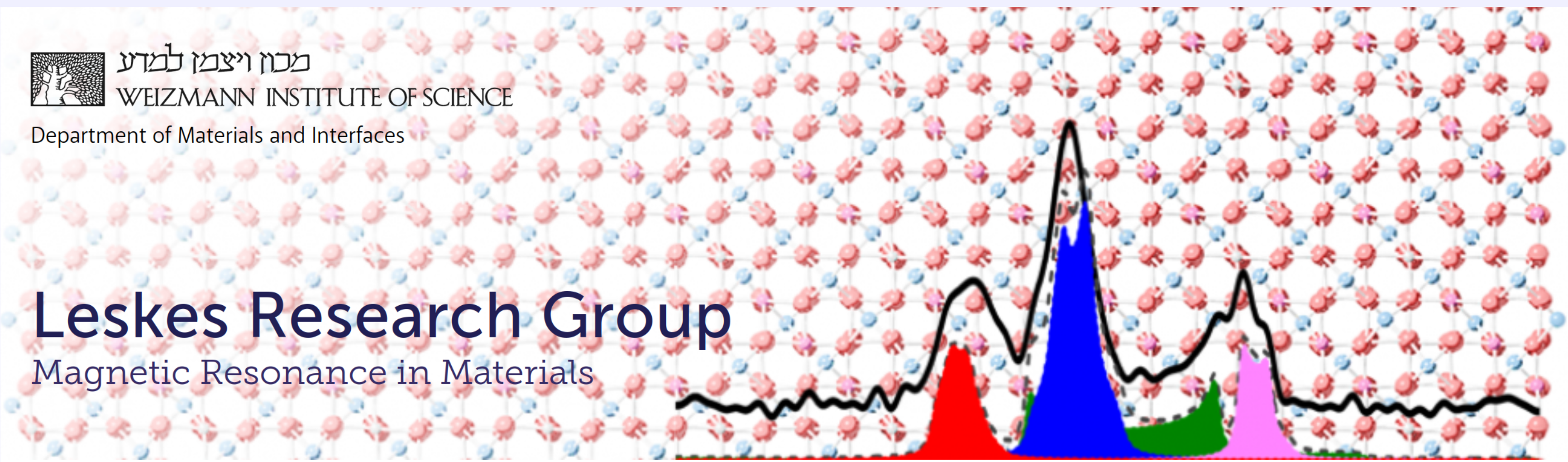
Rotation (Midpoint-esque) Presentation

Yuval Bernard

Leskes Group

April 2023

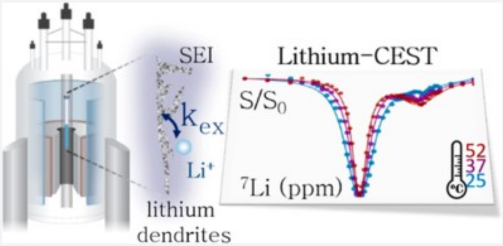


Direct Detection of Lithium Exchange across the Solid Electrolyte Interphase by ^{7}Li Chemical Exchange Saturation Transfer

David Columbus, Vaishali Arunachalam, Felix Glang, Liat Avram, Shira Haber, Arava Zohar, Moritz Zaiss, and Michal Leskes*

Cite This: *J. Am. Chem. Soc.* 2022, 144, 9836–9844

Read Online



ABSTRACT: Lithium metal anodes offer a huge leap in the energy density of batteries, yet their implementation is limited by solid electrolyte interphase (SEI) formation and dendrite deposition. A key challenge in developing electrolytes leading to the SEI with beneficial properties is the lack of experimental approaches for directly probing the ionic permeability of the SEI. Here, we introduce lithium chemical exchange saturation transfer (Li-CEST) as an efficient nuclear magnetic resonance (NMR) approach for detecting the otherwise invisible process of Li exchange across the metal–SEI interface. In Li-CEST, the properties of the undetectable SEI are encoded in the NMR signal of the metal resonance through their exchange process. We benefit from the high surface area of lithium dendrites and are able, for the first time, to detect exchange across solid phases through CEST. Analytical Bloch-McConnell models allow us to compare the SEI permeability formed in different electrolytes, making the presented Li-CEST approach a powerful tool for designing electrolytes for metal-based batteries.

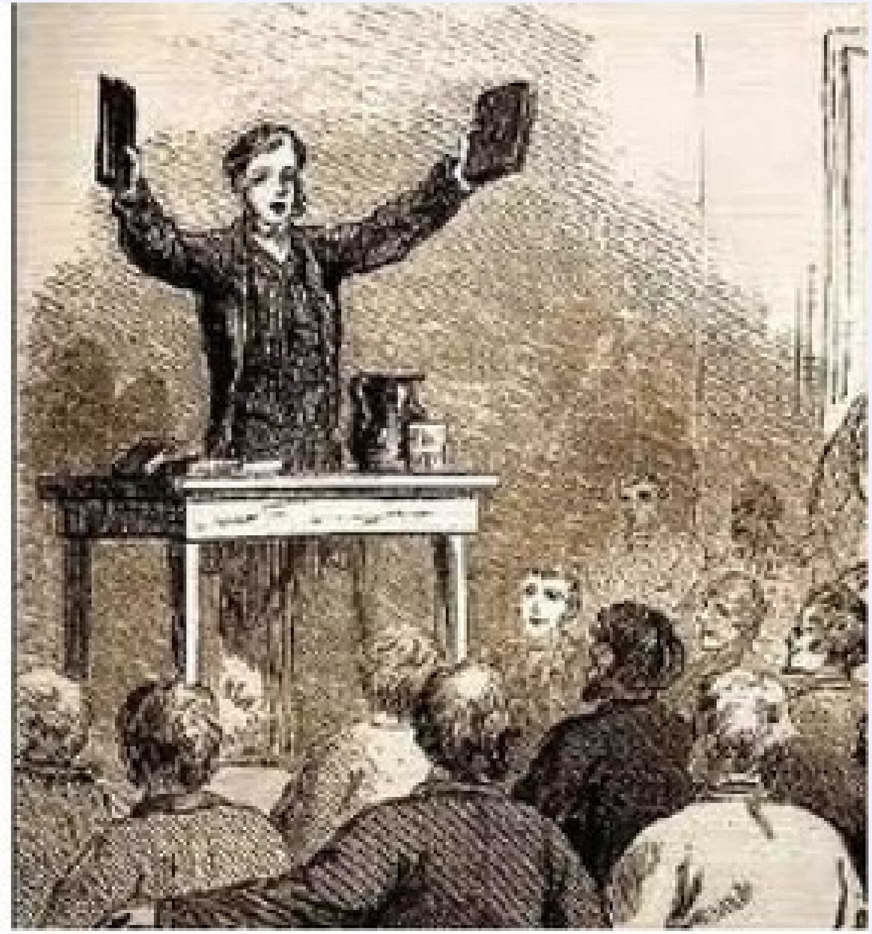
Outline — Part 1

- What is CEST?
- The typical continuous wave (CW) CEST experiment
- Bloch-McConnell equations
- Numerical vs. Analytical solution
- MATLAB simulation of exchange between multiple spin populations.

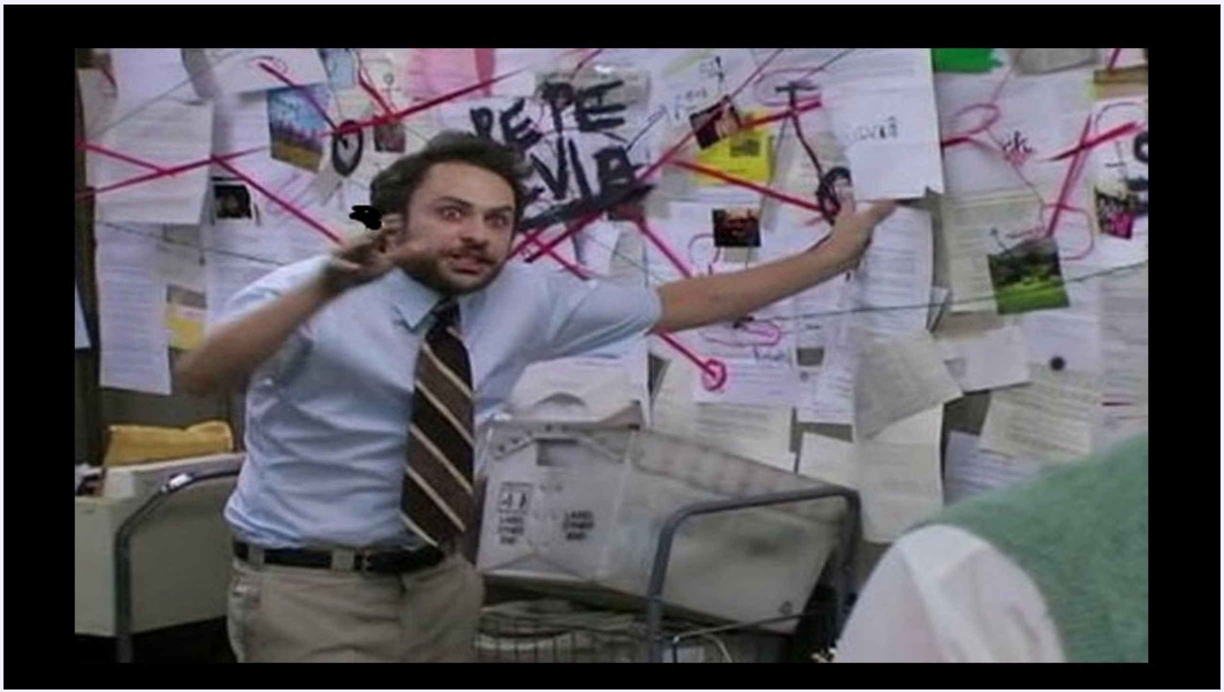
Outline — Part 2

- CEST fitting
- Frequentist or Bayesian approach?
- Monte Carlo simulation and Metropolis-Hastings algorithm
- Diagnostics, analysis & problems with the algorithm
- Hamiltonian Monte Carlo & Stan to the rescue
- Results so far (please help)

A Word of Caution



A Word of Caution



- Method mainly used in MRI, where generation of images with minimal concentration of contrast agents is critical. (As to avoid perturbation of the physiological environment to minimize toxicity.)

Introduction to CEST

- Method mainly used in MRI, where generation of images with minimal concentration of contrast agents is critical. (As to avoid perturbation of the physiological environment to minimize toxicity.)
- Suppose you wish to probe a dilute macromolecule in a tissue that has a weakly-bound hydrogen that can exchange with bulk water.

- Method mainly used in MRI, where generation of images with minimal concentration of contrast agents is critical. (As to avoid perturbation of the physiological environment to minimize toxicity.)
- Suppose you wish to probe a dilute macromolecule in a tissue that has a weakly-bound hydrogen that can exchange with bulk water.
- NMR being NMR™, the hydrogen is barely detectable in the $^1\mathrm{H}$ spectrum.

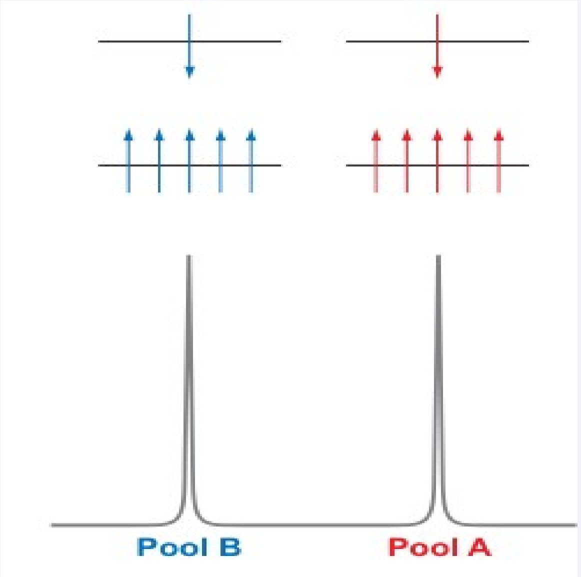
Introduction to CEST

- Method mainly used in MRI, where generation of images with minimal concentration of contrast agents is critical. (As to avoid perturbation of the physiological environment to minimize toxicity.)
- Suppose you wish to probe a dilute macromolecule in a tissue that has a weakly-bound hydrogen that can exchange with bulk water.
- NMR being NMR™, the hydrogen is barely detectable in the $^{1}\mathrm{H}$ spectrum.
- However, irradiating the dilute population to saturation can have a detectable effect on the NMR signal of the bulk water hydrogen population!

The CEST Mechanism

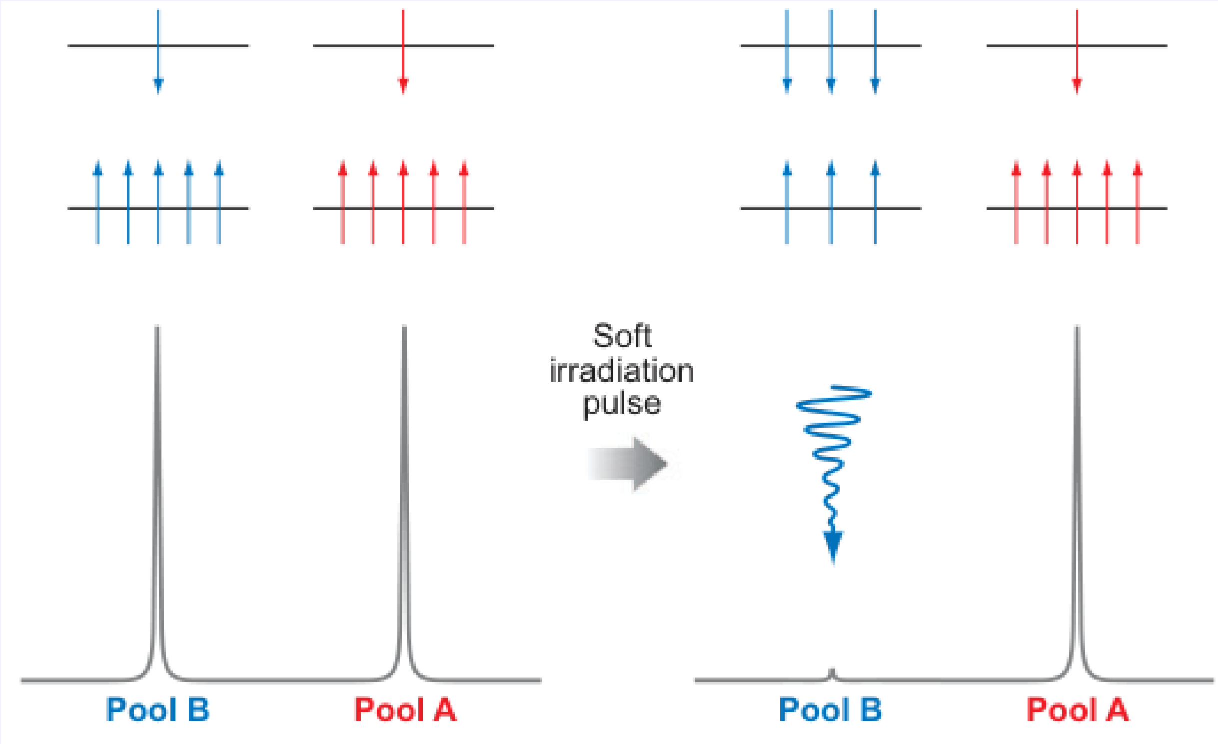
Consider the following NMR experiment:

- Two nuclei ensembles A and B exposed to different chemical environments are placed in an external magnetic field B_0 (e.g. 7.4T)
- Using a low intensity magnetic field, a $\frac{\pi}{2}$ pulse is applied to both ensembles to initiate free induction decay.



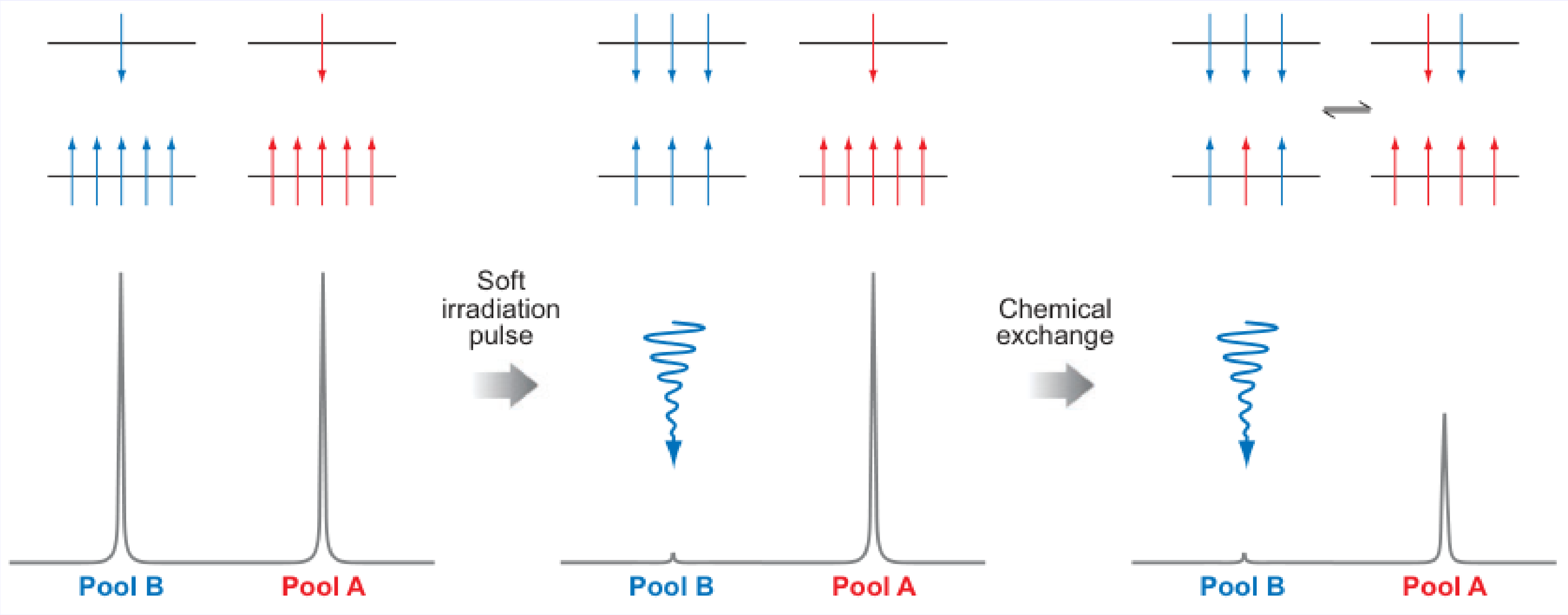
The CEST Mechanism (cont.)

- Using a low-intensity magnetic field B_1, Pool B is irradiated at its resonance frequency until the net magnetization reaches zero (saturation) .



The CEST Mechanism (cont.)

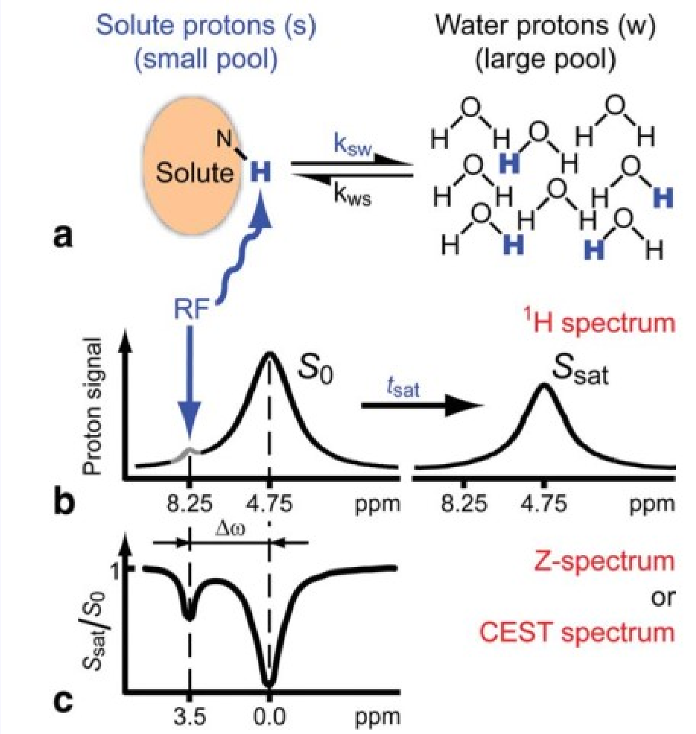
- Nuclei in pool B and pool A are in exchange throughout the process:
- Spins from pool B that are aligned against B_0 are transfered to pool A.
- Spins from pool A that are aligned with B_0 are transfered to pool B.

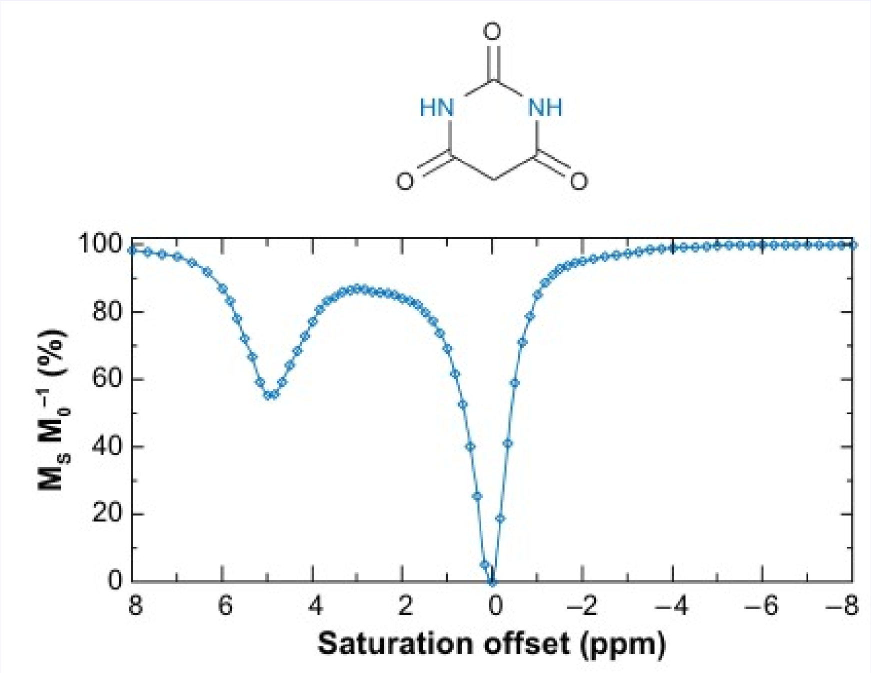


The CEST Experiment

- How would the CEST effect be influenced by changing the saturation irradiation frequency?
  - Straying away from the resonance frequency of pool B results in less effective saturation of its population, which in turn makes the chemical exchange less effective in attenuating pool A NMR signal.
  - *However*, Approaching the resonance frequency of pool A results in **direct saturation** and thus much more effective reduction of the signal.
- A typical CEST experiment involves swiping the saturation frequency and measuring the attenuation of the signal of pool A, to get a **Z-spectrum**.

A Typical 2-pool Z-spectrum





Information Gained From a CEST Experiment

Generally, The CEST effect is influenced by the following parameters:

- The relaxation times of each pool, $T_{1i}, T_{2i}, \quad i \in A, B$
- The back and forward exchange rate of nuclei (spins), $k_{A \rightarrow B}, k_{B \rightarrow A}$
- The fraction of pool B in solution (relative to pool A), $f_B = M_B^0 / M_A^0$
- The saturation field intensity and frequency, $B_1, \Delta\omega$
- The offset between the resonance frequency of the pools, $\Delta\omega_B$
- The saturation irradiation duration t_{sat}

Information Gained From a CEST Experiment

Generally, The CEST effect is influenced by the following parameters:

- The relaxation times of each pool, $T_{1i}, T_{2i}, \quad i \in A, B$
- The back and forward exchange rate of nuclei (spins), $k_{A\rightarrow B}, k_{B\rightarrow A}$
- The fraction of pool B in solution (relative to pool A), $f_B = M_B^0 / M_A^0$
- The saturation field intensity and frequency, $B_1, \Delta\omega$
- The offset between the resonance frequency of the pools, $\Delta\omega_B$
- The saturation irradiation duration t_{sat}

How do these parameters govern the CEST effect?

Bloch-McConnell equations!

Bloch-McConnell Equations

In 1946, Felix Bloch proposed a set of 3 differential equations that describe the time evolution of nuclear magnetization under the influence of a constant, z-aligned field B_0 and a field B_1 aligned on the x-axis and alternating at frequency ω.

Define: $\vec{B}_1(t) = \boldsymbol{i}\,(2B_1 \cos \omega t), \vec{B}_0 = \boldsymbol{k}\,B_0$.

If $\vec{M} = \boldsymbol{i}\,M_x + \boldsymbol{j}\,M_y + \boldsymbol{k}\,M_z$, then:

$$\frac{\mathrm{d}\vec{M}}{\mathrm{d}t} = \vec{M} \times \gamma\,[\vec{B}_0 + \vec{B}_1(t)] + \left(-\frac{M_x}{T_2}\,\boldsymbol{i} - \frac{M_y}{T_2}\,\boldsymbol{j} + \frac{M_z - M_0}{T_1}\,\boldsymbol{k} \right)$$

The second summand in the equation refers to **relaxation terms**.

Bloch-McConnell Equations (cont.)

The time dependence of $\vec{B}_1(t)$ can be eliminated by moving to a *rotating frame* rotating at frequency ω. The equation is then transformed:

$$\frac{\mathrm{d}\vec{M}}{\mathrm{d}t} = \vec{M} \times [\gamma B_1 \boldsymbol{i} + (\gamma B_0 - \omega)\, \boldsymbol{k}] + \left(-\frac{M_x}{T_2} \boldsymbol{i} - \frac{M_y}{T_2} \boldsymbol{j} + \frac{M_z - M_0}{T_1} \boldsymbol{k} \right)$$

Define $\Delta\omega = \gamma B_0 - \omega$ and $\gamma B_1 = \omega_1$.

The Bloch equations in the rotating frame are:

$$\begin{aligned}
\frac{\mathrm{d}M_x}{\mathrm{d}t} &= -\frac{1}{T_2} M_x + \Delta\omega\, M_y \\
\frac{\mathrm{d}M_y}{\mathrm{d}t} &= -\Delta\omega\, M_x - \frac{1}{T_2} M_y + \omega_1\, M_z \\
\frac{\mathrm{d}M_z}{\mathrm{d}t} &= -\omega_1\, M_y - \frac{1}{T_1} M_z + \frac{M_0}{T_1}
\end{aligned}$$

Bloch-McConnell Equations (cont.)

In 1957, Harden M. McConnell modified the Bloch equations to permit chemical exchange and RF saturation.

Given two pools, A and B, that undergo nuclei exchange under constant RF irradiation at frequency ω, the time-evolution of their spacial magnetization follows **Bloch-McConnell equations**:

- Each pool's x, y and z-magnetization develop "independently" according to Bloch equations.
- Each pool's x, y and z-magnetization is transfered between the pools according to the exchange rate.

Bloch-McConnell Equations (cont.)

$$
\begin{aligned}
\frac{\mathrm{d}M_x^A}{\mathrm{d}t} &= -\frac{M_x^A}{T_{2A}} + \Delta\omega_A\, M_y^A - k_{\mathrm{AB}}\, M_x^A + k_{\mathrm{BA}}\, M_x^B \\
\frac{\mathrm{d}M_y^A}{\mathrm{d}t} &= -\frac{M_y^A}{T_{2A}} - \Delta\omega_A\, M_x^A + \omega_1\, M_z^A - k_{\mathrm{AB}}\, M_y^A + k_{\mathrm{BA}}\, M_y^B \\
\frac{\mathrm{d}M_z^A}{\mathrm{d}t} &= -\frac{M_z^A - M_0^A}{T_{1A}} - \omega_1\, M_y^A - k_{\mathrm{AB}}\, M_z^A + k_{\mathrm{BA}}\, M_z^B \\
\frac{\mathrm{d}M_x^B}{\mathrm{d}t} &= -\frac{M_x^B}{T_{2B}} + \Delta\omega_B\, M_y^B - k_{\mathrm{BA}}\, M_x^B + k_{\mathrm{AB}}\, M_x^A \\
\frac{\mathrm{d}M_y^B}{\mathrm{d}t} &= -\frac{M_y^B}{T_{2B}} - \Delta\omega_B\, M_x^B + \omega_1\, M_z^B - k_{\mathrm{BA}}\, M_y^B + k_{\mathrm{AB}}\, M_y^A \\
\frac{\mathrm{d}M_z^B}{\mathrm{d}t} &= -\frac{M_z^B - M_0^B}{T_{1B}} - \omega_1\, M_y^B - k_{\mathrm{BA}}\, M_z^B + k_{\mathrm{AB}}\, M_z^A
\end{aligned}
$$

Solving Bloch-McConnell Equations

Let $\vec{M} = [M_x^A, M_y^A, M_z^A, M_x^B, M_y^B, M_z^B]^T$. Then:

$$\frac{\mathrm{d}\vec{M}}{\mathrm{d}t} = A\,\vec{M} + \vec{b}$$

$$A = \begin{bmatrix} -\left(\frac{1}{T_{2A}} + k_{\mathrm{AB}}\right) & \Delta\omega_A & 0 & k_{\mathrm{BA}} & 0 & 0 \\ -\Delta\omega_A & -\left(\frac{1}{T_{2A}} + k_{\mathrm{AB}}\right) & \omega_1 & 0 & k_{\mathrm{BA}} & 0 \\ 0 & -\omega_1 & -\left(\frac{1}{T_{1A}} + k_{\mathrm{AB}}\right) & 0 & 0 & k_{\mathrm{BA}} \\ k_{\mathrm{AB}} & 0 & 0 & -\left(\frac{1}{T_{2B}} + k_{\mathrm{BA}}\right) & \Delta\omega_B & 0 \\ -\Delta\omega_B & k_{\mathrm{AB}} & 0 & -\Delta\omega_B & -\left(\frac{1}{T_{2B}} + k_{\mathrm{BA}}\right) & \omega_1 \\ 0 & 0 & k_{\mathrm{AB}} & 0 & -\omega_1 & -\left(\frac{1}{T_{1B}} + k_{\mathrm{BA}}\right) \end{bmatrix}$$

$$\vec{b} = \left[0, 0, \frac{M_0^A}{T_{1A}}, 0, 0, \frac{M_0^B}{T_{1B}}\right]^T$$

Solving Bloch-McConnell Equations (cont.)

A system of linear ODEs with constant coefficients can be solved numerically using matrix exponential.

Given $\frac{\mathrm{d}\vec{M}}{\mathrm{d}t} = A\,\vec{M} + \vec{b}$, by variation of parameters,

$$\begin{aligned}\vec{M}(t) &= \mathrm{e}^{tA}\,\vec{M}(0) + \int_0^t \mathrm{e}^{(t-s)A}\,\vec{b}\,\mathrm{d}s \\ &= \mathrm{e}^{tA}\,\vec{M}(0) + \mathrm{e}^{tA}\left(\int_0^t \mathrm{e}^{-sA}\,\mathrm{d}s\right)\vec{b}\end{aligned}$$

Using the definition for the integral of a matrix exponential:

$$\int_0^t \mathrm{e}^{-sA}\,\mathrm{d}s = (I - \mathrm{e}^{-tA})\,A^{-1}$$

we get:

$$\vec{M}(t) = \mathrm{e}^{tA}\,(\vec{M}(0) + A^{-1}\,\vec{b}\,) - A^{-1}\,\vec{b}$$

Simulation Of a Z-Spectrum

- By calculating $\vec{M}(t)$ for t_{sat}, we get $M_z^A(t_{\mathrm{sat}})$ — the saturated signal of pool A after irradiation at $\Delta\omega_A$!
- By solving Bloch-McConnell equations for different saturation frequencies we can simulate a CEST experiment and plot the Z-spectrum.
- Fitting a Z-spectrum according to Bloch-McConnell equations is a powerful tool to characterize the dilute pool and extract its relaxation times, exchange rate with the large pool, and its fraction in the solution.

Simulation Of a Z-Spectrum

- By calculating $\vec{M}(t)$ for t_{sat}, we get $M_z^A(t_{\mathrm{sat}})$ — the saturated signal of pool A after irradiation at $\Delta\omega_A$!
- By solving Bloch-McConnell equations for different saturation frequencies we can simulate a CEST experiment and plot the Z-spectrum.
- Fitting a Z-spectrum according to Bloch-McConnell equations is a powerful tool to characterize the dilute pool and extract its relaxation times, exchange rate with the large pool, and its fraction in the solution.

Why isn't this method frequently implemented in fitting?

Why resort to analytical solutions with approximations?

Limitations

- Matrix exponential calculation is (in principle) very computationally expensive.
- Simulation of a Z-spectrum with N data point requires performing N matrix exponential calculations.

4. The four horsemen of incomputability

This wisdom took decades to crystallise – from the point of view of numerical computing, some operations in magnetic resonance are in the whole separate category of horrible. Every time you use them, a kitten dies. A kitten with big eyes (Fig. 2).

1. **Diagonalisation.** We were all brought up with energy level diagrams – get the Hamiltonian, solve the secular equation... this works for a two-spin system. With more spins, two nasty properties of the diagonalisation operation manifest themselves:

   (a) Spin Hamiltonians are always sparse, but their eigenvectors are not – memory and bandwidth will be needed store and move every single one of those $16{\cdot}2^N$ bytes.

78 *I. Kuprov / Journal of Magnetic I*

vector multiplications, meaning that the cost is negligible relative to the other mathematics that is going on. The sage advice here: *avoid operations that cost more than a few matrix-vector products.*

4. **Matrix exponential.** We have yet to vanquish this one. There are places in magnetic resonance where it is unavoidable,

Ilya Kuprov — Defeating the matrix, 2019

Limitations (cont.)

- Applying a fitting algorithm to a solution of a set of ODEs is difficult.
- Conventional fitting procedures (e.g. non-linear least squares) rely heavily on the initial estimate of the fitting parameters and on the sensitivity of the design matrix (A in BM equations) to noise in the data.

Limitations (cont.)

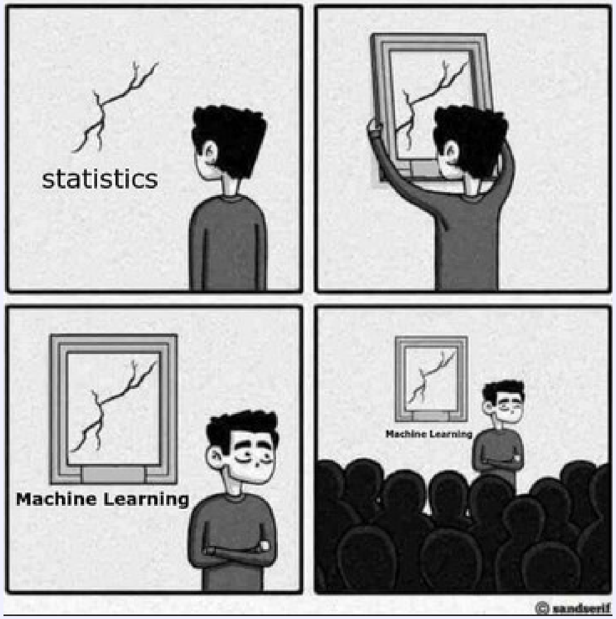
- Applying a fitting algorithm to a solution of a set of ODEs is difficult.
- Conventional fitting procedures (e.g. non-linear least squares) rely heavily on the initial estimate of the fitting parameters and on the sensitivity of the design matrix (A in BM equations) to noise in the data.

- Analytic approaches provide a formula for the Z-spectrum and are much more easy to implement.
- *However*, approximations do not always hold in Li-CEST, where transverse relaxation times are very short (ms to $\mu\mathrm{s}$).
- Also, extracting all parameters may require repetitions of the experiment.

Everything considered, in Li-CEST the **numerical** approach seems most reliable.

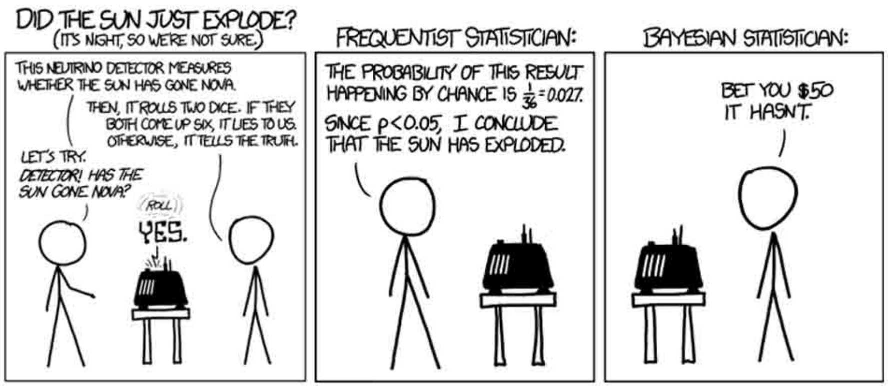
Which means that probability theory statistics must be applied rigorously.

(Call it Machine Learning to make it sound better)



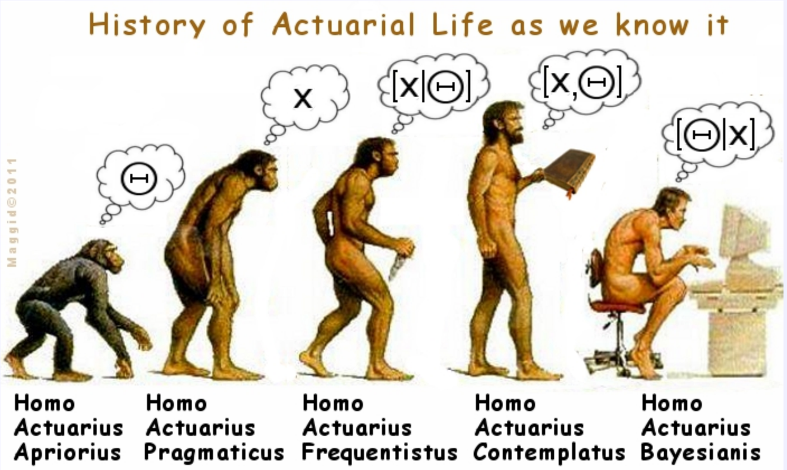
Are You Frequentist?

- Least squares fitting is equivalent to **maximum likelihood estimation**, which is conventionally **Frequentist** by nature.
- **Frequentism**:
    - Unknown model parameters are *fixed*. The model is *deterministic*.
    - Optimal parameters recreate the system's response with the *highest frequency*.



Or Are You Bayesian?

- **Bayesianism:**
  - Unknown model parameters are *randomly distributed*.
  - Parameters are given **prior distributions** according to belief/information
  - Evidence from data *updates* our belief.
  - Optimal parameters are taken from the **posterior distribution**.



Bayes theorem states that given a data set $(\boldsymbol{x}, \boldsymbol{y})$ and model f with parameters $\boldsymbol{\theta}$,

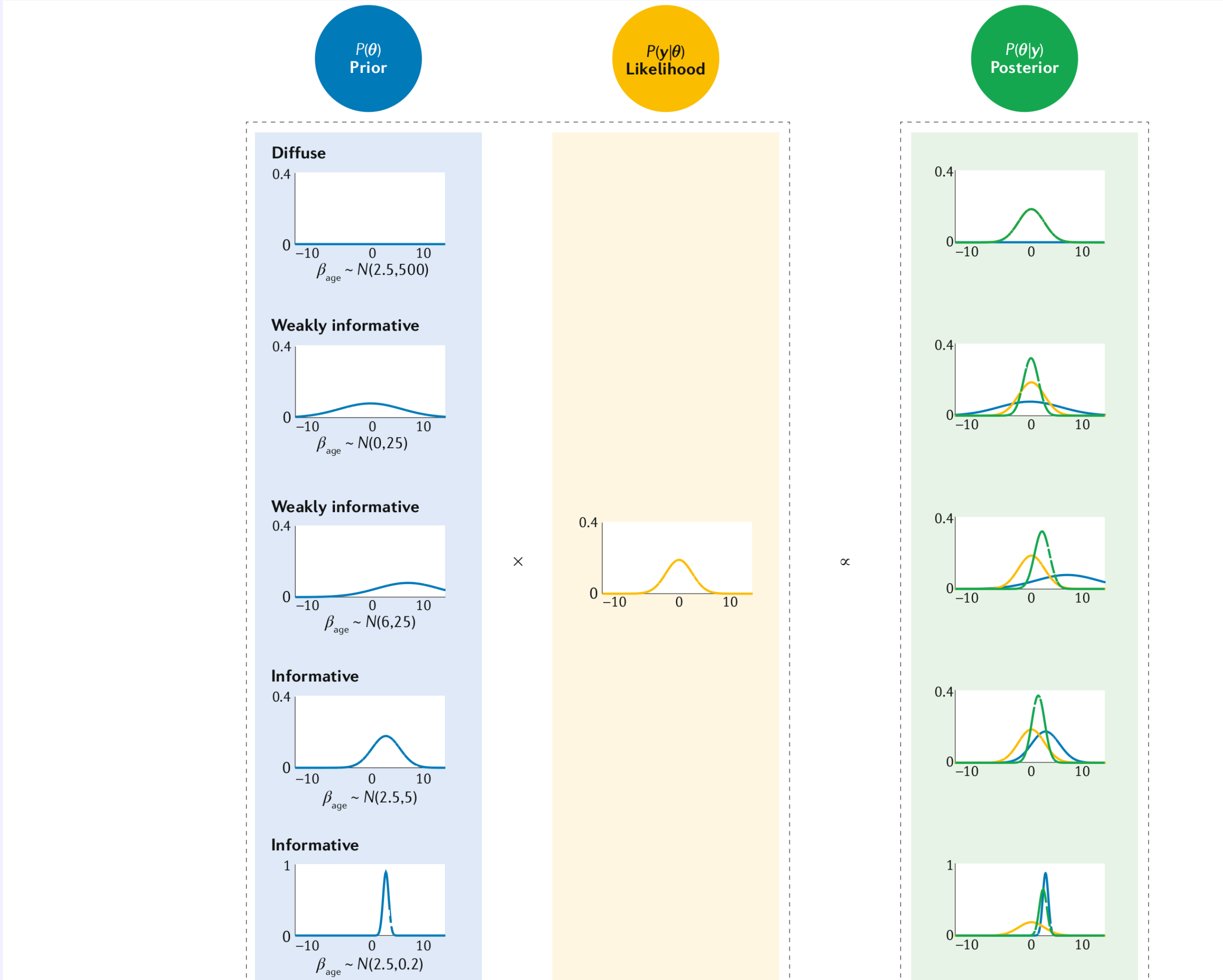
$$P(\boldsymbol{\theta}|\boldsymbol{y}) \propto P(\boldsymbol{y}|\boldsymbol{\theta}) \cdot P(\boldsymbol{\theta})$$

- $P(\boldsymbol{\theta})$ — Encodes information **prior** to measurement.
- $P(\boldsymbol{y}|\boldsymbol{\theta})$ — **Likelihood** of the model to recreate the data.
- $P(\boldsymbol{\theta}|\boldsymbol{y})$ — Encodes all information **a posteriori** to measurement.

The minimum mean-squared error estimation of a model parameter θ^i is its **expected value** — an **integral**.

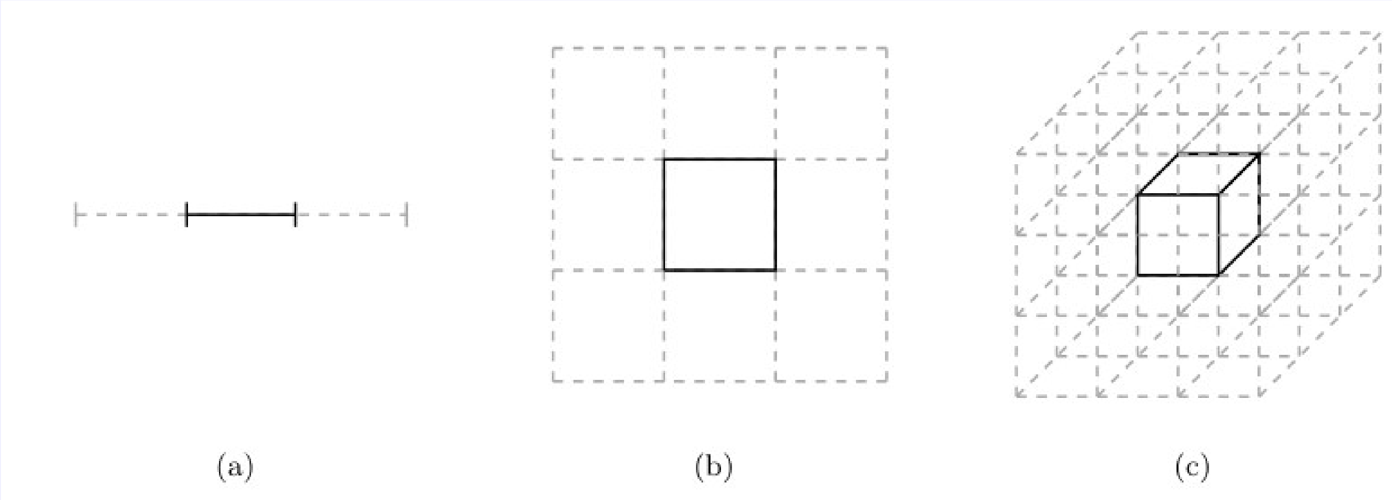
$$\theta^i_{\mathrm{MMSE}} = \mathbb{E}(\theta^i|\boldsymbol{y}) = \int \theta^i \, P(\theta^i|\boldsymbol{y}) \, \mathrm{d}\theta^i$$

Bayes Theorem In Action



Achilles' Heel of Bayesian Inference

- All Bayesian computations reduce to **integration** over parameter space.
- In high-dimensional spaces, there is much more volume outside any give neighborhood than inside.

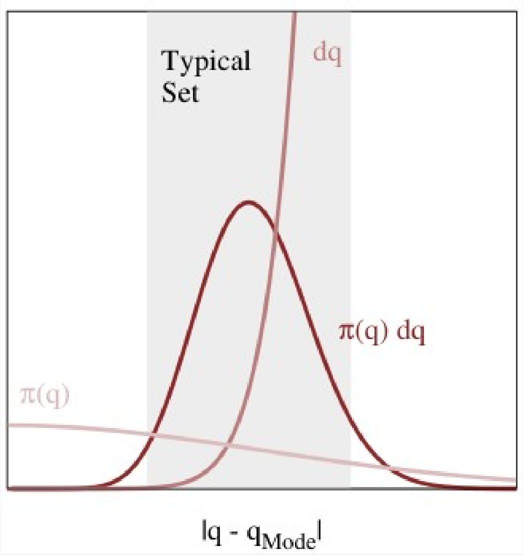


- Volume grows exponentially away from the posterior mode.

- The immediate neighborhood around the mode features large densities, but its volume is very small.
- Most volume sits on the tail of the probability distribution.

Conclusion: both are negligible in expectation calculation

- The only significant contributions come from the **typical set**.



The Typical Set in High-Dimensional PDFs

- Estimation of expectations is done by averaging over the **typical set**.
- The typical set is a **surface** around the mode where the probability is concentrated.

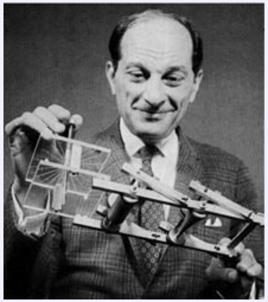
- Modern Bayesian Inference tackles the challenge of **efficient exploration** of the typical set.

Modern Bayesian Inference

- Efficient and robust parameter estimation in Bayesian Inference relies on:
  - Identification of the typical set within parameter space.
  - Generating samples from the typical set and calculating expectations.

- Luckily, development of the first atomic bomb had the same problem.

Manhattan Project and Monte Carlo



- In the late 1940s, *Stanislav Ulam, John von Neumann, et. al* developed a numerical method to calculate high-dimensional integrals as they were investigating neutron diffusion in the core of a nuclear weapon.
- As the work was secret, the project required a code name.
- Their colleague, Nicholas Metropolis, suggested the name **Monte Carlo**, after the Monte Carlo Casino, where Ulam's uncle would borrow money from relatives to gamble.

- Ulam suggested the use of **stochastic simulation** and **Markov chains** in their research, which was the first instance of **Markov Chain Monte Carlo**.

Note. Ulam conjured the method while playing solitaire.

Monte Carlo Integration

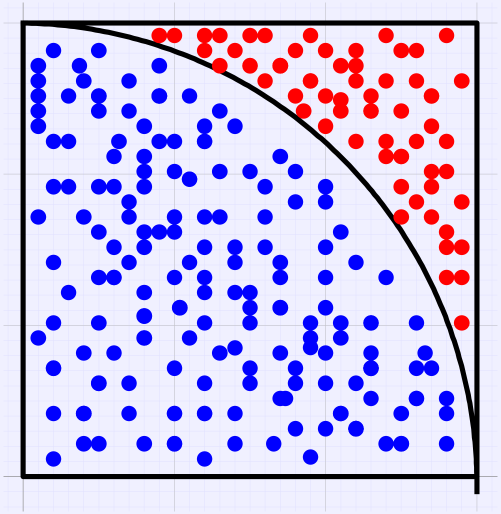
- Monte Carlo simulation is about **sampling** and **scoring**.
- Sampling is done randomly from some distribution.
- Scoring is book-keeping of samples that passed a certain test criterion.

Averaging over successful scoring histories provides an estimate of the integral.

Example: Estimate π

Goal: Estimate π by the ratio of the area under a quadrant.

- Draw a square and a quadrant
- **Sample** points randomly within the square
- **Score** points within the quadrant (distance from origin less than 1)
- **Estimate integral**: Ratio of inside count to total sample count is $\frac{\pi}{4}$.
  Multiply by 4 to estimate π



Markov Chains

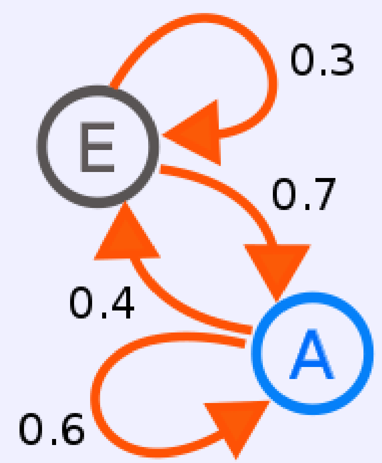
- **Markov chains** may be used to generate samples from complex distributions.
- A Markov chain is a sequence of random samples $\boldsymbol{x}_i$ for which the value $\boldsymbol{x}_{i+1}$ depends only on the previous state $\boldsymbol{x}_i$.

Suppose the system can move between states E and A.

Each transition $i \to j$ has a certain probability T_{ij}.

the matrix $\boldsymbol{T}$ encodes chances for the **next** transition:

$$\boldsymbol{T} = \begin{bmatrix} T_{\mathrm{AA}} & T_{\mathrm{AE}} \\ T_{\mathrm{EA}} & T_{\mathrm{EE}} \end{bmatrix} = \begin{bmatrix} 0.6 & 0.4 \\ 0.7 & 0.3 \end{bmatrix}$$



If $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_n]$ lists the probabilities of being in states $\boldsymbol{x} = [x_1, \ldots, x_n]$, then

$$\boldsymbol{\pi}^{(k+1)} = \boldsymbol{\pi}^{(k)} \boldsymbol{T}$$

By induction,

$$\boldsymbol{\pi}^{(k)} = \boldsymbol{\pi}^{(0)} \boldsymbol{T}^k$$

As n becomes large enough, $\boldsymbol{\pi}^{(k)}$ represents the **steady-state** of the system, or the **stationary probability distribution**.

The local weather may be "sunny" ($x_1 = \mathrm{s}$), "cloudy" ($x_2 = \mathrm{c}$) or "rainy" ($x_3 = \mathrm{r}$).

The transition matrix is:

$$\boldsymbol{T} = \begin{bmatrix} T_{\mathrm{ss}} & T_{\mathrm{sc}} & T_{\mathrm{sr}} \\ T_{\mathrm{cs}} & T_{\mathrm{cc}} & T_{\mathrm{cr}} \\ T_{\mathrm{rs}} & T_{\mathrm{rc}} & T_{\mathrm{rr}} \end{bmatrix} = \begin{bmatrix} 0.3 & 0.4 & 0.3 \\ 0.3 & 0.2 & 0.5 \\ 0.4 & 0.1 & 0.5 \end{bmatrix}$$

Today is a cloudy day. Out of the next three days, on which day is it least likely to rain?

Method: Compare $T_{\mathrm{cr}}^{(k)}$ for $k = 1, 2, 3$.

Example: Weather in Transylvania (cont.)

From $\boldsymbol{T}$, the probability of rain tomorrow is 50.0%.

$$\boldsymbol{T}^2 = \begin{bmatrix} 0.33 & 0.23 & 0.44 \\ 0.35 & 0.21 & 0.44 \\ 0.35 & 0.23 & 0.42 \end{bmatrix}, \boldsymbol{T}^3 = \begin{bmatrix} 0.3434 & 0.2254 & 0.4312 \\ 0.3430 & 0.2258 & 0.4312 \\ 0.3430 & 0.2254 & 0.4316 \end{bmatrix}$$

From $\boldsymbol{T}^2$, the probability of rain on the second day is 44.0%.

From $\boldsymbol{T}^3$, the probability of rain on the third day is 43.12%.

$$\boldsymbol{T}^4 = \begin{bmatrix} 0.343137 & 0.225490 & 0.431372 \\ 0.343137 & 0.225490 & 0.431372 \\ 0.343137 & 0.225490 & 0.431373 \end{bmatrix}$$

Note: $\boldsymbol{T}^{(4)}$ gives the **stationary probability density** to five decimal places.

Why Use Markov Chains for Sampling?

- The unique property of Markov chains is that it is impossible to predict the next state of the system, as the transition is random.
- However, once enough transitions occur the system approaches a steady state and the statistical properties of the equilibrium distribution are revealed.
- In **Markov Chain Monte Carlo**, a Markov chain is constructed with the **posterior** as the equilibrium distribution of the chain.
- Then, once the chain has converged to steady state, any further steps of the chain are effectively samples from the target posterior distribution!

Exploration of the Typical Set via Markov Chains

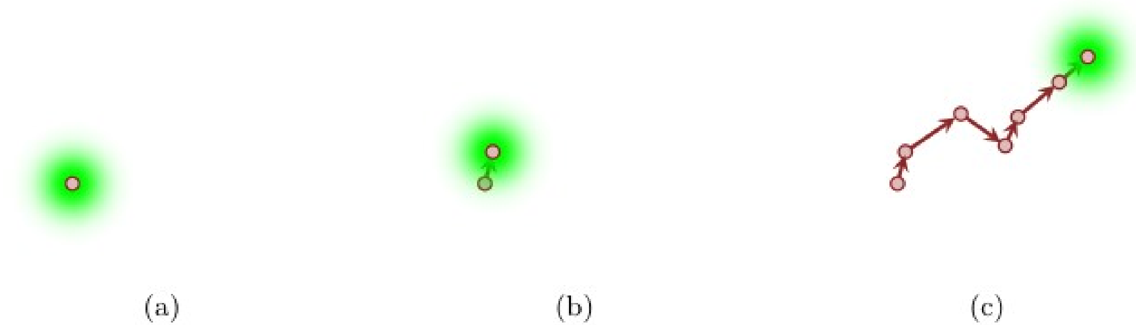


Fig 5. *(a) A Markov chain is a sequence of points in parameter space generated by a Markov transition density (green) that defines the probability of a new point given the current point. (b) Sampling from that distribution yields a new state in the Markov chain and a new distribution from which to sample. (c) Repeating this process generates a Markov chain that meanders through parameter space.*

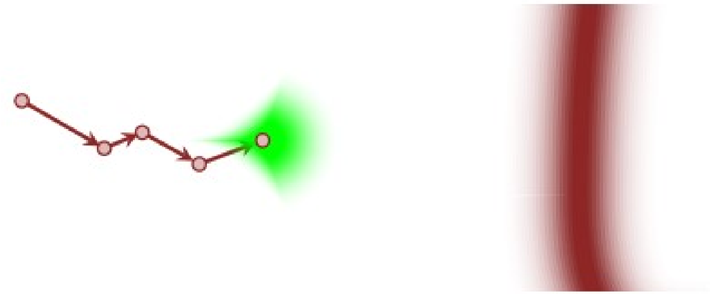


Fig 6. *When a Markov transition (green) preserves the target distribution, it concentrates towards the typical set (red), no matter where it is applied. Consequently, the resulting Markov chain will drift into and then across the typical set regardless of its initial state, providing a powerful quantification of the typical set from which we can derive accurate expectation estimators.*

Let $\hat{f}_N$ be the estimate of some statistic property f after N steps.

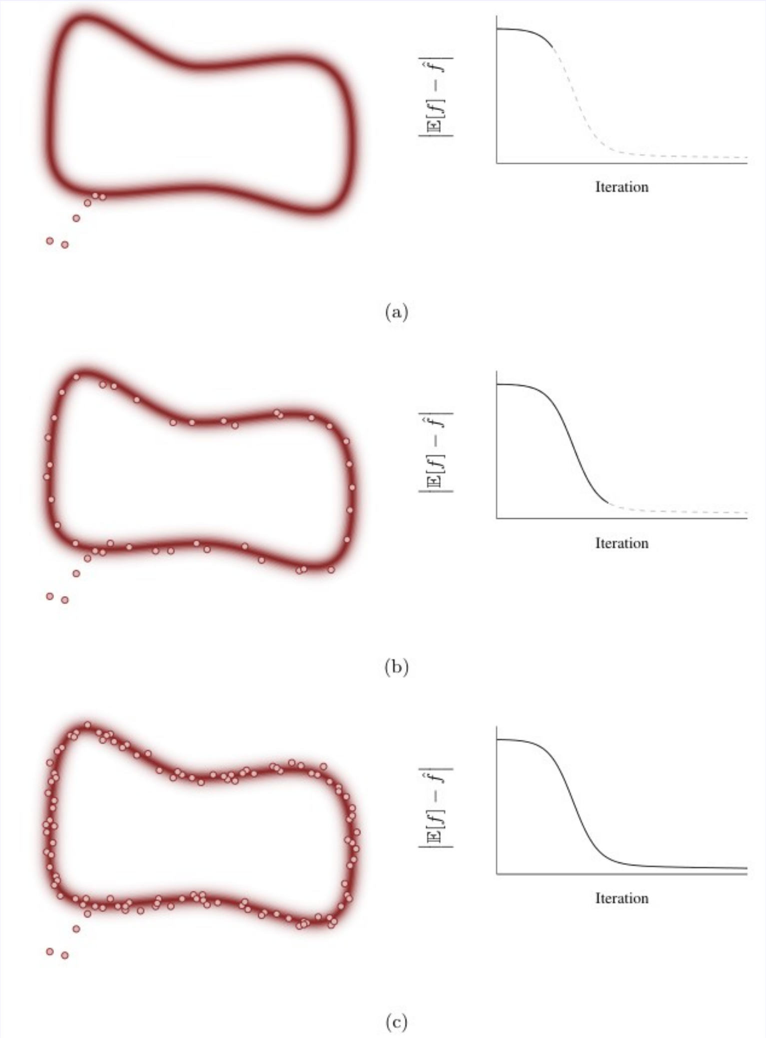
$$\hat{f}_N = \frac{1}{N} \sum_{n=1}^{N} f(q_n)$$

As $N \to \infty$, the estimate must converge to the true expectation $\mathbb{E}[f]$:

$$\lim_{N \to \infty} \hat{f}_N = \mathbb{E}[f]$$

The convergence process can be monitored through $|\mathbb{E}[f] - \hat{f}_N|$.

Convergence of Histories to Expectation (cont.)



a) The Markov chain reaches the typical set.

b) The Markov chain completes the first sojourn of the typical set.

c) Each completed sojourn discoveres more fine details of the typical set and increases the accuracy of the calculation.

Constructing a Markov Transition

Random Walk Metropolis Algorithm (1970)

- Given an initial estimate q of some parameter and a posterior PDF $\pi(q)$,
- Propose a new estimate q' from a Normal distribution centered around q.

$$\mathbb{Q}(q'|q) = \mathcal{N}(q'|q, \Sigma)$$

- Accept the proposal with probability

$$a(q'|q) = \min\left(1, \frac{\mathbb{Q}(q|q')\,\pi(q')}{\mathbb{Q}(q'|q)\,\pi(q)}\right)$$

- The Gaussian is symmetric, so

$$a(q'|q) = \min\left(1, \frac{\pi(q')}{\pi(q)}\right)$$

Critique of Metropolis Algorithm

- Random Walk Metropolis is easy to implement.
- The proposal distribution is biased towards large volumes.
- The Metropolis correction rejects proposals that lead to very small densities.

$\Rightarrow$ The constructed Markov transitions concentrate towards the typical set.

Critique of Metropolis Algorithm

- Random Walk Metropolis is easy to implement.
- The proposal distribution is biased towards large volumes.
- The Metropolis correction rejects proposals that lead to very small densities.

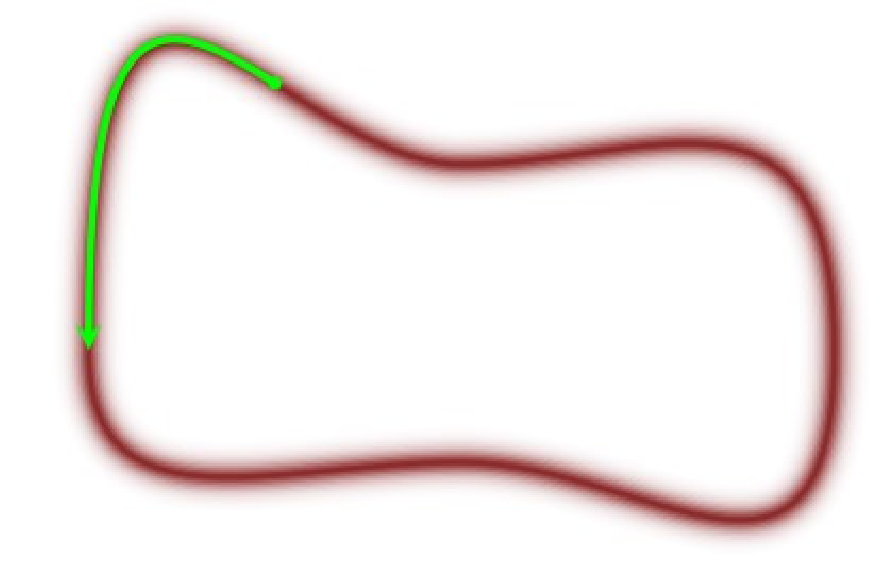
$\Rightarrow$ The constructed Markov transitions concentrate towards the typical set.

How does Random Walk Metropolis scale with increasing dimension?

The Curse of High Dimensionallity

- With increasing dimension the volume exterior to the typical sets overwhelmes the volume interior, and almsot every proposal will fall far-off the typical set and get rejected.
- Because the algorithm is run for a finite number of steps, in most cases there will be only partial exploration of the typical set, which would give biased results.

- The only way to efficiently explore the typical set in finite steps is to **exploit the geometry** of the typical set.
- Ideally, transitions should follow the contours of high probability density.



Hamiltonian Monte Carlo is an algorithm that automatically does this.