DLR Proposal (2022–2025)

(joint proposal of JMU/JGU):

"Development of the data analysis software for the Compton Spectrometer and Imager (COSI)"

Status of the Würzburg working packages

Thomas Siegert

University of Würzburg

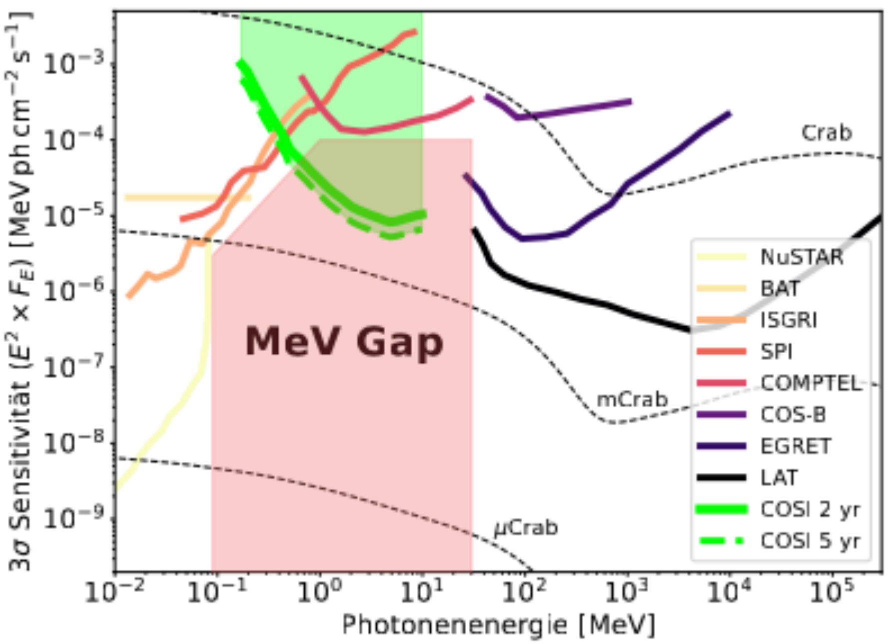
January 30/31 2025, COSI Germany Meeting, Mainz

Scientific and Technical Goals

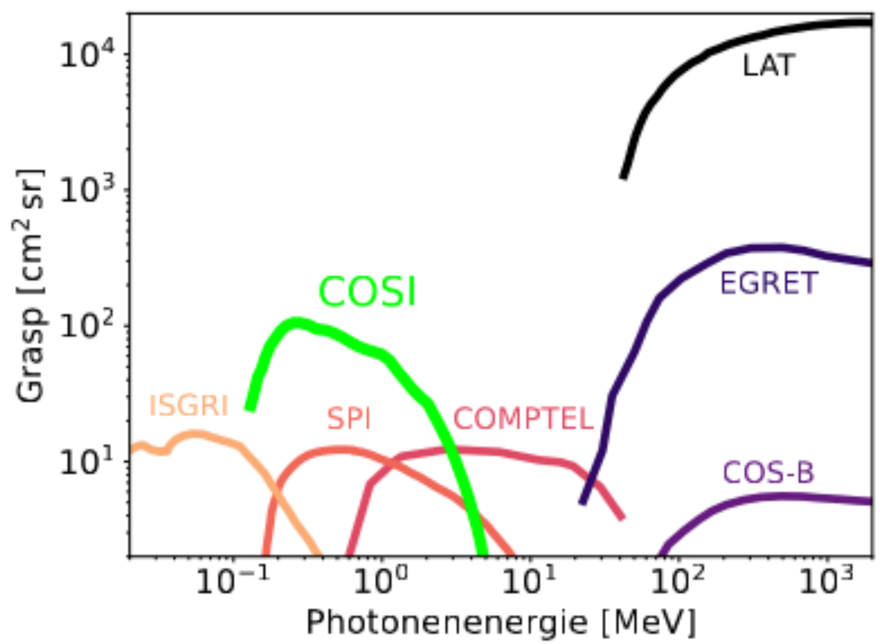
- There is (2022) no stand-alone software for the low- to high-level data analysis for Compton telescopes — now (2025) there is !
- The **primary objectives** of our (i.e. JMU/JGU) working group are:
  1. **“From detector signals over interaction points and energy deposits to photons with energies and scattering angles”**
     ➡ Compton telescopes are no ‘cameras’, we need to reconstruct the events **(1) classically**, and **(2) by machine learning**
  2. **“From photons with energies and scattering angles to sky coordinates and spectra”**
     ➡ After reconstruction, we need to handle the new data space through **(1) image deconvolution**, and **(2) forward-folding**
  3. **“Identification and modelling of background photons”**
     ➡ We measure mostly background / underground, which we want to handle by **(1) empirical/heuristic fits**, **(2) detailed MC simulations**, and **(3) statistical marginalisation**

State of the Art / Problems

Continuum sensitivity of MeV instruments



Grasp effective area × field of view



Definition of Work Packages

University / Group	Work Packages (WPs)		
Mainz	Event reconstruction with machine learning **WP3.1, 3.2**	GRB (& transient) localisation **WP3.3**	Background simulations **WP2.1**
Würzburg	Development of data analysis framework (COSIpy) **WP1.1**	Image reconstruction **WP1.2**	Background modelling **WP2.2, 2.3**

WP1.1 Software Framework COSIpy

- **Definition of likelihood function** of point sources and diffuse emission
  - 2022: Persistent sources, transients, marginalized background, …
  - **2025: Single sources (point sources of diffuse), GRBs**
  - **Missing: long-term transients (novae, solar flares, …), fits with multiple sources**
- **Making the response function usable**
  - 2022: $R(E_i,Z,A \rightarrow E_f,\varphi,\psi,x)$ for spectra and imaging; $R(E_i,Z,A,\Pi,\alpha \rightarrow E_f,\varphi,\psi,x)$ for polarisation
  - **2025: Spectra and imaging done (Hiroki, paper in review)**
  - **Missing: Polarisation likelihood (currently only ASAD method; Eliza)**
- **Efficient convolution into COSI data space**
  - 2022: Need to re-apply response every time step because we have a moving telescope; ~15 s for ~1 deg resolution
  - Re-use of already-calculated "pixels" for diffuse emission?
  - **2025: solved to large extent (see Hiroki's talk) for imaging**
  - **Missing: fast polarisation response for spectral-spatial-polarisation fits**

WP1.1 Software Framework COSIpy

- **Validation of methods**
  - 2022: Unit test, residuals, etc.
  - **2025: Validation via source injector**
  - **Missing: Unit tests, residuals in different dimensions**
- **Including polarization**
  - 2022: Low priority for now, will be very difficult for persistent sources
  - **2025: Only few people working on polarization (Eliza with ASAD method)**
  - **Missing: polarization for transients and persistent sources (likelihood fits)**
- **Extension for large data set** (high-performance computing)
  - 2022: Might be impossible for end-users to analyse all the data → COSI-team generated end-user products with HPC
  - **2025: Might still be impossible to do that**
  - **Missing: Option to restrict datasets to regions of interest (response adaption, etc.)**

WP1.2 Imaging with COSIpy

- **General deconvolution of 5D data space into 2D image**
  - 2022: Spectro-spatial deconvolution: proper application of transpose response matrix with energy information (assumed spectrum?)
  - **2025: Done (Hiroki).**
- **Inclusion of** fixed or flexible **background model** in deconvolution
  - 2022: Imaging requires a set of possible options for background (fitting each imaging iteration, restriction to positive fluxes, etc.)
  - **2025: Fixed template with 1 scaling parameter (MC simulation, Savitri); continuum background method (filling gaps in CDS, Chris); line background with adjacent energy bins (Saurabh)**
  - **Missing: Time variability; multiply background parameters; Options in COSIpy (with defaults) for background handling**
- **Developing/refining of the Richardson-Lucy method**
  - 2022: Class in COSIpy for RL algorithm
  - **2025: Done (Hiroki)**

WP1.2 Imaging with COSIpy

- **Developing/refining of the Multi-Resolution Expectation Maximization (MREM) method**
  - 2022: Class in COSIpy for MREM algorithm
  - **2025: Can be implemented within Hiroki's imaging framework**
- **Developing/refining of the Maximum Entropy method**
  - 2022: Class in COSIpy for MaxEnt algorithm
  - **2025: Can be implemented within Hiroki's imaging framework**
- **Tests of Information Field Theory**
  - 2022: If enough time (or believed useful), test class for IFT (Torsten Ensslin, MPA)
  - **2025: Not sure if we want that…**

WP2.2 Background in Orbit

- **Simulation of known background components from previous instruments**
  - 2022: Components-wise models, including protons, neutrons, activation, SAA, solar flares, atmosphere, photons, CXB, …; Shared efforts between Mainz and Würzburg
  - **2025: Model by Savitri based on Cumani+2019 and improvements**
  - **Missing: Gamma-ray line background? Application to balloon (Savitri)**
- **Re-evaluation in COSI orbit**
  - 2022: Tests for specified COSI orbit, with correct rocking of COSI; Requires (also for DC simulations) correct orientation file for two year mission
  - **2025:** Done? (Saurabh, orbit file)
- **Parametrizing individual models and implementation into COSIpy**
  - 2022: MEGAlib simulations packaged in a COSIpy-readable format; parametrization of different components (in addition to amplitude)
  - **2025: Different components available for different rigidities, etc. (Savitri)**
  - **Missing: Implementation of these models in COSIpy**

WP2.3 Background Marginalisation

- **Investigation of constant dimensions via simulations**
  - Background **amplitude** constant during orbit?
  - Masking out South Atlantic Anomaly?
  - Background **spectrum** constant?
  - Background **scattering** angle distribution constant?
- **Additional likelihood function with background marginalisation**
  - Explicitly writing down the math and statistics how this is down
  - Tests with simulations if background can efficiently be marginalised for point sources, diffuse emission, only transients (lightcurve), …
  - Efficient calculation of likelihood function if math too heavy (approximations)
- **Implementation in COSIpy**
  - Optional background handling via marginalisation, probably with caveats and warnings

WP2.3 Background Marginalisation

- **Investigation of constant dimensions via simulations**
  - 2022: Background **amplitude** constant during orbit?
  - **2025: Within some time?**
  - 2022: Masking out South Atlantic Anomaly?
  - **2025: Probably required (Savitri)**
  - 2022: Background **spectrum** constant?
  - **2025: Probably not a lot?**
  - 2022: Background **scattering** angle distribution constant?
  - **2025: Missing?**
- **Additional likelihood function with background marginalisation**
  - 2022: Explicitly writing down the math and statistics how this is done; Tests with simulations if background can efficiently be marginalised for point sources, diffuse emission, only transients (lightcurve), …; Efficient calculation of likelihood function if math too heavy (approximations)
  - **2025: Missing all that (except for GRBs, Saurabh)**
- **Implementation in COSIpy**
  - **2025: Missing all that (except for GRBs, Saurabh)**

WP2.1 Background MC Simulations

- Search for instruments in comparable orbit; interpolation for COSI orbit
  - NuSTAR, Fermi/GBM, CGRO/COMPTEL&BATSE, …
  - Evaluation for high resolution spectroscopy with COSI (need gamma-ray line background)
- Using spectra for simulations of prompt and delayed emission
  - Particle simulations to estimate the bremsstrahlung component (continuum, prompt), and particle background itself
  - Estimate of anti coincidence shield efficiency
  - Secondary particles from cosmic-ray excitation (delayed lines)
  - Tertiary components (particle production in satellite, annihilation line, etc.)
- Investigation of prompt and delayed background components for artifacts, verification with balloon data

WP3.1 Event Reconstruction

- **Collecting algorithms and implementation** towards a pipeline for usage with simulated data
  - Agreement on which SMEX mass model will be used
  - Improvement of Andreas' current versions
  - Testing likelihood approaches
- **Creation of training and test data** from dedicated simulations; **implementation of model candidates**
  - Heavy simulations to obtain large data base
  - Expected simulation time: 10 months
  - Application/test of algorithms
- **Supervision of data challenges and model selection**
  - Configuration files for Chris Karwin and Andreas for new event reconstruction to be used in data challenges
- **Model optimisation and integration in analysis chains**
  - Standardising event reconstruction for use in response generation

WP3.2 Background and Event Classification

- **Collecting algorithms and implementation** towards a pipeline for usage with simulated data
  - Identification of background events from raw data
  - Application of convolutional neural network
  - Identification of “what is background”: delayed photons vs. emission from the bottom / Earth, etc.
- **Creation of training and test data** from dedicated simulations; **implementation of model candidates**
  - Same simulations used as in WP3.1
- **Supervision of data challenges and model selection**
  - DC test datasets with background events removed vs test datasets with normal procedures
- **Model optimisation and integration in analysis chains**
  - Possible implementation of a second dataset plus response for data analysis

WP3.3 GRB/Transient Localisation

- **Introduction to the topic**
  - What are GRBs and how are they simulated in MEGAlib
  - Connection to COSI GRB Science group (currently triggers and some catalog simulations already performed)
- **Collection of algorithms to test, implementation of pipeline for simulation data**
  - Testing CNNs in raw data
  - Testing CNNs in reconstructed data space
  - Trigger criteria for when/how algorithm could be applied (onboard?)
- **Creation of training and test data from catalogs** (Fermi/GBM, BATSE, etc.), implementation of reconstruction approaches
  - Simulation of sets of objects: short GRBs, long GRBs, solar flares, …
- **Supervision of data challenges**
  - Special data challenge for GRBs/transients upcoming
- **Model optimisation and integration in analysis chains**
  - Exact (slow) vs approximate (fast) localisation comparisons in DC

Background marginalisation (transients)

- **Likelihood for BG measurement:**

$$L(B_i, t_b | b_i) = \frac{(t_b b_i)^{B_i}}{B_i!} e^{-t_b b_i}$$

t_b : **off observation time**

b_i : **background rate [ph/s]**

B_i : **detected counts (off-source)**

- **Likelihood for on-source measurement:**

$$L(D_i, B_i, t_d, t_b | \theta, b_i) = \frac{[t_d(m_i(\theta) + b_i)]^{D_i}}{D_i!} e^{-t_d(m_i(\theta)+b_i)} \times \frac{(t_b b_i)^{B_i}}{B_i!} e^{-t_b b_i}$$

- **BG marginalisation:**

$$L(D_i, B_i, t_d, t_b | \theta) = \int_0^\infty db_i \, L(D_i, B_i, t_d, t_b | \theta, b_i)$$

t_d : **on observation time**

m_i : **model rate [ph/s]**

θ : **model parameters**

D_i : **detected counts (on-source)**

- **Profile likelihood:**

$$\frac{d}{db_i} L(D_i, B_i, t_d, t_b | \theta, b_i) = 0$$

$$\Longrightarrow b_{i,max}(\theta) = \frac{1}{2(t_b + t_d)} \left(B_i + D_i m_i(\theta)(t_b + t_d) + \sqrt{[B_i + D_i + m_i(\theta)(t_b + t_d)]^2 + 4m_i(\theta)D_i(t_d - t_b)} \right)$$

Plugging back into likelihood for on-source measurement!

Background marginalisation (transients)

Plugging back into likelihood for on-source measurement!

- **Profile likelihood (continued):**

$$\Longrightarrow b_{i,max}(\theta) = \frac{1}{2(t_b + t_d)}\left(B_i + D_i m_i(\theta)(t_b + t_d) + \sqrt{[B_i + D_i + m_i(\theta)(t_b + t_d)]^2 + 4m_i(\theta)D_i(t_d - t_b)}\right)$$

- **Likelihood for on-source measurement:**

$$L(D_i, B_i, t_d, t_b \,|\, \theta, b_i) = \frac{[t_d(m_i(\theta) + b_i)]^{D_i}}{D_i!} e^{-t_d(m_i(\theta)+b_i)} \times \frac{(t_b b_i)^{B_i}}{B_i!} e^{-t_b b_i}$$

t_b : off observation time

b_i : bac[illegible]te [ph/s]

B_i : detected counts (off-source)

t_d : on observation time

m_i : model rate [ph/s]

θ : model parameters

D_i : detected counts (on-source)

$$\ln L(D_i, B_i, t_d, t_b \,|\, \theta) =$$

$$= D_i \ln[t_d(m_i(\theta) + b_{i,max}(\theta)] - t_d(m_i(\theta) + b_{i,max}(\theta)) - \ln D_i! +$$

$$+ B_i \ln[t_b b_{i,max}(\theta)] - t_b b_{i,max}(\theta) - \ln B_i!$$

- **Final step:** $\frac{d}{d\theta} \ln L(D_i, B_i, t_d, t_b \,|\, \theta) = 0 \Longrightarrow \theta = \ldots$