ChatGPT

From Cosmic Rays to DNA Damage: Foundations for Space Radiation Effects (Phase 1)

Galactic Cosmic Ray Environment and Stochastic Modeling

Space is pervaded by **galactic cosmic rays (GCR)** – charged particles accelerated to near light speed by supernovae and other astrophysical processes [1]. GCR consist predominantly of high-energy **protons (~87%)**, with about 12% alpha particles (helium nuclei) and a small (~1%) admixture of heavier ions from lithium up to uranium [2]. Their energy spectra peak around a few hundred MeV per nucleon and extend to many GeV/nucleon [3]. Notably, GCR intensities are **anti-correlated with solar activity**: during solar minimum the flux of these particles is highest (solar magnetic shielding is weakest), whereas at solar maximum the GCR flux is suppressed [4]. In practical terms, this means long-term space missions must consider a varying radiation field over the ~11-year solar cycle. GCR are often modeled as an **isotropic, random field** of particles impinging on a spacecraft or human tissue, and a common assumption is that arrival of particles follows a **Poisson process** in time [5]. Under this assumption, the number of cosmic ray hits on a given detector or tissue area in a fixed time follows a Poisson distribution, and the spacing between successive particle hits is memoryless (exponential) – a reasonable approximation given the independent, sparse nature of cosmic ray interactions in space. Such stochastic modeling is essential for capturing the **probabilistic nature** of radiation exposure; it provides not just an average dose rate, but a distribution of possible fluxes and event frequencies that inform uncertainty estimates [6] [7].

To quantitatively characterize the space radiation environment, NASA has developed empirical models and tools incorporating decades of balloon, satellite, and ISS measurements [5] [8]. **NASA's OLTARIS** (On-Line Tool for the Assessment of Radiation in Space) is a web-based system that allows users to obtain environment spectra and compute doses behind shielding. OLTARIS uses the **Badhwar-O'Neill GCR model** (an analytic fit to historical data) with enhancements from **CREME96** (Cosmic Ray Effects on Micro-Electronics, 1996 revision) to generate differential fluxes [9] [10]. CREME96 is a widely used model originally developed to predict single-event upsets in electronics; it provides flux spectra for all relevant ions and incorporates solar modulation effects [9]. By selecting mission parameters (e.g. location, date or solar cycle phase) in OLTARIS, one can obtain GCR spectra (often given as differential flux in particles/cm^2/sr/MeV) for use in downstream analyses [11] [12]. These spectra can serve as input to a **probabilistic radiation field simulation**: essentially sampling random particles from the flux distribution and assigning random arrival times (via a Poisson process) to simulate "runs" of radiation exposure. This forms the first step in *Milestone 1*, stochastically modeling the space radiation field. The **output of this step** is a library of randomly generated radiation "events" (each event with a specific particle type, energy, direction, and time) representing the space environment [13] [7]. Such a synthetic event list can then drive Monte Carlo transport simulations to produce **spatiotemporal dose maps with uncertainty bounds**, as required by the project goals [14] [6].

Galactic cosmic ray energy spectra under different conditions. Panel (A) shows differential flux vs. energy per nucleon for various GCR ions in free space during solar minimum (solid lines) and solar maximum (dashed lines). The flux of lower-energy (hundreds of MeV) particles is notably higher at solar minimum. Panel (B) shows how shielding (20 g/cm^2 Al) dramatically reduces the flux of lower-energy particles entering a tissue site (here, the blood-forming organ) [15] [16].

In addition to GCR, space radiation includes **solar energetic particles (SEP)** from solar flares/CMEs and trapped radiation belts around Earth. These components are important for low Earth orbit or solar storm scenarios, but GCR – especially the heavy ions (HZE particles) – dominate for deep-space chronic exposure [17]. NASA's models (e.g. CREME96 and newer CRÈME-MC [18] [19]) account for both GCR and sporadic SEP event spectra. For Phase 1, the focus is on GCR as the **background radiation field**. It is worth noting that while protons and He are most abundant, the **biological impact per particle increases with atomic number and energy**. A single iron nucleus (Z=26) can produce a dense trail of ionizations in tissue far more damaging than a proton's track [17] [20]. This is reflected in standard radiation metrics: for instance, heavy ions contribute only a minor fraction of the particle fluence but a disproportionately large fraction of the **dose and biological dose equivalent** (which weights dose by a quality factor for damage) [21] [22]. In one analysis, protons were ~90% of GCR fluence yet heavy ions (Z>2) contributed over half of the dose equivalent because of their high linear energy transfer [21]. **Understanding the composition and energy spectrum of the radiation field is thus foundational**, as it informs how many particles of each type must be simulated and what energies they carry into the target – the starting point for any mechanistic model of radiation effects.

Radiation Transport and Energy Deposition (LET Distributions)

Once the external radiation environment is characterized, the next step is to model how these particles **transport through matter and deposit energy** in biological targets (e.g. tissue-equivalent material or actual tissue geometry). When high-energy charged particles penetrate matter, they interact via electromagnetic forces with atomic electrons and nuclei, losing energy along their path. The **linear energy transfer (LET)** is a key quantity describing this: it is the average energy $\mathrm{d}E$ lost by the particle per unit path length $\mathrm{d}l$ in the medium [23]. LET is usually expressed in keV/μm for biological materials. Conceptually, LET is similar to **stopping power** – it measures how densely the particle expends its energy. Low-LET radiation like X-rays or fast electrons deposits energy sparsely (many small energy transfers spread out), whereas high-LET radiation like heavy ions deposits energy **densely along a track** [24]. The ICRU formally defines LET as the mean energy lost due to collisions with electrons per unit distance [23]. Importantly, this is an average; the actual energy deposition is **stochastic**, comprising many individual ionization events. A heavy ion's path can be thought of as a **shower of microscopic energy transfers** – tens of thousands of ionizations and excitations occur along a single **track** through a cell, concentrated near the trajectory for a high-LET particle [17] [20]. In dosimetry, one often distinguishes unrestricted LET (total energy loss per length) from restricted LET (excluding energy of secondary electrons above a cutoff), but for our purposes LET gives a general measure of track density [25]. High-LET particles produce **clusters of ionization** (densely ionizing "blobs"), whereas low-LET radiation produces isolated ionizations ("spurs") spaced far apart in the nanometer scale. This distinction underlies why heavy ions tend to cause complex DNA damage: their energy deposition events can concentrate multiple lesions in a small volume.

Mathematically, we can model the **energy deposition kernel** for a particle track – essentially a function describing how energy from a single particle is distributed in the surrounding medium. In radiation therapy physics, dose from broad beams is sometimes calculated by convolving a fluence with a pre-computed energy deposition kernel [26]. At the scale of cellular targets, an energy deposition kernel could describe, for example, the radial distribution of dose around the path of a heavy ion. High-LET ions generate a **core of intense ionization** along their path and a penumbra of δ-ray electrons extending outward a few micrometers [27] [28]. In contrast, a fast electron (low-LET) deposits energy in a more diffuse, zig-zagging path with fewer dense clusters. To build an accurate model, one must account for the **stochastic nature of these interactions**: the number of ionizations in a segment of track is Poisson-distributed, and the step-by-step energy loss fluctuates (described by theories of straggling) [29] [28]. Tools like NASA's HZETRN and Monte Carlo codes handle this by either deterministic methods or random sampling.

Radiation transport codes generally fall into two categories: (1) **Deterministic codes** (e.g. HZETRN, High-Z and Energy Transport code) which solve the Boltzmann transport equation for particles propagating through matter, and (2) **Monte Carlo codes** (e.g. GEANT4, FLUKA, MCNP) which simulate millions of individual particle histories. Deterministic methods are highly efficient for thick shields and complex spectra since they propagate particle fluences continuously, but they rely on mean properties and approximations (such as continuous slowing-down or limited angle scattering) [30] [31]. Monte Carlo methods, by contrast, **track each interaction explicitly** – random distances between collisions are sampled based on interaction cross-sections, and secondary particles are generated and tracked in turn [32]. This approach can capture all the physics (energy straggling, angular deflections, nuclear fragmentations, secondary radiation) with high fidelity, at the cost of significant computation. In the context of space radiation, NASA historically used HZETRN for rapid assessment of GCR shielding and dose behind spacecraft materials [30] [33]. HZETRN was developed starting in the 1980s when computing power was limited [30] [34], but it has evolved (through 3D-HZETRN) to include improved nuclear collision models and even to interface with CAD geometries [35] [36]. Today, **hybrid approaches** are common: using deterministic codes for bulk transport and Monte Carlo for detailed local dosimetry or verification [37] [38]. OLTARIS itself integrates HZETRN for its dose calculations behind user-defined shielding, and its development adhered to NASA's modeling standards (NASA-STD-7009) to ensure verification & validation [33] [38]. Meanwhile, in the research community, Monte Carlo **track-structure simulations** have become invaluable for microdosimetry – simulating particle tracks on the scale of DNA. For example, the next-generation version of CREME (CRÈME-MC) was designed to incorporate **Geant4** as its physics engine to explicitly model energy loss fluctuations and nuclear interactions that the older CREME96 handled with analytical approximations [19] [39]. This shift reflects a general trend: **physics-based Monte Carlo models** now complement empirical models, providing greater accuracy for modern micro- and nano-scale analyses of radiation effects [40] [41].

For *Milestone 1*, we need to construct a **Monte Carlo framework for particle tracks through tissue-equivalent material**. In practice, this means for each sampled radiation event (from the environment model), we simulate the particle traversing a volume of water or tissue phantom, and record the energy deposition along its path. Each step involves sampling a random free path length from an exponential distribution (characteristic of a Poisson process in which interactions occur with a constant mean free path) – this is a classic application of **Markov processes** in radiation transport, since the probability of a collision in the next infinitesimal path segment is memoryless. When a collision occurs, a **differential cross-section** comes into play: for example, given a collision, the probability distribution of energy $ \Delta E$ transferred or the angle θ of deflection is governed by differential cross-sections $d\sigma/d\Omega$ or $d\sigma/dE$. By sampling from these distributions (using tabulated physics data or formulas like the Bethe equation for electronic stopping and the Molière theory for scattering), the simulation updates the particle's energy, direction, and produces secondaries as needed [42] [43]. Repeating this until the particle's energy falls below a cutoff or it exits the volume yields one complete **particle track** with its associated energy deposition profile. Aggregating many such tracks builds up statistical distributions of deposited energy – for instance, we can derive the **LET spectrum** in the material (frequency of tracks with a given LET) [44] [45], or spatial dose maps (3D energy deposition patterns). *Milestone 1* specifically mentions *"energy deposition kernels that capture LET distributions"*. One way to interpret this is that we will compute how energy is spread out (radially, longitudinally) around a particle's path for various LET particles. For example, we might create a radial dose kernel for a 1 GeV proton vs. a 1 GeV Fe ion and observe the higher LET of Fe leads to a higher local energy density in the core of the track. These kernels can then be used to **convolve with particle fluence** in a larger geometry or to rapidly estimate microdosimetric quantities without simulating every track in detail [26] [46].

The outcome of the Monte Carlo transport simulation is a **spatiotemporal map of energy depositions** – essentially a 4D record (3D position + time) of "hits" in the tissue volume, or equivalently a sequence of

energy deposition events with known locations and times. This is the bridge between physics and biology: energy deposited in nanometer volumes causes chemical ionizations and excitations that can damage biomolecules. It's important to note the **scale**: a single 1 GeV iron ion through a cell nucleus (~10 μm) might deposit on the order of 10^5 keV in that nucleus [17], which could cause dozens of DNA double-strand breaks in one traversal. By contrast, a 1 GeV proton (lower LET) might pass through the same nucleus leaving only a few keV, potentially causing zero or one DSB. This stark difference highlights why heavy ions are a unique concern for deep space travel – they create a **shower of damage** in a single event. Our Monte Carlo framework will capture these stochastic differences event-by-event. Modern computational tools will assist greatly: for example, one could use the Julia language's strengths in scientific computing to implement this. Julia offers packages like **MonteCarlo.jl** (for generic Monte Carlo simulation) and a rich **DifferentialEquations.jl** suite for solving motion and interaction equations. Although no off-the-shelf Julia package will magically perform Geant4-level radiation transport, Julia's speed and libraries for random number generation and parallelism can be leveraged. We might generate track step lengths by sampling exponential distributions (using Distributions.jl) and even solve equations of motion if needed (for e.g. charged particle deflection in magnetic fields, though that is a minor factor for straight-line trajectories in tissue). The **SciML** ecosystem in Julia is also relevant beyond transport: it provides capabilities for **discrete stochastic simulations (Markov jumps)**, **ODEs**, and even **stochastic differential equations (SDEs)** [47] [48]. This means once energy depositions are obtained, we have the tools to simulate downstream chemical reactions or even biological signaling with the same platform. In summary, by combining a **stochastic space radiation source model** with a **Monte Carlo transport simulation**, we will obtain the fundamental physical input to Phase 1: where, when, and how much energy is imparted in a biological system by space radiation. The next step is connecting this energy deposition to molecular damage.

Radiation-Induced Molecular Damage: DNA Lesions and Damage Distribution

Ionizing radiation affects biological tissue by altering atoms and molecules – primarily through **ionization (ejecting electrons)** and excitation of molecules. The most critical target is DNA, the carrier of genetic information. Radiation can damage DNA in numerous ways, classified broadly into **single-strand breaks (SSBs)**, **double-strand breaks (DSBs)**, and **base damages or cluster lesions**. A **single-strand break** involves a break in the sugar-phosphate backbone of one strand of the DNA double helix. These are common (cells might experience tens of SSBs per gray of low-LET radiation) and typically repairable by base excision repair. A **double-strand break** is more severe: both strands are broken, either by a single track (if two nearby ionizations occur across from each other or within ~10 bp) or by two independent tracks in proximity. **Clustered DNA lesions** refer to two or more damages (strand breaks or base modifications) within one or two helical turns of DNA (approximately 10–20 base pairs) [49] [50]. These clusters, also called **locally multiple damaged sites**, are a **signature of ionizing radiation** because the dense ionization from a track can create multiple damages in a short segment of DNA [51] [52]. For example, a single heavy ion traversal might produce a DSB accompanied by several nearby base oxidations – a complex lesion cluster that is difficult for the cell to repair accurately. In fact, classical radiation biophysics (dating back to Ward in the 1980s) proposed that the **lethality of radiation** is largely due to clustered damage that cannot be simply "patched" by normal repair enzymes [52] [53]. Modern studies continue to support that **clustered DNA lesions are highly mutagenic and repair-resistant**, often leading to double-strand break formation or misrepair [54] [55].

High-LET radiation is particularly effective at creating **complex DNA damage**. As the MDPI review by Mladenova et al. (2022) notes, there is growing evidence that **heavy ions induce DSBs that are more complex or even clustered together**, challenging the cell's repair machinery [56]. Such **complex DSBs** may have chemically modified ends or be part of a cluster with nearby breaks, which tends to engage

error-prone repair pathways and can lead to chromosome aberrations [57] [58]. By contrast, low-LET X-rays tend to cause more isolated, simpler breaks that are more often repaired by high-fidelity pathways. Our Phase 1 model will need to capture not just the number of DSBs, but also the **quality** or complexity of those breaks. One approach (in a mechanistic model linking physics to biology) is to define **damage thresholds** based on local energy deposit: for instance, a rule could be that if a certain amount of energy (e.g. > 500 eV) is deposited in a DNA segment of a few nanometers, a DSB occurs, whereas smaller deposits cause SSB or base damage [54]. Detailed track-structure simulations like **PARTRAC** or **Geant4-DNA** follow physically each ionization near DNA and can output the distribution of damages. Since our aim is an integrative model, we might implement a simpler **probabilistic damage mapping**: using the energy deposition data from the Monte Carlo step, we can overlay a model of the cell's DNA geometry and flag damage events. For example, we could divide the cell nucleus into chromatin domains and, knowing the DNA density, calculate a Poisson probability of SSB or DSB per unit energy deposited in that volume [59] [60]. This would result in a **list of DNA damage events** with genomic positions.

To build confidence, we will lean on experimental radiobiology findings. Textbook data (e.g. Hall & Giaccia's *Radiobiology for the Radiologist*) indicate that about **40 DSBs per cell per Gy** are induced by X-rays (low LET), whereas an equal dose of heavy ions produces many fewer tracks but each can yield **clustered DSBs** and complex lesions. In fact, a single iron ion of sufficient energy traversing a nucleus can produce on the order of 10 DSBs (hence the concept of "kill radius" for heavy ions) [17]. Therefore, it's not just total dose but track structure that matters. Our model will explicitly simulate each ion's **damage pattern**. If an ion passes through the nucleus, we mark the sites where its energy spikes occurred. If the energy spike overlaps a DNA segment, we create a damage event. This requires mapping physical space to the genome coordinate space – addressed below – but conceptually it's building a **damage matrix**: for each radiation track, for each DNA segment (e.g. gene or chromatin bin), increment damage counts. The *output of Milestone 2* will be **probabilistic damage maps** – essentially the likelihood or frequency of DNA lesions across the genome given a radiation exposure [61] [62].

It's important to separate **direct vs. indirect effects** of radiation on DNA. **Direct effects** mean the radiation's charged particles directly ionize the DNA molecule or its immediate hydration shell, breaking bonds. **Indirect effects** refer to damage by reactive chemical species (free radicals) produced by radiation in water or other molecules, which then diffuse to damage DNA. In typical aqueous biology, about two-thirds of DNA damage from X-rays is indirect (mediated by hydroxyl radicals), whereas for high-LET radiation the direct component is larger [63] [64]. The primary actor in indirect DNA damage is the **•OH (hydroxyl radical)**, generated by radiolysis of water [64]. Other species like H• (hydrogen atom) and e^-_{aq} (solvated electron) also form, but •OH is the most reactive and abundant radical causing DNA base lesions and strand breaks [64]. Indirect effect unfolds in the chemical stage of radiation action: within about 10^{-12} to 10^{-6} seconds after the initial ionizations, radicals diffuse and react. To model this, we incorporate ***reaction-diffusion equations*** for radical species (Milestone 2). Essentially, we treat the locations of energy depositions as sources of radicals and then simulate how those radicals move and interact. A common approach is to use ***reaction kinetics and diffusion equations***: for example, one could write partial differential equations for the concentration $c(r,t)$ of OH radicals, which include a diffusion term $D \nabla^2 c$ (Fick's law) and reaction terms that consume OH (e.g. reactions with DNA or scavengers). However, because initial radical positions are highly localized along tracks, the system is not well-described by uniform concentrations – it's a }**point cloud of radicals** that diffuse outward. This is where a **Brownian motion** perspective is useful: each radical executes a random walk from its creation site, which can be modeled by either solving the diffusion PDE or by Monte Carlo simulation of individual radical trajectories. The latter is essentially a **stochastic diffusion process** – one can generate random displacements for radicals over small time steps (with root-mean-square displacement $\sqrt{6Dt}$ in 3D for diffusion coefficient D) to emulate Brownian motion, and include random reactions when radicals encounter DNA. Brownian motion is a classic example of a Markov

process (the future position depends only on the current position, with no memory, and increments are Gaussian for normal diffusion). The probabilistic theory here, dating back to Einstein's 1905 work, underpins why diffusion can be treated with random walks or equivalently with the diffusion equation.

NASA and other researchers have developed sophisticated models for the radiation chemistry stage. For instance, Plante et al. introduced a **Green's function method to solve diffusion-kinetic equations** for radicals, allowing partial analytical treatment of how radicals migrate and react in tracks [65] [66]. Recombination of radicals, competition between reactions, and the presence of chemical scavengers (like glutathione in cells) all play a role. For our purposes, a reasonable starting point is to focus on **DNA damage yield from radicals**: we can implement a simpler reaction-diffusion system where we have a certain probability per •OH radical to damage DNA if it is within a certain distance. For example, if an OH radical is generated within a few nanometers of DNA and not scavenged within ~nanoseconds, it might cause a base damage or SSB. One can draw on experimental diffusion-controlled reaction rates: OH has a diffusion constant on the order of 2.8×10^{-9} m^2/s in water and a lifetime of ~nanoseconds in a cell due to reactions. In a modeling context, an **SDE (stochastic differential equation)** or **Gillespie algorithm** can be used to simulate these chemical reactions. Julia's DiffEq library includes **Jump processes** for exactly this kind of chemical kinetics simulation (stochastic reactions) [67] [68]. Alternatively, we could solve deterministic reaction-diffusion PDEs using Finite Difference or Finite Element methods (which Julia can do via packages or by linking to existing C/C++ solvers). Since Phase 1 emphasizes mechanistic modeling, we might implement a **radical propagation module** that takes each energy deposition site, spawns a number of OH radicals (radiolysis yields can be taken from literature, e.g. ~0.2 μM per Gy for OH), and then propagates them with diffusion until they either react with a target or decay. The result of this module will be an added set of DNA damages (mostly SSB and base damage) that occur slightly after the initial physical hits. Reaction-diffusion modeling also introduces the concept of **spatiotemporal overlap** of tracks: if two radiation tracks pass nearby in time (< microseconds), their radical clouds can overlap, potentially causing nonlinear enhancement of damage (though for space radiation at chronic low flux this is rare).

In summary, **Milestone 2**'s "molecular damage landscape" is achieved by combining the **direct deposition-induced damage** (from track structure) with the **indirect radical-mediated damage**. We will produce a comprehensive map of DNA lesions. This map can be conceptualized as a tensor or matrix: dimensions could be (genomic location) × (damage type), containing probabilities or counts [69]. For instance, one axis might index the genome in, say, 1 Mbp bins or gene regions, and the other axis lists lesion types (SSB, DSB, oxidized base, etc.), and entries are expected values of lesions per cell per Gray, or something similar. The phrasing *"tensor fields over genomic space"* [70] suggests we want a spatially resolved damage probability distribution – effectively, a function $D(g)$ giving the probability density of damage at genomic coordinate g. This naturally brings us to the role of **chromatin accessibility** in modulating damage.

Chromatin Structure and Damage Distribution across the Genome

Not all DNA in the nucleus is equally exposed to damage. The DNA is packaged into chromatin, which exists in states ranging from **euchromatin (open, accessible)** to **heterochromatin (condensed, shielded)**. It has been experimentally observed that **chromatin context influences the yield of radiation-induced breaks** [71] [72]. Open chromatin is more accessible not only to transcription factors but unfortunately also to damaging agents like radicals or even the radiation track itself, whereas densely packed nucleosomes can provide a degree of physical protection. A 2020 study by Sellami et al. demonstrated that **nucleosome-bound DNA is partly shielded from radiation**, leading to fewer observable breaks in nucleosome-dense regions [72] [73]. In other words, **open chromatin regions**

(OCRs) are more vulnerable to ionizing radiation than nucleosome-dense areas [74] [75]. This makes intuitive sense: a tightly wound DNA around histones presents a smaller target and is partly occluded by protein, whereas linker DNA in open regions is an easier target. Additionally, the **indirect effect** is mitigated in condensed chromatin because histone proteins readily scavenge radicals. Indeed, simulations (RITRACKS code) indicate that radicals diffusing in a nucleosome environment often react with amino acids in histones before reaching the DNA [76] [77]. This leads to a fascinating periodic pattern: DNA wrapped on nucleosomes can show a **periodic damage pattern** (every ~10 bp, corresponding to how DNA faces outward from the histone) [78] [79]. The net effect is that heterochromatin (tightly packed) tends to suffer slightly fewer initial DSBs, but those it does incur may be harder to repair (because the chromatin must relax for repair machinery to access). Meanwhile, euchromatic regions (open, often gene-rich) might accumulate more breaks initially [71]. It has been shown, for example, that **heterochromatin is more protective against DSB formation by IR** – human stem cells with more open chromatin sustain more breaks than differentiated cells with more heterochromatin [80].

Our model will incorporate **chromatin accessibility data** to modulate damage probabilities. Public datasets like **ATAC-seq (Assay for Transposase-Accessible Chromatin)** or DNase hypersensitivity maps provide genomic coordinates of open chromatin across the genome. For a given genomic location, we can obtain an **accessibility score**. The plan is to **map damage probability distributions onto genomic coordinates using these accessibility weights**. Practically, if our initial physics-based damage map says "X DSB in this 1 Mbp region on average," we might redistribute those X across the region's genes or sub-regions proportional to an accessibility function. Regions with high chromatin accessibility (euchromatic, often active regulatory regions) would get a higher share of the damage probability. Conversely, densely packed regions (heterochromatin, e.g. centromeres or telomeres in certain cells) would get a lower weight. This way, we refine the damage map to reflect biological reality: DNA in open state is more likely to be damaged per unit energy deposited. In essence, we introduce an **ontological layer**: the genome isn't a uniform string in our model, but a landscape of varying vulnerability informed by epigenomic data. This addresses the user's note that Phase 1 should incorporate Phase 2 knowledge where relevant – here, understanding chromatin's role (which is often discussed in the context of DNA damage response and gene regulation) helps inform the initial damage distribution. Additionally, by mapping damage to specific genomic features, we set the stage for Phase 2: for example, if many breaks occur in gene promoters or in certain chromosomes, the downstream **gene regulatory network (GRN) activation** can be modeled accordingly (some genes might be more likely to be disrupted or mis-expressed due to damage in regulatory regions). We effectively create a **bridge from physical damage to biological effect** by localizing where hits occur in the genome. Studies have even begun correlating radiation-induced DNA breaks with 3D genome organization and transcriptional activity [71] [81]. Our approach aligns with those trends – acknowledging that the **3D chromatin structure** (which parts of the genome are buried in nuclear lamina vs. which are in open transcription factories) will bias the damage pattern and thereby influence which signaling pathways are activated.

Concretely, implementing this requires merging the Monte Carlo damage output with a genome annotation. We will likely discretize the genome into bins (or use gene annotations) and calculate a **damage probability per bin**. Then using a vector of accessibility scores (from ATAC-seq peaks or a binary open/closed mask), we adjust those probabilities. For example, if bin A has double the ATAC-seq signal of bin B, we might assign bin A twice the probability of getting a break (for a given energy deposition) than bin B. Calibration could be done if experimental data on break distribution (e.g. via techniques like BLESS or γ-H2AX ChIP-seq which map DSBs) are available – but even without that, qualitative weighting is possible. The result is a **genome-wide damage landscape** that is more biologically informed. It will identify "hotspots" of damage (often correlating with open chromatin, which might include gene-rich regions) and "cold regions" (like structural heterochromatin).

Finally, we must consider how Phase 1 outputs feed Phase 2 of the workflow. In Phase 2, we deal with **DNA damage response signaling** (e.g. activation of p53 network, DNA repair pathways). The **absolute number of DSBs and their complexity** will be a key driver for p53 pathway activation [82] – cells have sensors (ATM, ATR kinases) that are activated by DSBs. So from our Phase 1, we will provide as input to Phase 2 the **time-series of DSB counts** (and possibly their distribution if needed). For example, if our radiation exposure yields ~20 DSBs in a cell nucleus within a minute, we pass that to the p53 model as an initial trigger (ATM sees DSBs, triggers p53). Moreover, the *location* of breaks might influence which genes are later transcribed or if certain chromosomal rearrangements occur, but in early DNA damage signaling, it's mostly the count and complexity (which influences how long repair takes and whether some breaks are unrepairable) that matter. We might simplify by summing over the genomic distribution to get total DSBs, but keep the distribution for later linking to longer-term effects like mutation accumulation or specific gene hits. Another Phase 2 link is **damage clustering and repair pathway choice**: our model outputs whether breaks are isolated or clustered. Phase 2 can use that to decide if repair is likely to be error-free (homologous recombination for an isolated clean DSB) or error-prone (alternative end-joining for a messy cluster). Thus, Phase 1 provides the **initial conditions for cellular response models**. By building Phase 1 on sound physical and mathematical principles (cosmic ray physics, Monte Carlo transport, diffusion kinetics) and grounding it in literature (NASA standards for radiation, radiobiology textbooks for damage yields, and up-to-date research on chromatin effects), we ensure that the input to downstream biological modeling is credible and well-understood. The narrative we've constructed – from cosmic ray origins, through energy deposition, to molecular lesions – establishes a *chronological and conceptual foundation* for tackling how those lesions propagate through cellular networks in subsequent phases.

References (Literature & Tools): The journey above has drawn on key sources that serve different purposes. Foundational textbooks and reviews (e.g. ICRU reports [83], radiobiology reviews [56], Ward's classic ideas on clustered damage [52]) provide the *conceptual prerequisites*. They teach us what LET means, why clustered DNA damage is important, how diffusion of radicals works, etc. On the other hand, implementation-guiding materials include NASA's technical reports and tools documentation (OLTARIS user guide [11] [12], space radiation transport papers [30] [33]) which give concrete equations, models, and data for performing the calculations. The Julia package documentation and SciML resources [47] [48] are also implementation guides, showing what kinds of equations can be solved and how. By synthesizing these sources, we ensure Phase 1 of the project stands on a solid interdisciplinary footing – embracing physics (cosmic ray and nuclear interaction theory), *stochastic processes* (Poisson arrivals, Markov chains for transport, Brownian motion for diffusion), *chemistry* (radical reactions), and *biology* (DNA damage and chromatin structure). This comprehensive understanding is not only academically satisfying but practically necessary for the ambitious goal: simulating how space radiation leads to biological effects from the molecular scale upward. With Phase 1's framework in place, we are prepared to move into Phase 2, where the cell's dynamic responses to the computed DNA damage can be modeled with a high degree of realism and mechanistic insight.

---

[1] NASA's Cosmicopia -- Cosmic Rays -- Galactic Cosmic Rays
https://cosmicopia.gsfc.nasa.gov/gcr.html

[2] [3] [4] [5] [9] [10] [30] [31] [32] [33] [34] [35] [36] [37] [38] ntrs.nasa.gov
https://ntrs.nasa.gov/api/citations/20200002751/downloads/20200002751.pdf

[6] [13] [14] [61] [62] [69] [70] [82] workflow-condensed.md
file://file-EtuWvyqCBk1UvUpcb6CQUL

[7] workflow.md
file://file-V1XRtTJrtqk3HT2hkJrakQ

8 11 12 oltaris.larc.nasa.gov
https://oltaris.larc.nasa.gov/static_pages/OLTARIS_TP_216722/sendfile

15 16 NASA's first ground-based Galactic Cosmic Ray Simulator: Enabling a new era in space radiobiology research | PLOS Biology
https://journals.plos.org/plosbiology/article/figure?id=10.1371/journal.pbio.3000669.g003

17 20 65 66 ntrs.nasa.gov
https://ntrs.nasa.gov/api/citations/20140000785/downloads/20140000785.pdf

18 19 27 28 29 39 40 41 Welcome to the CRÈME site — CREME-MC site
https://creme.isde.vanderbilt.edu/

21 22 NASA's first ground-based Galactic Cosmic Ray Simulator: Enabling a new era in space radiobiology research | PLOS Biology
https://journals.plos.org/plosbiology/article/figure?id=10.1371/journal.pbio.3000669.g002

23 24 25 83 pdg.lbl.gov
https://pdg.lbl.gov/2023/reviews/rpp2023-rev-radioactivity.pdf

26 46 [PDF] Calculation of photon energy deposition kernels and electron dose ...
https://physics.carleton.ca/~drogers/pubs/papers/Ma05.pdf

42 43 63 64 76 77 78 79 Simulation of Radiation-Induced DNA Damage and Protection by Histones Using the Code RITRACKS
https://www.mdpi.com/2673-6284/13/2/17

44 45 NASA's first ground-based Galactic Cosmic Ray Simulator: Enabling a new era in space radiobiology research | PLOS Biology
https://journals.plos.org/plosbiology/article/figures?id=10.1371/journal.pbio.3000669

47 48 67 68 GitHub - SciML/DifferentialEquations.jl: Multi-language suite for high-performance solvers of differential equations and scientific machine learning (SciML) components. Ordinary differential equations (ODEs), stochastic differential equations (SDEs), delay differential equations (DDEs), differential-algebraic equations (DAEs), and more in Julia.
https://github.com/SciML/DifferentialEquations.jl

49 Delayed repair of radiation induced clustered DNA damage
https://pmc.ncbi.nlm.nih.gov/articles/PMC3112496/

50 51 54 Radiation-induced clustered DNA lesions: Repair and mutagenesis
https://www.sciencedirect.com/science/article/abs/pii/S0891584916310954

52 Formation of Clustered DNA Damage after High-LET Irradiation
https://academic.oup.com/jrr/article-pdf/49/3/203/2711250/jrr-49-203.pdf

53 Clustered DNA Damage and its Complexity: Tracking the History
https://meridian.allenpress.com/radiation-research/article/202/2/385/501602/Clustered-DNA-Damage-and-its-Complexity-Tracking

55 Quantifying DNA damage induced by ionizing radiation and ...
https://www.sciencedirect.com/science/article/pii/S1936523320303144

56 57 58 DNA Damage Clustering after Ionizing Radiation and Consequences in the Processing of Chromatin Breaks
https://www.mdpi.com/1420-3049/27/5/1540

59 60 A Monte Carlo simulation surveying direct DNA damage using ...
https://www.sciencedirect.com/science/article/abs/pii/S0969806X25004323

71 Chromatin structure influences the sensitivity of DNA to γ-radiation
https://www.sciencedirect.com/science/article/pii/S0167488908002590

72 75 Ionizing radiation damage and repair from 3D-genomic perspective
https://www.cell.com/trends/genetics/pdf/S0168-9525(22)00181-0.pdf

73 74 81 Ionizing radiation damage and repair from 3D-genomic perspective
https://www.sciencedirect.com/science/article/pii/S0168952522001810

80 Effect of Chromatin Structure on the Extent and Distribution of DNA Double Strand Breaks Produced by Ionizing Radiation; Comparative Study of hESC and Differentiated Cells Lines
https://www.mdpi.com/1422-0067/17/1/58