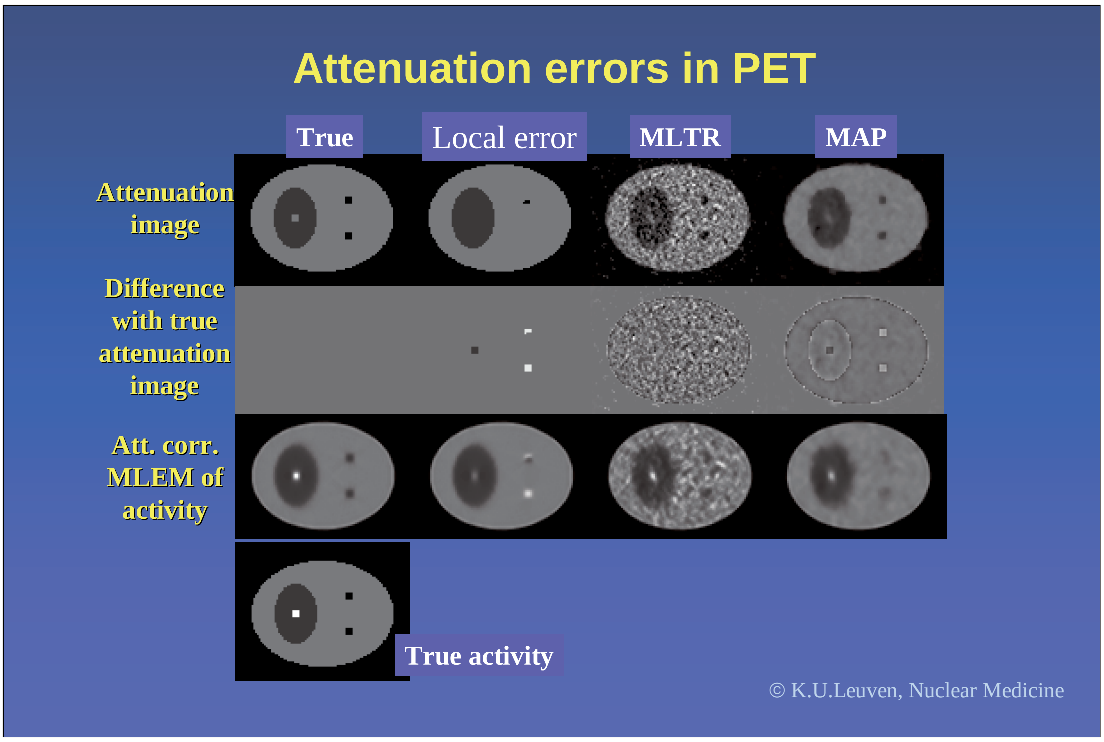
Row 4: MLEM-reconstruction of tracer distribution with classical attenuation correction

Row 5: MLEM-reconstruction of tracer distribution with MAP-based attenuation correction.



Attenuation Correction

Evaluation of attenuation correction: [69]-[78].



Attenuation errors in PET

True **Local error** **MLTR** **MAP**

Attenuation image

Difference with true attenuation image

Att. corr. MLEM of activity

True activity

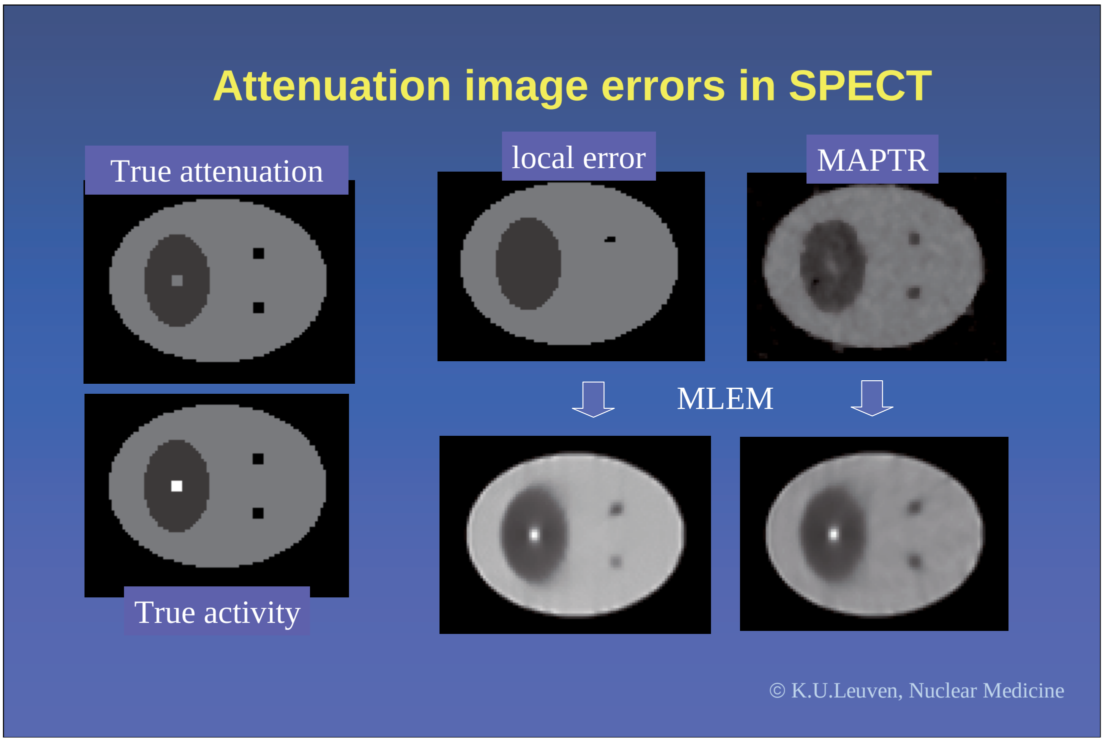
Simulated elliptical phantom with a single “lung”, and a hot dense region in that lung. In addition, the tissue region contains two “air bubbles” (no attenuation, no activity).

This simulation shows that a few systematic errors in an otherwise perfect attenuation map may lead to severe artifacts in the attenuation corrected reconstruction:

the hot spot has reduced activity because its attenuation is underestimated

the cold spots with zero attenuation (e.g. air in colon or stomach) are reconstructed as hot spots.

The MAP-reconstructed attenuation map has errors in every pixel. However, this reconstruction agrees better with the transmission data, making it more useful for attenuation correction [55].



Attenuation image errors in SPECT

True attenuation

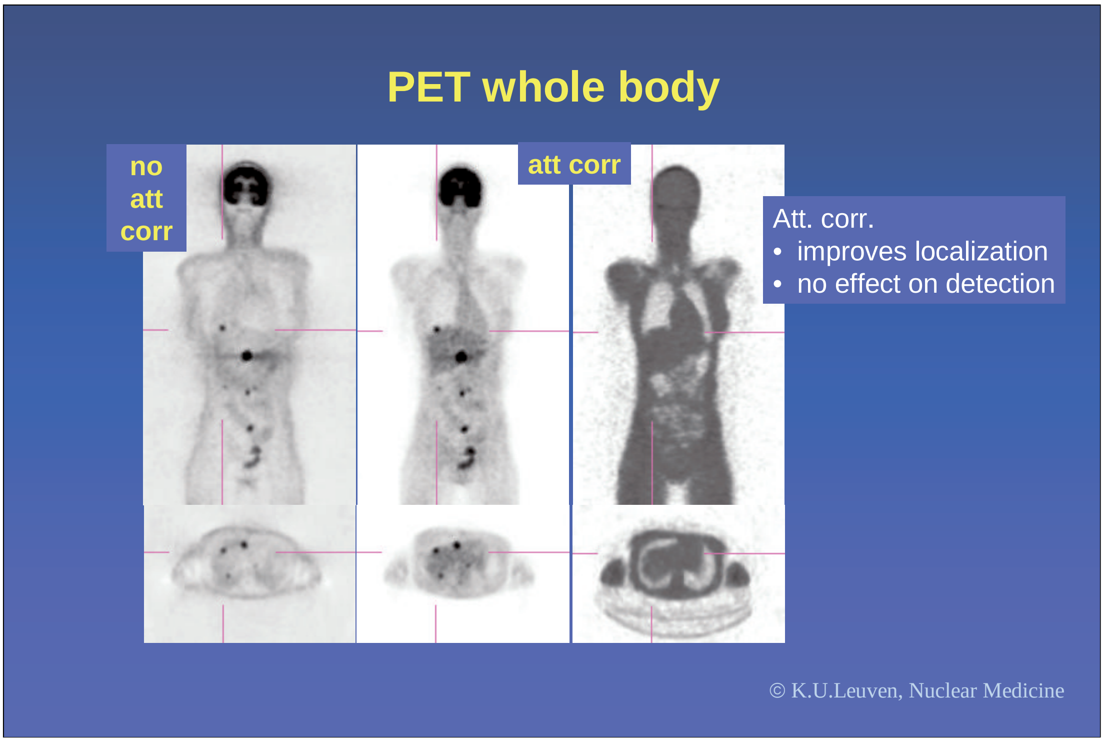
local error

MAPTR

MLEM

True activity

In SPECT, local systematic errors have not the same dramatic effect as in PET.

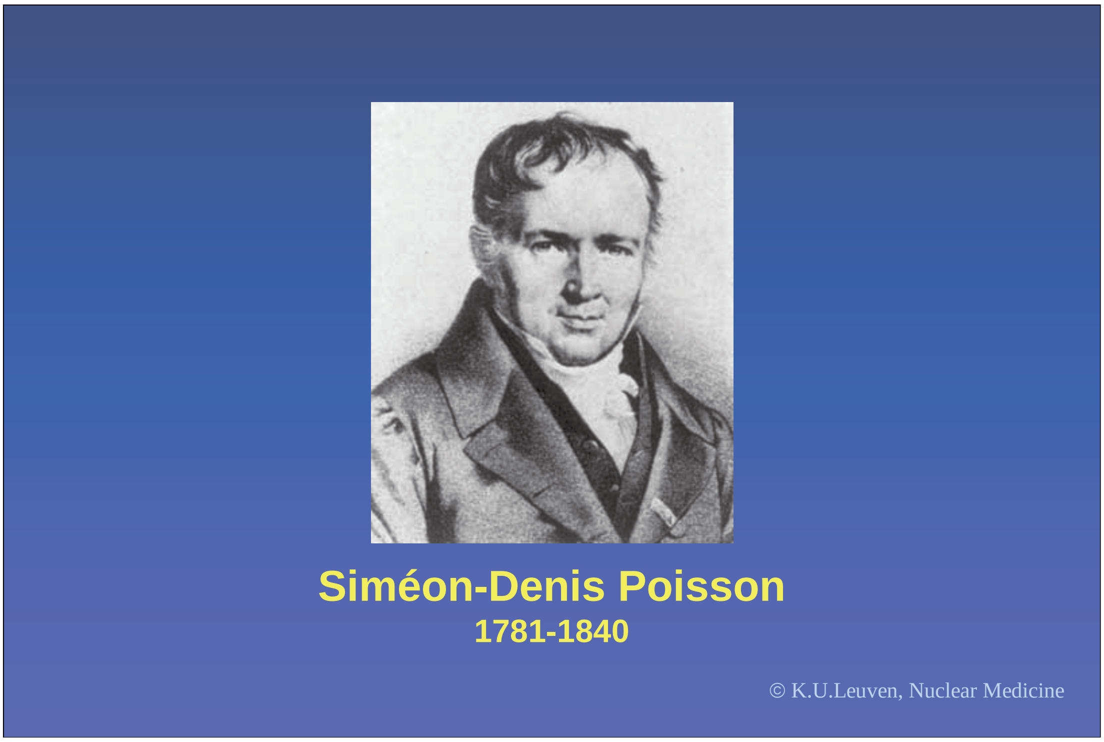


PET whole body

no att corr

att corr

Att. corr.
- improves localization
- no effect on detection



Siméon-Denis Poisson

1781-1840