Medical Imaging

Iterative Reconstruction with ML-EM

Klaus Mueller

Computer Science Department

Stony Brook University

Algebraic/gradient methods do not model statistical effects in the underlying data
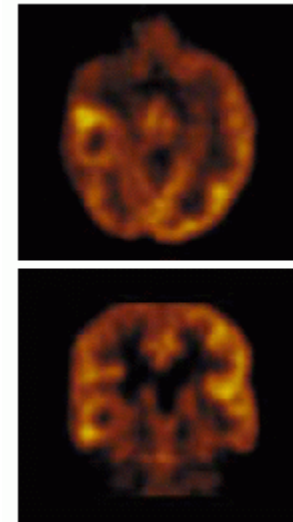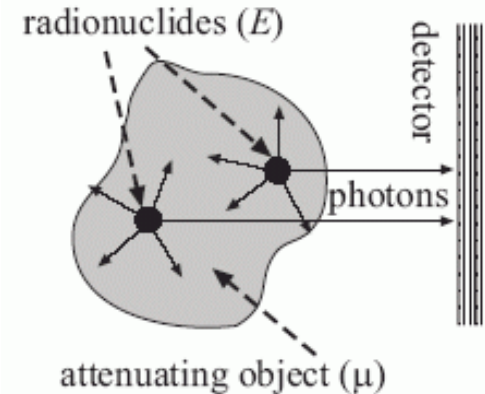
- this is OK for CT (within reason)

However, the emission of radiation from radionuclides is highly statistical

- the direction is chosen at random
- similar metabolic activities may not emit the same radiation
- not all radiation is actually collected (collimators reject many photons)
- in low-dose CT, noise is also a significant problem

Need a reconstruction method that can accounts for these statistical effects

- Maximum Likelihood – Expectation Maximization (ML-EM) is one such method

Overall Concept of ML-EM

Setup:

- there are three types of variables: observed data, unobserved data, and model parameters
- due to this, there is a many-to-one mapping of parameters $\rightarrow$ data

Goal:

- estimate the model parameters using the observed data

Solution:

- use an iterative solver that finds an optimal solution (but not necessarily an accurate one)
- possible algorithms are: Newton-type (for example, conjugate gradient), ART, EM
- EM does not require the computation of gradients and it is also stable (will always converge)
- EM will converge to a solution of maximum likelihood (but not necessarily the global maximum)
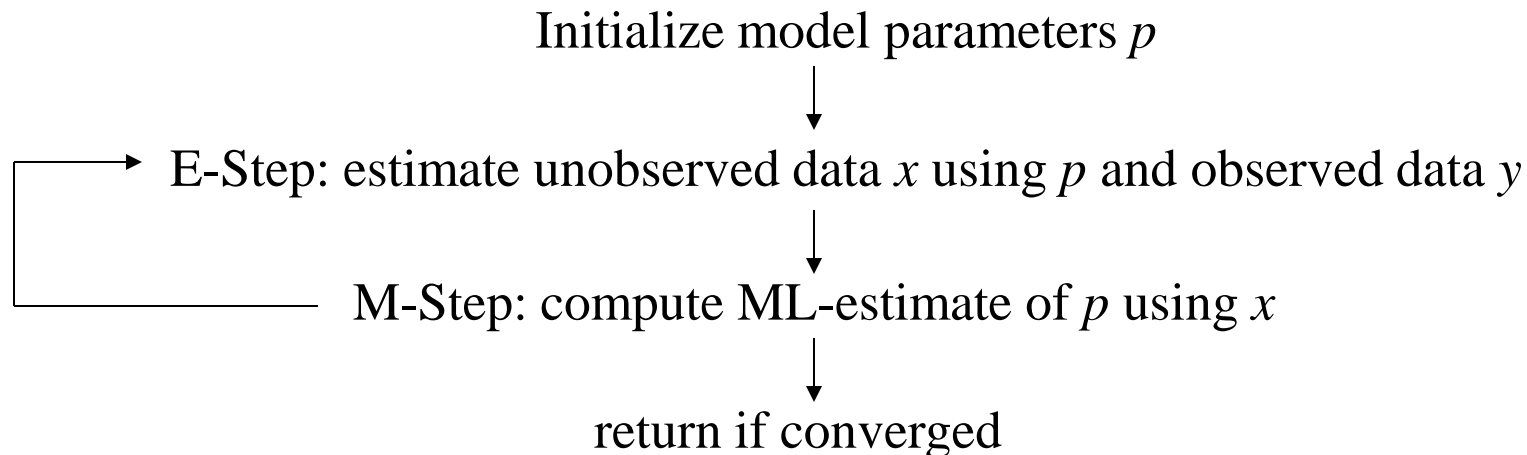
Initialization step: choose an initial setting of the model parameters

Then proceed to EM, which has two steps, executed iteratively:

- E (expectation) step: estimate the unobserved data from the current estimate of the model parameters and the observed data
- M (maximization) step: compute the maximum-likelihood estimate of the model parameters using the estimated unobserved data

Stop when converged

Initialize model parameters p

E-Step: estimate unobserved data x using p and observed data y

M-Step: compute ML-estimate of p using x

return if converged
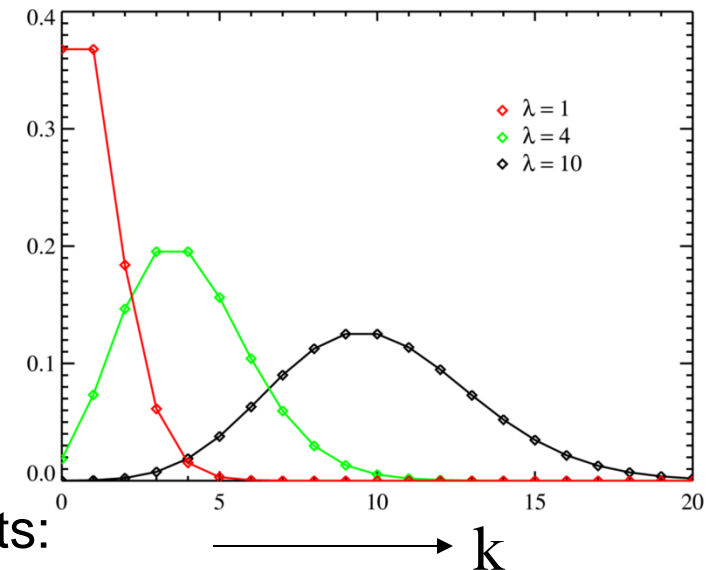
The Poisson Distribution

Also called the *law of rare events*

- it is the binomial distribution of *k* as the number of trials *n* goes to infinity

$$\lim_{n \to \infty} \Pr(X = k) = \lim_{n \to \infty} \binom{n}{k} p^k (1-p)^{n-k}$$

- with *p* = λ / *n*

$$f(k; \lambda) = \frac{e^{-\lambda} \lambda^k}{k!}$$

λ: expected number of events (the mean) in a given time interval

Some examples for Poisson-distributed events:

- the number of phone calls at a call center per minute
- the number of spelling errors a secretary makes while typing a single page
- the number of soldiers killed by horse-kicks each year in each corps in the Prussian cavalry
- the number of positron emissions in a radio nucleotide in PET and SPECT
- the number of annihilation events in PET and SPECT

The observed data, y(d), are the detector readings

The unobserved data, x(b), are the photon emission activities in the pixels (the tissue)

- these give rise to the detector readings
- they follow a Poisson distribution

The model parameters, $\lambda(b)$, are the metabolic properties of the pixels (the tissue)

- these cause the emissions
- they represent the expectations (means) of the resulting Poisson distribution causing the readings at the detectors

Note: to conform with the literature we shall use the terms (object) pixels and (detector) bins here

Since there is a many-to-one mapping, many objects are probable to have produced the observed data

- the object reconstruction (the *image*) having the highest such probability is the *maximum likelihood estimate* of the original object

We shall use the following variables:

- $x[b:1…B]$: the discretized image with B pixels (boxes)
- $y[d:1…D]$: the data (the projections), D is the number of detector bins (tubes)

The reconstruction is based on some mathematical model that relates parameters, observed and unobserved data

The mathematical model is based on the assumption that the emissions occur according to a spatial Poisson point process in the region of interest (field of view) in the source.

- for each box *b* there is a Poisson distributed random variable x(b), with mean $\lambda(b)$, that can be generated independently:

$$f(n|\lambda(b)) = P(n(b) = n|\lambda(b)) = e^{-\lambda(b)} \frac{\lambda(b)^n}{n!}$$

Suppose now, that an emission in the *b*-th box is detected in the *d*-th tube with known probability

p(b,d) = P (event detected in tube d | event emitted in box b)

- the probability matrix *p(b,d)* is assumed to be known from the detector array geometry and other characteristics of the system (attenuation, scatter, etc)
- the probability of an event in box *b* to be detected by the scanner is given by:

$$\sum_{d=1}^{D} p(b,d) \leq 1$$

- thus, photons do go undetected

The variables *y(d)* are independent and Poisson, with expectation λ(d) where:

$$f(y|\lambda(d)) = P(y(d) = y) = e^{-\lambda(d)} \frac{\lambda(d)^y}{y!}$$

Since the *x(i)* are independent Poisson variables, a linear combination of these variables is also Poisson distributed.

- the expectation of the observed data is therefore linked with the model parameters:

$$\lambda(d) = E[y(d)] = \sum_{b=1}^{B} \lambda(b) p(b,d)$$

λ(d) then expresses the expectation (under the Poisson probability model for emission) to observe the given counts in the detector tubes if the true density is λ(b).

For any set of detector data y(d) there are many different ways the photons could have been generated.

- thus there is a many-to-one mapping x(b,d) to y(d)
- this mapping is Poisson distributed with mean:

$$\lambda(b,d) = \lambda(b)p(b,d).$$

The likelihood function is then:

$$l_x(\lambda) = f(x|\lambda) = \prod_{\substack{b=1,\dots,B \\ d=1,\dots,D}} e^{-\lambda(b,d)} \frac{\lambda(b,d)^{x(b,d)}}{x(b,d)!}$$

and the log-likelihood is:

$$L_y(\lambda) = \log l_y(\lambda) =$$

$$\sum_{\substack{b=1,\dots,B \\ d=1,\dots,D}} -\lambda(b)p(b,d) + x(b,d)\log\lambda(b) + x(b,d)\log p(b,d)$$

$$-\log x(b,d)!$$

For the E-step we compute:

$$x^{[k+1]}(b,d) = E[x(b,d)|\mathbf{y},\lambda^{[k]}] = E[x(b,d)|y(d),\lambda^{[k]}]$$

since

- x(b,d) is Poisson with mean $\lambda^{[k]}$(b,d)
- $y(d) = \sum_{b=1}^{B} x(b,d)$ is Poisson with mean $\lambda^{[k]}(d) = \sum_{b=1}^{B} \lambda^{[k]}(b,d)$
- one can write, after some manipulations:

$$x^{[k+1]}(b,\ d) = \frac{y(d)\lambda^{[k]}(b,d)}{\sum_{b'=1}^{B}\lambda^{[k]}(b',d)}, \qquad b = 1,2,...,B, d = 1,2,...,D$$

In the M-step we maximize $\lambda^{[k]}(b)$ using $x^{[k+1]}$:

$$0 = \frac{\partial}{\partial \lambda(b)} \sum_{\substack{b=1,\ldots,B \\ d=1,\ldots,D}} -\lambda(b)p(b,d) + x^{[k+1]}(b,d)\log \lambda(b) +$$

$$x^{[k+1]}(b,d)\log p(b,d) - \log x^{[k+1]}(b,d)!$$

$$\rightarrow \lambda^{[k+1]}(b) = \sum_{d=1}^{D} x^{[k+1]}(b,d)p(b,d)$$

Combining the E and the M steps:

$$\lambda^{[k+1]}(b) = \lambda^{[k]}(b) \sum_{d=1}^{D} \frac{y(d)p(b,d)}{\sum_{b'=1}^{B} \lambda^{[k]}(b')p(b',d)}$$

Algorithm

1. Start with an initial estimate $\lambda^{[0]}(b)$ satisfying $x(0)(b) > 0$, $b=1,2,...,B$

2. If $\lambda^{[k]}(b)$ denotes the estimate of $\lambda(b)$ at the k-th iteration, define a new estimate $\lambda^{[k+1]}(b)$ by:

$$\lambda^{[k+1]}(b) = \lambda^{[k]}(b) \sum_{d=1}^{D} \frac{y(d)p(b,d)}{\sum_{b'=1}^{B} \lambda^{[k]}(b')p(b',d)}$$

3. If the required accuracy for the numerical convergence has been achieved, then stop

There are a number of properties:

- since the correction is multiplicative, all images produced are *non-negative* (compare the algebraic reconstruction methods)
- the algorithm is also *self-normalizing*: the redistribution of the activity in the image cells which occurs after each iteration is done without any net increase or decrease in the total activity

The EM algorithm for emission tomography can provide a physically accurate reconstruction model

- it allows the direct incorporation of many physical factors, which, if not accounted for, can introduce errors in the final reconstruction.
- these factors can be included in the transition matrix *p(b,d):*
  - detector response
  - attenuation correction information
  - scatter modeling
  - positron range and angulation effects
  - time-of-flight information
  - and others

Comparing EM to Filtered Backprojection

- advantages:
    - less streak artifacts and noise
    - projection data can be irregularly spaced
- disadvantages:
    - much slower due to iterative scheme
    - may produce streak artifacts at the edges

Stopping criterion:

$$RN(x) = \sqrt{\sum_{j=1}^{J}\left(v(j) - \tilde{y}(j)\right)^2}$$

- stop when this residual goes to a small number, or grows again
- typically, noise and the edge artifacts lead to growth in the residual after reaching he optimum point

Acceleration:

- most popular is Ordered Subsets (OS) EM
- here the update is done after every subset of projections
- note: this method does not converge to a maximum likelihood solution, except for the case of noise-free data.

Acceleration can provide faster convergence towards high-likelihood estimates, however, it does not guarantee a better image quality

- various modified MLE techniques exist, which incorporate *a priori* information to characterize the source distribution and the data noise.

Bayesian Priors

Concept:

- Recall that EM tries to maximize the probability:

$$P(data|image) \quad (or\ P(y|x))$$

    to observe the measured data given a current estimate of the activity distribution in the source.

- according to Bayes' theorem, one can write:

$$P(image|data) = \frac{P(data|image)\ P(image)}{P(data)}$$

- here we express the conditional probability that the image is true given the set of measured data in terms of the usually computed *P(data|image)*, a normalization constant *P(data)* and the prior probability *P(image)*.

- now, the most probable image, given the set of measured data can be obtained by maximizing the right side of the above equation, called the *a posteriori* probability distribution of the image

Bayesian Priors

Makes use of some „reasonable" *a priori* information in order to prevent the image deterioration that occurs when maximizing *P(data|image)* in unconstrained EM.

- this a priori information is incorporated into the *P(image)* term
- it represents an *a priori* estimate of how the resulting image is expected to be
- these can be smoothness constraints or partial specified topological information
- they are regularization schemes

Examples:

- the use of Gibbs priors penalizes large deviations between the estimates of the image vector for neighboring pixels, except for edges which divide regions
- others: Gaussian priors, Poisson and Gamma priors, Good's roughness prior:
- all model the local continuity of images
- another type of prior: registered anatomical MRI images

For example: the Gibb's Prior

$$U(\mathbf{x}) = exp\left(-\beta \sum_{[n,m] \in \Omega} w_{n,m} V(x(n) - x(m))\right)$$

V is a potential function

Basic idea:

- add the log-energy function, now $u(\mathbf{x}) = log(V(\mathbf{x}))$, to the log-likelihood function $l(\mathbf{x})$ as a penalty function
- it expresses any prior knowledge on the smoothness or other property of the estimate $\mathbf{x}$
- produces a log-posterior function $l(\mathbf{x}) + u(\mathbf{x})$
- the expression $l(\mathbf{x}) + u(\mathbf{x})$ now represents the likelihood of the *a posteriori* probability distribution
- taking the logarithms yields:

$$u(\mathbf{x}) = -\beta \sum_{[n,m] \in \Omega} w_{n,m} V(\mathbf{x})$$

$$logP(\mathbf{x}|\mathbf{y}) = logP(\mathbf{y}|\mathbf{x}) + logP(\mathbf{x}) = l(\mathbf{x}) + u(\mathbf{x})$$

Bayesian Priors

Strategy:

- maximize this expression by taking partials with respect to each *x(i)* and set the result to zero
- solve with standard EM by assuming (specifying) a certain degree of „smoothness" of the resulting images as an a priori information

There are many variants of this concept, for example:

- Maximum A Posteriori Probability EM algorithm (MAP-EM) by Levitan and Herman
    - assumes a multivariate Gaussian a priori probability distribution for the image
- Fast Maximum A Posteriori with Entropy (FMAPE) algorithm by Nunez and Llacer
    - uses an entropy prior with an adjustable „contrast parameter"
- Bayesian Image Processing (BIP) method by Liang and Hart
- Entropy Image Processing method by Liang
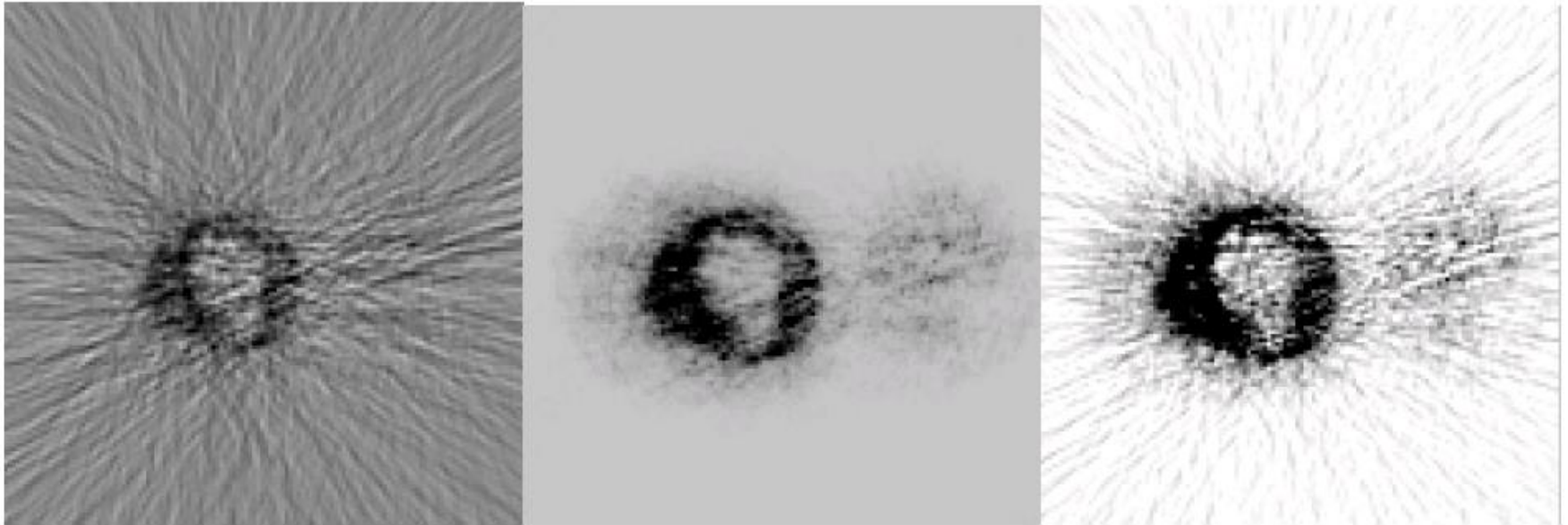   - both make use of entropy analysis on the strength correlations of the image pixels

Fig. 2: One slice (zoom factor: 1.5, 128x128 image matrix) from a study of a patient with a right-leg necrotic sarcoma. *Left*: FBP image (0.5 cutoff frequency, ramp filter). *Center*: OSEM image (4 subsets, 3 iterations). *Right*: Error-image (quantitative subtraction of the FBP image from the OSEM image).
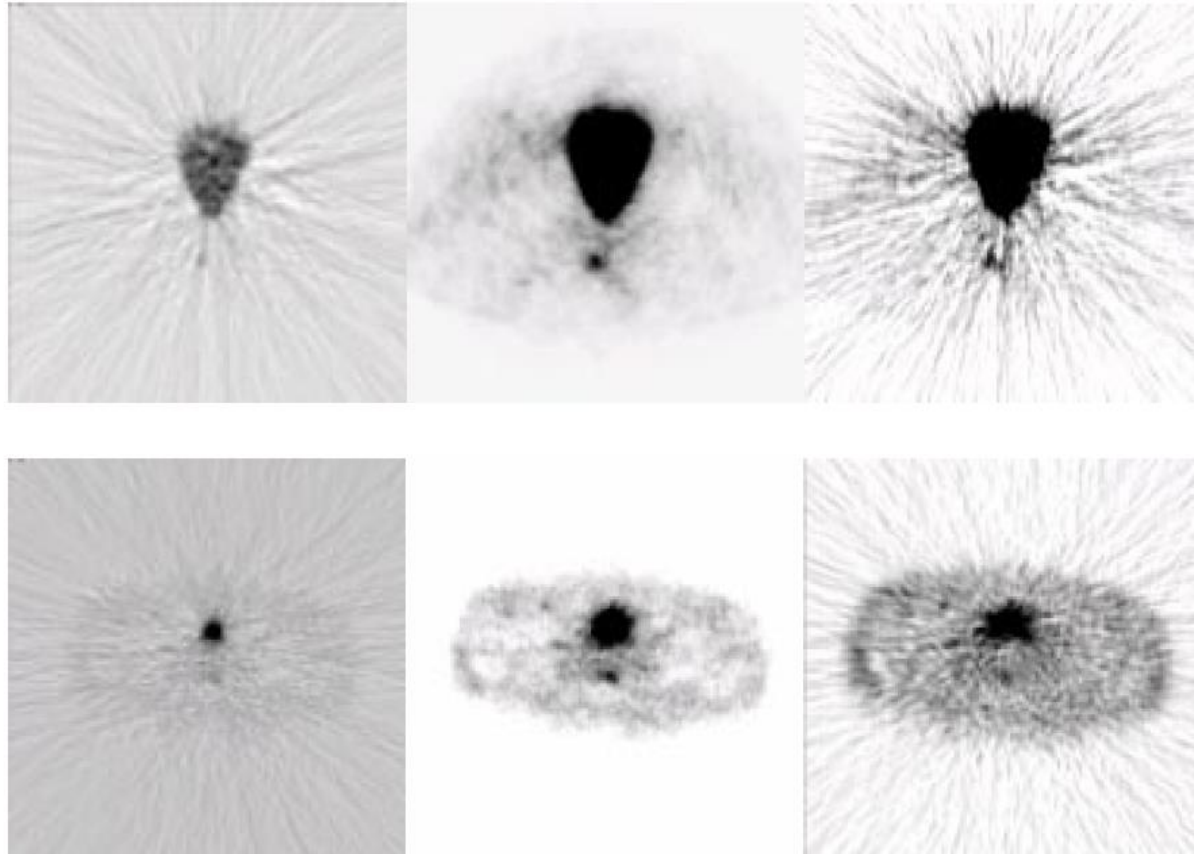
Fig. 3: One slice (zoom factor: 1.5, 128x128 image matrix) from two studies of residual colorectal carcinomas. *Left*: FBP image (0.5 cutoff frequency, ramp filter). *Center*: OSEM image (4 subsets, 3 iterations). *Right*: Error-image (quantitative subtraction of the FBP image from the OSEM image). Apart from the strong streak artifacts visible in the FBP images, the high-activity concentration of the bladder obscures the presence of the lesion.

Recall that the SPECT camera constantly rotates

- this gives rise to a continuous acquisition
- binning the detected events into discrete angles and detectors leads to discretization artifacts (blurring)

List-mode acquisition keeps the events in a list, recording timestamp, acquisition location and detector angle

- ML-EM can still be used to reconstruct, one just has to use the stored data parameters

$$x_j^{(k+1)} = \frac{1}{N} \cdot \sum_{n=1}^{N} \frac{P(l_n|j).x_j^{(k)}}{\sum_{j=1}^{J} P(l_n|j).s_j.x_j^{(k)}}$$

- $P(l_n,j)$ is the probability of list event l_n being emitted at source bin j
- $x^{[k]}_j$ is the expected number of photons emitted from source bin j per unit of time for the k-th iteration,
- s_j is the sensitivity for that bin j
- N is the number of list-mode events
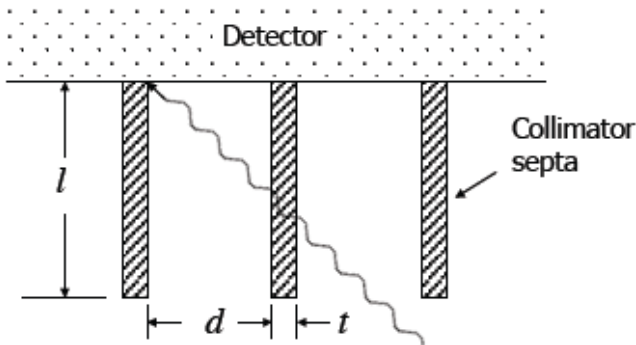- J is the number of source bins.

Collimators typically absorb well over 99.95% of all photons emitted from the patient

Most gamma cameras can acquire data using multiple energy windows.

- allows for simultaneous imaging of different radioisotopes, for example Tc-99m (140 keV) and I-131 (364 keV).
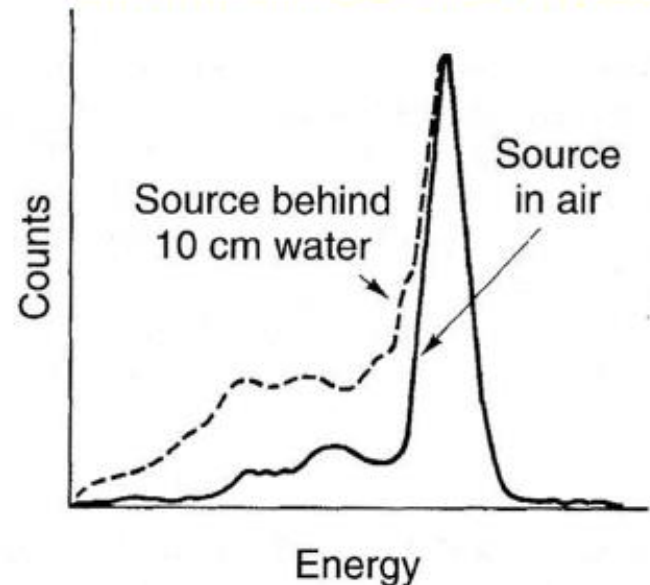
Collimators - Septal Penetration

Detector

Collimator septa

l

$\leftarrow d \rightarrow \leftarrow t$

Minimum septa thickness, t, for <5% septal penetration:

$$t \geq \frac{6d/\mu}{l - (3/\mu)}$$

Gamma Camera Energy Spectra

Counts

Source behind 10 cm water

Source in air

Energy

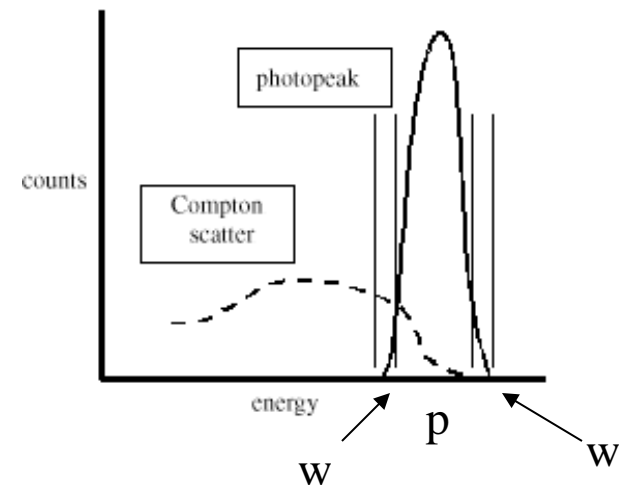Compton scatter is part of the attenuation process (up to 30%)

- scattered photons are diverted from their original path with some loss in energy
- need to reduce the number of the scattered events that are detected
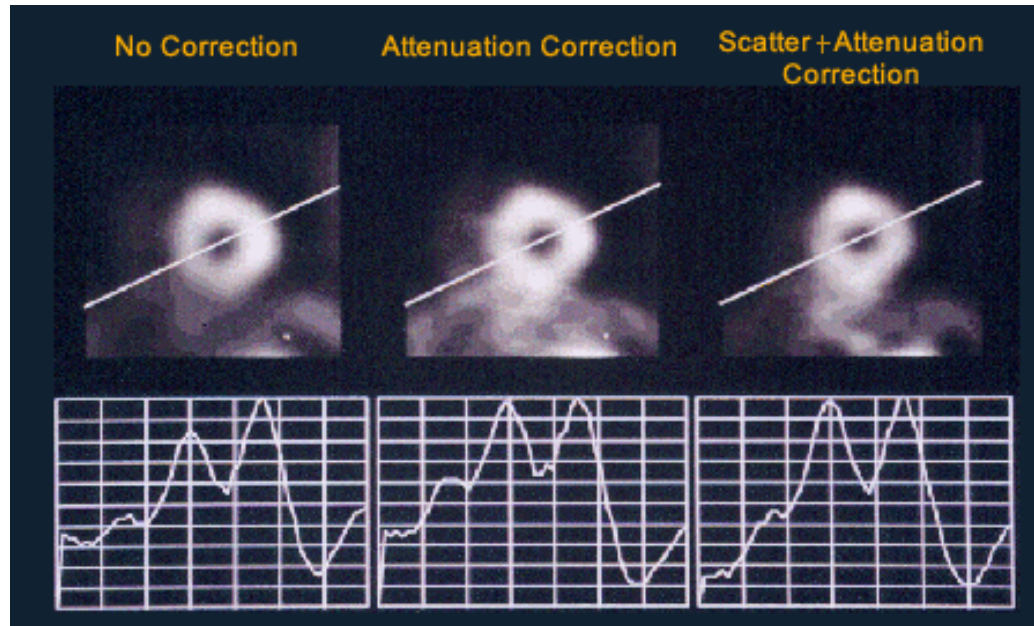
An effective technique is windowing

- triple or dual energy windows
- however, scatter modeling in the reconstruction can be more effective

The triple or dual window technique

- acquires additional images using energy windows either just below or on each side of the photopeak
- it assumes that scatter recorded in these windows will be similar to the scatter in the photopeak
- the scatter in the photopeak can be estimated by subtraction
- scatter=(count in narrow windows / 2) · p/w

Windowing for Scatter Correction

Without attenuation correction (left) the posterior wall has reduced counts. After attenuation correction the posterior wall is improved but myocardial to ventricle contrast is reduced (middle) and there is greater influence from abdominal activity. Scatter correction restores contrast (right) and reduces scatter from extra-cardiac structures
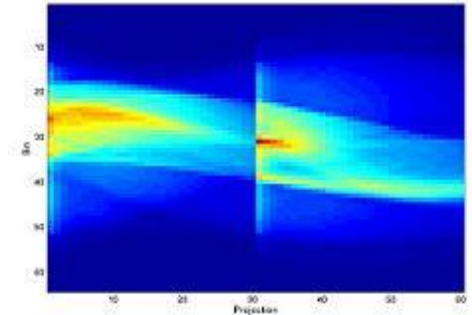
Conventional approaches assume constant emissions

- but temporal changes in emissions due to wash-out from the organ (cardiac, brain, kidney perfusion) can provide additional information (for example, tumor analysis)
- can reveal additional information about the underlying physiological processes and enhance the diagnostic possibilities of SPECT
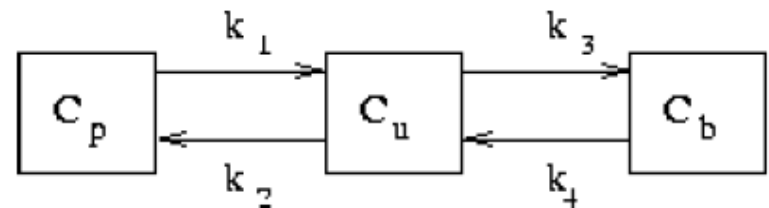
Reconstruction, using EM

- reconstruct as separate 3D volumes
- reconstruct as a combined 4D volume
- reconstruct the time courses via kinetic model

The kinetic (departmental) model: look at time-courses

- $k1$ and $k2$:uptake/washout from blood to tissue and vice versa
- $k3$ and $k4$: conversion from ordinary to a phosphorylated form in the tissue, and vice versa
- solution via differential equations
- may also model as B-splines

References

EM:

- G. Kontaxakis, L. Strauss, "Maximum Likelihood Algorithms for Image Reconstruction in Positron Emission Tomography"
- T. Moon, "The Expectation-Maximization Algorithm"

List mode:

- H. Barrett, T. White, and L. Parra, "List-mode likelihood," Journal of the Optical Society of America A-Optics Image Science and Vision, vol. 14, pp. 2914-2923, November 1997.
- Bouwens, Van de Walle, Gifford, Lemahieu and Dierckx, "3D List-Mode Reconstruction for SPECT," 6th Intern'l Meeting on Fully Three-Dimensional Image Reconstruction in Radiology & Nuclear Medicine

dSPECT and dPET:

- M. E. Phelps, "PET: Molecular Imaging and Its Biological Applications," New York: Springer, 2004.
- M. N. Wernick, J. N. Aarsvold (Eds.), "Emission Tomography: The Fundamentals of PET and SPECT," Elsevier, 2004
- I. Samil Yetik, Jinyi Qi, "Direct Estimation of Kinetic Parameters From the Sinogram with an Unknown Blood Function," *2006 IEEE International Symposium on Biomedical Imaging.*