Bayesian electron spectrum reconstruction from dose-depth profiles

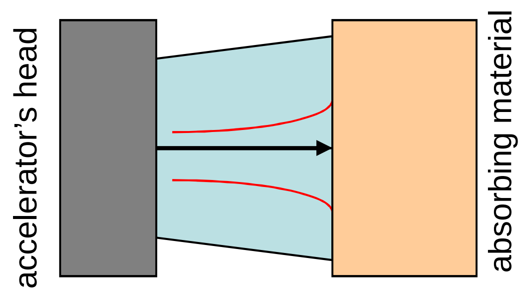
V. I. Dimitrov, IAC-ISU

Bayesian

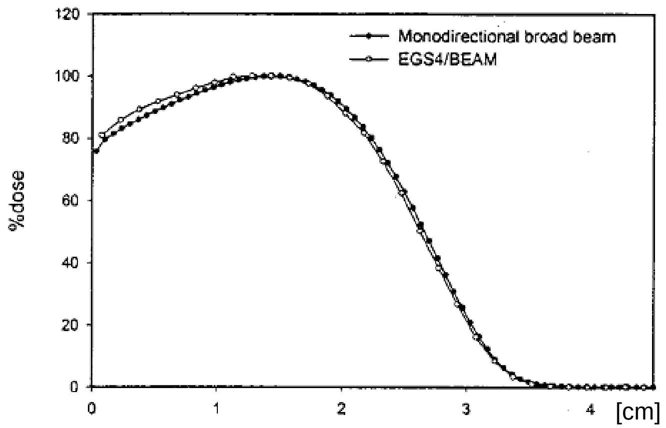
Function: *adjective*

being, relating to, or involving statistical methods that assign probabilities or distributions to events (as rain tomorrow) or parameters (as a population mean) based on experience or best guesses before experimentation and data collection, and that apply Bayes' theorem to revise the probabilities and distributions after obtaining experimental data.

The Physical Problem



a) space-charge repulsion
b) stray magnetic fields



Dose-depth profile (6MeV electrons in water)

$$D(x) = \int_0^{\varepsilon_{\max}} d\varepsilon\, W(x,\varepsilon) S(\varepsilon)$$

$$S(\varepsilon) = ?$$

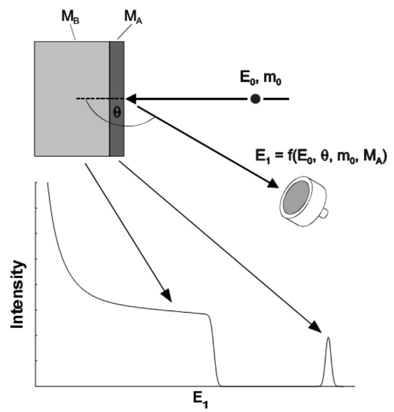
Fredholm integral equation (1st kind): an ill-posed mathematical problem!

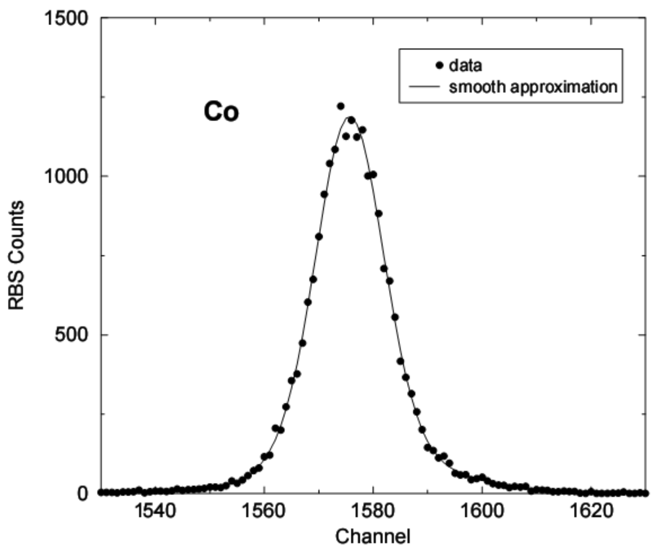
Additional complications due to beam's size and emittance/divergence exist as well.

The Status Quo

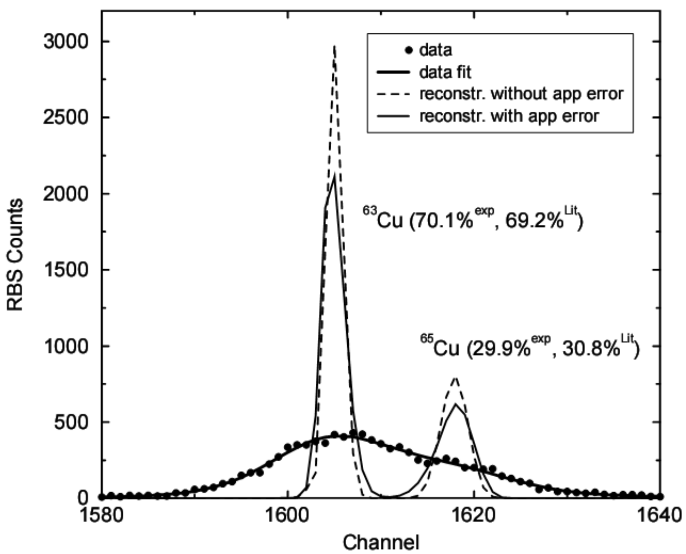
i) Bayesian methods in similar problems:

Rutherford Back-Scattering (H or He ions ~1MeV):
(e.g. R. Fisher et al., Phys. Rev. **E55** p.6667 (1997))





Apparatus Function Determination
(mono-isotopic Co has a single narrow line)

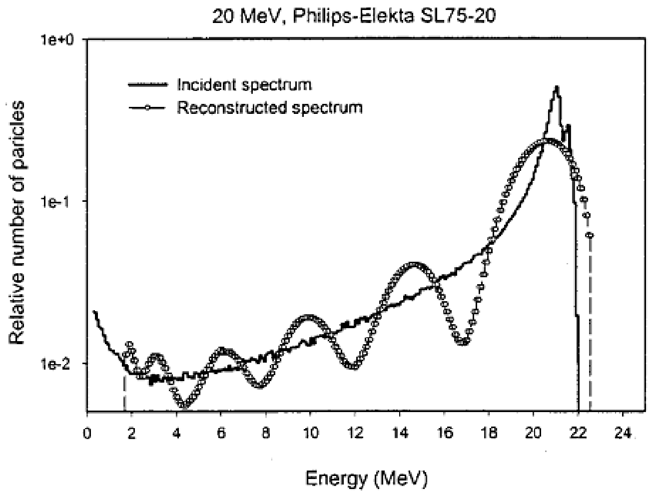


Spectrum Deconvolution
(Cu layer on Si substrate)

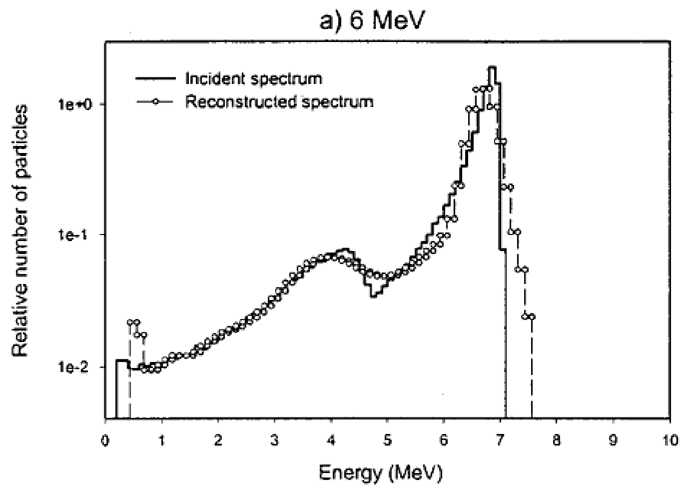
The Status Quo

ii) Ad hoc methods in electron spectrum reconstruction problems

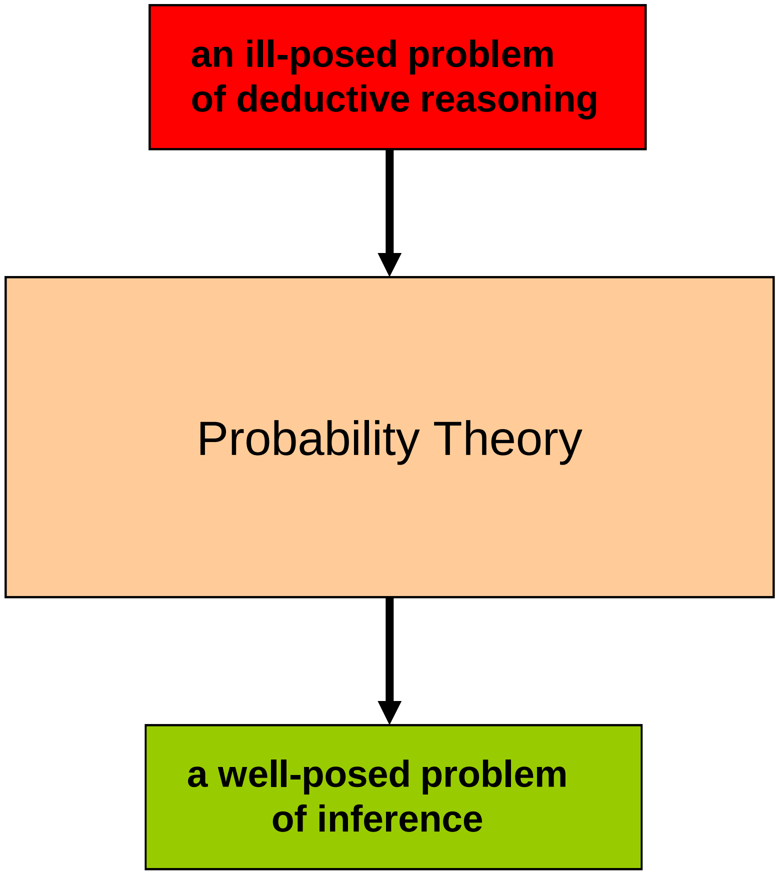
e.g. A. Chvetsov et al. Med. Phys. **29** p. 578 (2002)



No regularization, brute force



Tichonov regularization and spectrum splitting into smooth and peaked parts



$$\int_{0}^{\varepsilon_{\max}} d\varepsilon\, W(x,\varepsilon) F(\varepsilon) = 0$$

$$S(\varepsilon) \rightarrow S(\varepsilon) + F(\varepsilon)$$

An example of inference:

a) If A is true, than B is true as well (prior information);
b) A is false (data from experiment);
c) B is *less plausible* (than before the experiment)

Basic Probability Theory I

Notations:

$P(A)$: Probability of *A* being true

$P(\bar{A})$: Probability of *A* being false (.not.A being true)

$P(A \mid B)$: Probability of *A* being true provided that *B* is true

$P(A \mid BC)$: Probability of *A* being true provided that both *B* and *C* are true

Basic Probability Theory II

"common sense reduced to calculation" (Laplace)

Range: $$P(A) \in [0,1]$$

Sum rule: $$P(A) + P(\overline{A}) = 1$$

Product rule: $$P(AB\,|\,C) = P(A\,|\,BC)P(B\,|\,C)$$

Bayes' theorem (in its simplest form) is an immediate consequence of the above product rule and the commutativity of propositions:

$$P(AB) = P(BA)$$

<u>Bayes' Theorem</u>

H – hypothesis
D – data
I – prior information

$$P(HD \mid I) = P(H \mid DI)P(D \mid I) = P(D \mid HI)P(H \mid I) = P(DH \mid I)$$

sampling distribution / likelihood

posterior probability

prior probability (prior)

$$P(H \mid DI) = \frac{P(D \mid HI)P(H \mid I)}{P(D \mid I)}$$

Bayes' Theorem as a Learning Prescription

result of the experiment

updated knowledge

our state of knowledge before the experiment

$$P(H \mid DI) = \frac{P(D \mid HI)P(H \mid I)}{P(D \mid I)}$$

normalization

In 1946, R.T. Cox proved that any *consistent* scheme of logical inference must be *equivalent* to probability theory as described

Least-Informative Priors

a) Discrete probabilities: *Principle of Insufficient Reason*

b) Continuous probabilities: *Symmetries / Invariance requirements*

Example: $N(t) = N_0 \exp(-\lambda t) \qquad P(\lambda)d\lambda \sim d\lambda$

$$N(t) = N_0 \exp(-t/\tau) \qquad P(\tau)d\tau \sim d\tau \sim d\lambda/\lambda^2$$

$$P(\lambda)d\lambda \overset{!}{=} P(a\lambda')d(a\lambda') \Rightarrow P(\lambda)d\lambda \sim \frac{d\lambda}{\lambda} \sim \frac{d\tau}{\tau}$$ (Jeffrey's prior)

MAXENT principle (Jaynes'1957, but originally Gibbs'1902):

$$S = -\int dx P(x) \ln \frac{P(x)}{P_0(x)} \rightarrow \max$$

(Kullback-Leibler ('51) relative entropy)

The uniqueness of entropy

One seeks a “ranking” scheme *R(p)* for probability distributions *p(y)*:

i) Locality: $$R(p) = \int dy\ f(p(y))$$

ii) Invariance: $$R(p) = \int dy\ \mathrm{p}(y)\ f(\frac{p(y)}{m(y)})$$

iii) Consistency for independent systems:

$$R(p_1 p_2) = R(p_1) + R(p_2)$$

$$\int dy_1 \mathrm{d}y_2\ \mathrm{p}_1(y_1)\mathrm{p}_2(y_2)\ f(\frac{\mathrm{p}_1(y_1)\mathrm{p}_2(y_2)}{m_1(y_1)m_2(y_2)}) = \int dy_1\ \mathrm{p}_1(y_1)\ f(\frac{\mathrm{p}_1(y_1)}{m_1(y_1)}) + \int \mathrm{d}y_2\ \mathrm{p}_2(y_2)\ f(\frac{\mathrm{p}_2(y_2)}{m_2(y_2)})$$

for $\int dy\, p(y) = 1 \qquad \Rightarrow f(p) = \ln(p)$

thus $$R(p) = \int dy\ \mathrm{p}(y) \ln(\frac{p(y)}{m(y)})$$

The Likelihood

Gaussian likelihood function (just one of many possible):

$$P(D|HI) = \frac{\exp(-\frac{\chi^2}{2})}{\prod_{i=1}^{N} \sqrt{2\pi}\sigma_i}$$

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{D_i - F_i(H)}{\sigma_i} \right)^2$$

Here, σ_i is the error of the measurement of the i-th data point D_i and $F_i(S)$ is the calculated value of D_i assuming H.

Application to the problem at hand

i) Discretization:

$$S(\varepsilon)d\varepsilon \rightarrow s_i \Delta\varepsilon_i$$
$$s_i = S(\varepsilon_i)$$

$$S(\varepsilon) = \sum_{i=1}^{N} s_i F_i(\varepsilon)$$
$$\int d\varepsilon F_i(\varepsilon) F_j(\varepsilon) = \delta_{ij}$$
$$\sum_{i=1}^{\infty} F_i(\varepsilon) F_i(\varepsilon') = \delta(\varepsilon - \varepsilon')$$

$$D(x) = \int_0^{\varepsilon_{\max}} d\varepsilon\, W(x,\varepsilon) S(\varepsilon)$$
$$S(\varepsilon) = ?$$

Fredholm equation

discretization

Too fine a mesh (too big a basis) carries the danger of overfitting (ringing).

$$d_i = \int dx D(x) G_i(x)$$
$$W_{ij} = \int\int dx d\varepsilon W(x,\varepsilon) G_i(x) F_j(\varepsilon)$$
$$s_i = \int d\varepsilon S(\varepsilon) F_i(\varepsilon)$$

$$d_i = \sum_j W_{ij} s_j$$
$$s_i = ?$$

Matrix inversion

Application to the problem at hand

ii) The "hypothesis":

Every set of s_i gives us a spectrum $$S(\varepsilon) = \sum_{i=1}^{N} s_i F_i(\varepsilon)$$

N degrees of freedom

$$H : s_1 \in \{s^{(1)}, ds\} \cap s_2 \in \{s^{(2)}, ds\} \cap \cdots s_N \in \{s^{(N)}, ds\}$$
$$P(H \mid X) = P(s_1 \in \{s^{(1)}, ds\} \cap s_2 \in \{s^{(2)}, ds\} \cap \cdots s_N \in \{s^{(N)}, ds\} \mid X)$$

iii) The "prior":

The choice of the prior is where the art in this science is!

For the sake of example, the joint Jeffrey's prior would be

$$P(H \mid I) \sim \left[\prod_{i=1}^{N} | s^{(i)} | \right]^{-1}$$

Application to the problem at hand

ii) The Bayes theorem and the inferred spectrum:

$$P(\{s\}\,|\,DI) \sim \frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{d_i - D_i(s)}{\sigma_i}\right)^2\right)}{\prod_{k=1}^{M}\sigma_k \prod_{i=1}^{N} |\, s^{(i)} \,|}$$

The most probable spectrum:

$$P(\{s\}\,|\,DI) \to \max|_{\{s\}}$$

The inferred spectrum:

$$\bar{s}_i = \int d^N s \; s_i \frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{d_i - D_i(s)}{\sigma_i}\right)^2\right)}{\prod_{k=1}^{M}\sigma_k \prod_{i=1}^{N} |\, s^{(i)} \,|}$$

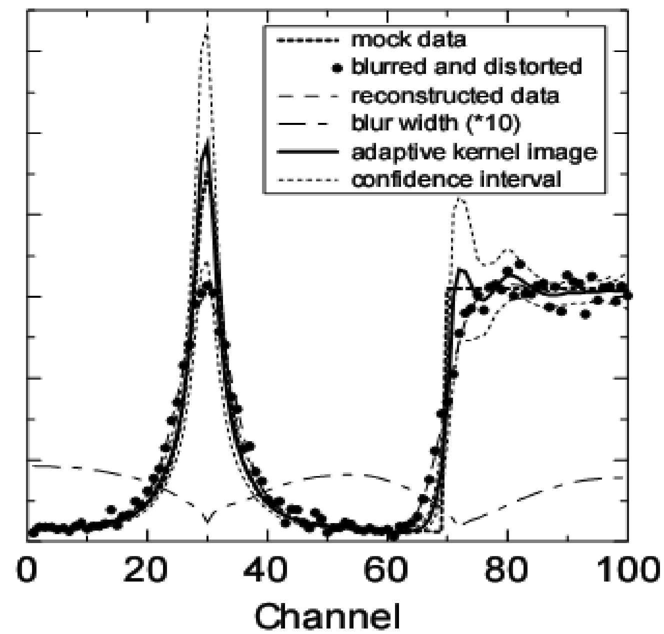
The multiple integral is best taken by Monte Carlo methods. If the results are to be trusted, the most probable and the average spectra should be fairly similar.

Application to the problem at hand

iv) The Bayesian approach produces estimates of the quality of the result as well!

$$(\Delta \bar{s}_i)^2 = \int d^N s \, (s_i - \bar{s}_i)^2 \, \frac{\exp\left(-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{d_i - D_i(s)}{\sigma_i}\right)^2\right)}{\prod_{k=1}^{M}\sigma_k \prod_{i=1}^{N} | s^{(i)} |}$$

R.Fisher et al. 1996

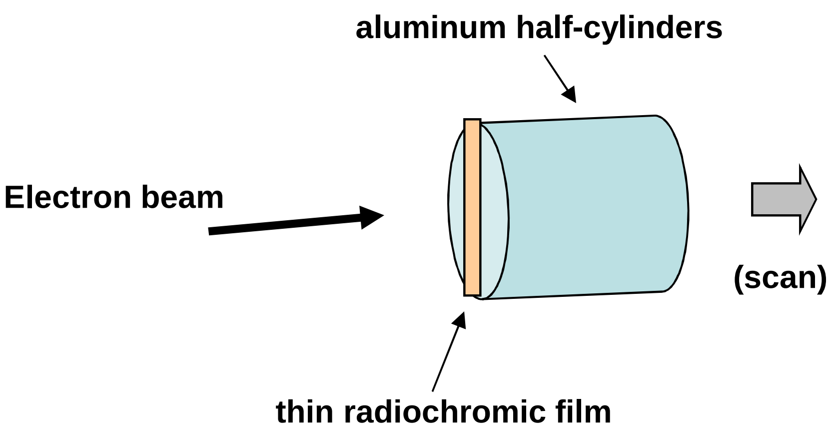


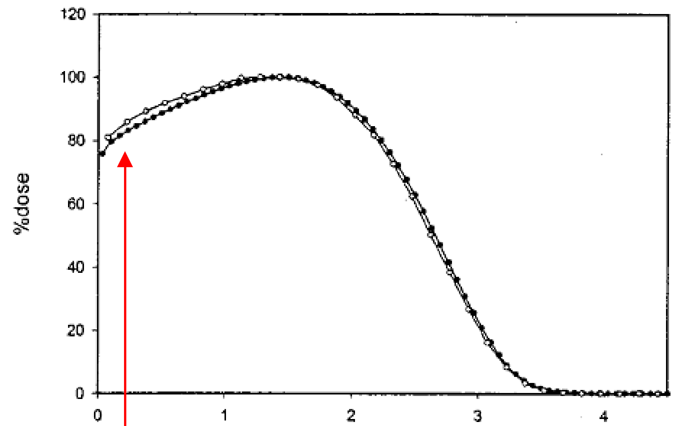
It can be shown that whenever

$$\frac{| \Delta \bar{s}_i |}{| s_i |} << 1$$

the most-probable and the average spectra are “close”.

Experimental

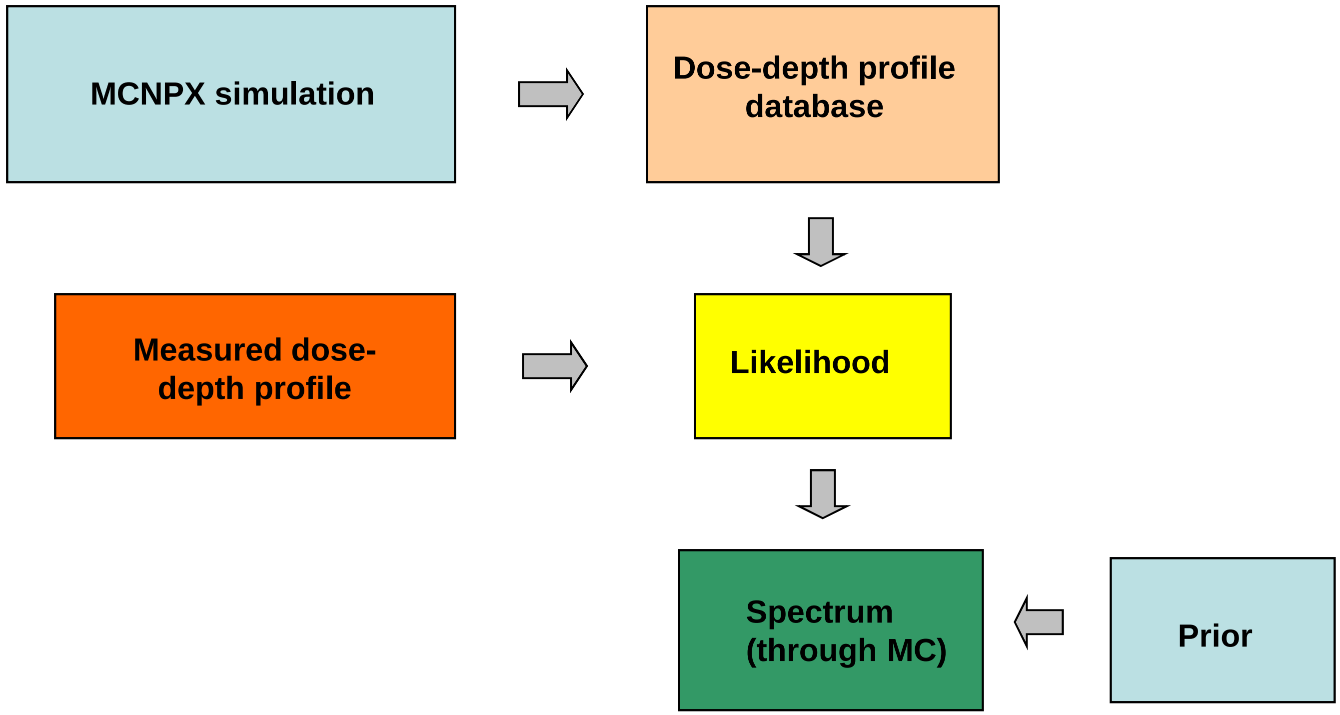




Dose-depth profile

The initial part of the profile (for small depths) is overly sensitive to electron beam divergence and therefore should be discarded.

Theoretical



Work to be done

i) modification of the MCNPX code to track electrons in external magnetic and electric fields;
ii) calculation of dose-depth profile database(s) for a range of electron energies and beam divergences;
ii) Bayesian deconvolution algorithm development and software implementation;
iv) code(s) validation and testing.

Interested students please contact V. Dimitrov at **282-5472**