Real-time custom computing for adaptive radiotherapy

Wayne Luk
Department of Computing
Imperial College

CCAP Plenary Meeting 7
16 December 2020

Acknowledgement: U. Oelfke, A. Wetscherek (ICR), N. Voss, M. Barbone (Imperial)

Cancer treatment

- radiotherapy – common treatment for cancer
  - kill cancer using high radioactive dose
  - avoid damage of surrounding healthy tissue

Cancer treatment

- radiotherapy – common treatment for cancer
  - kill cancer using high radioactive dose
  - avoid damage of surrounding healthy tissue

- current common practice
  - imaging: locate cancer region in MRI machine
  - planning: organise the radiation dose
  - therapy: in medical linear accelerator (Linac)

Image source: www.radiologyinfo.org

Adaptive radiotherapy

- organ movement: what you see is *not* what is there
  - between imaging and therapy: different dates
  - during therapy: e.g. breathing, bladder filling
  - radiate healthy tissues rather than cancer!

Adaptive radiotherapy

- organ movement: what you see is *not* what is there
  - between imaging and therapy: different dates
  - during therapy: e.g. breathing, bladder filling
  - radiate healthy tissues rather than cancer!
- MR Linac: imaging + therapy
  - UK's first: Inst. of Cancer Research
  - adapt to organ movement in real-time
- benefits
  - more targeted treatment: fewer sessions
  - less damage: surrounding healthy tissues

Image source: https://www.icr.ac.uk

Adaptive radiotherapy

- challenge: all 3 steps together in real time!
  - imaging
  - planning
  - therapy
- MR Linac: imaging + therapy
  - UK's first: Inst. of Cancer Research
  - adapt to organ movement in real-time
- benefits
  - more targeted treatment: fewer sessions
  - less damage: surrounding healthy tissues

Image source: https://www.icr.ac.uk

Vision: real-time image analysis + treatment planning

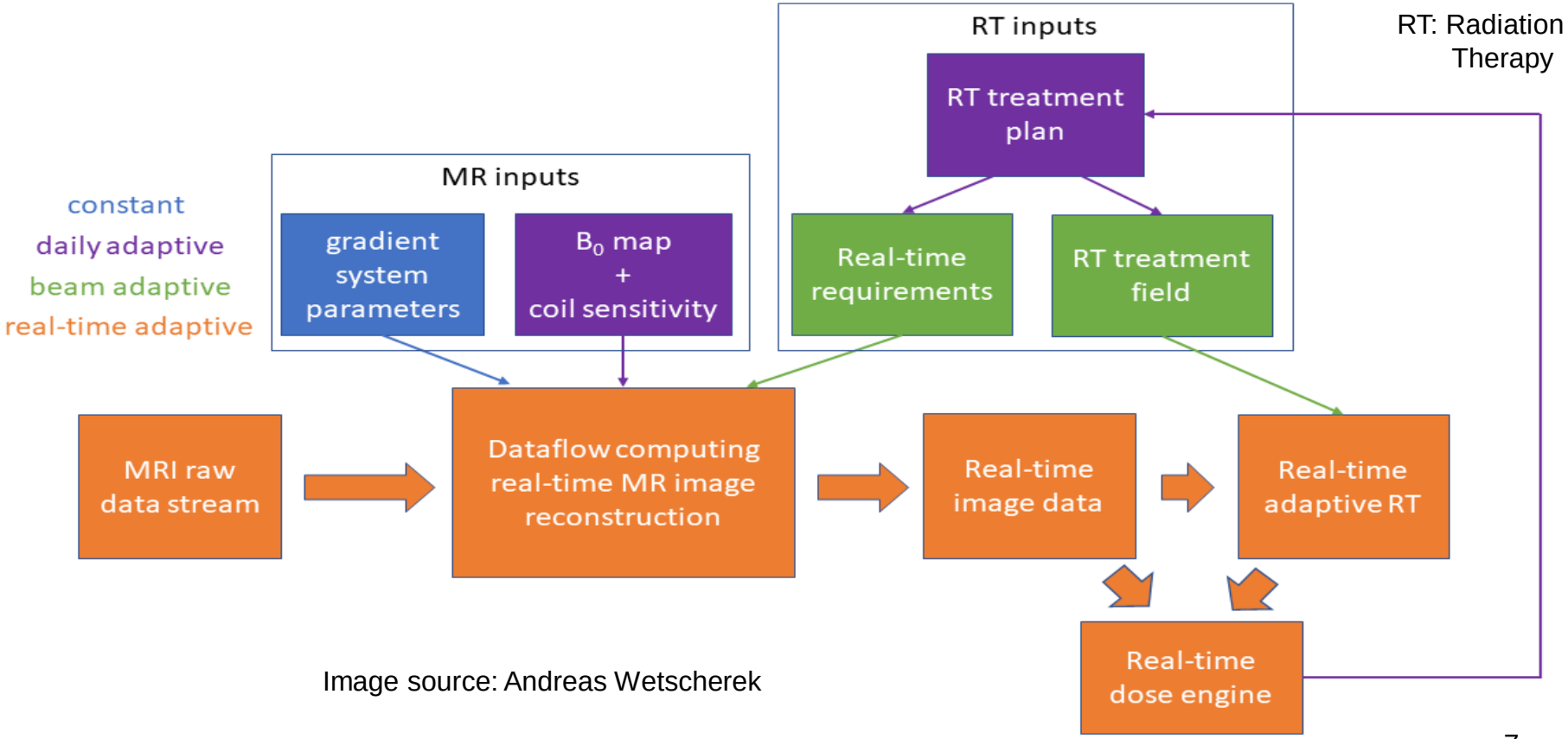


Image source: Andreas Wetscherek

Vision: real-time image analysis + treatment planning

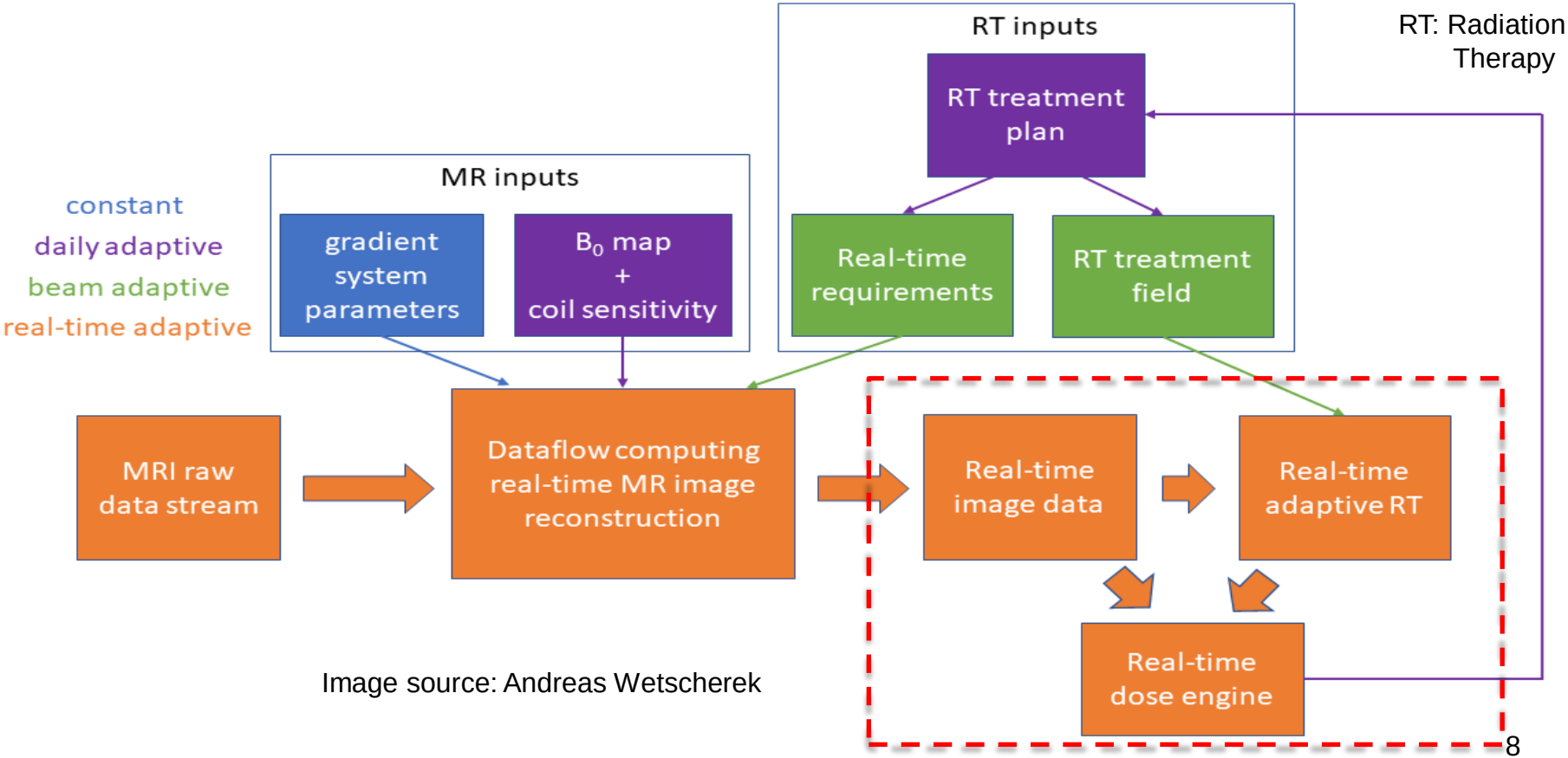


Image source: Andreas Wetscherek

Treatment planning: Monte Carlo dose simulation

- dose effect: simulate 100 million particles within 1 second
- patient body: represented as cube, discretised in voxels
- particles are sent into a cube
  - particle: energy + fuel => travel distance + direction vector

Treatment planning: Monte Carlo dose simulation

- dose effect: simulate 100 million particles within 1 second
- patient body: represented as cube, discretised in voxels
- particles are sent into a cube
  - particle: energy + fuel => travel distance + direction vector
- when fuel runs out
  - select particle interaction
  - aggregate voxel dose
  - produce new particle with energy, fuel, random direction
- when energy runs out, particle gets absorbed / removed

Custom computing: what is it?

- conventional computing: fit program to processor

Custom computing: what is it?

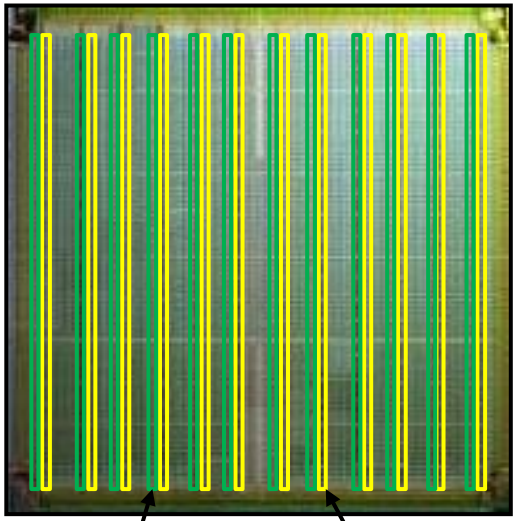
- conventional computing: fit program to processor

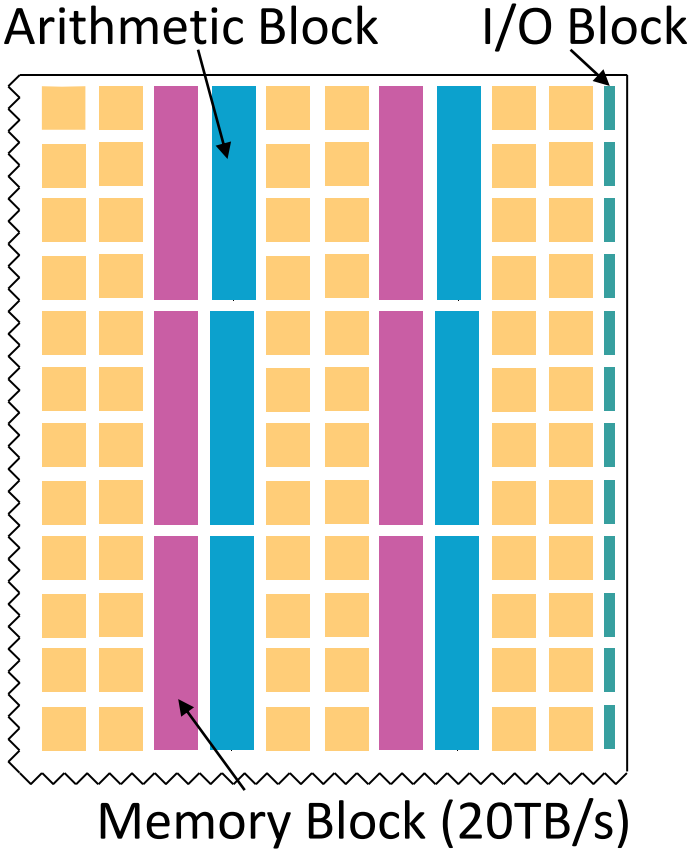
- custom computing: fit processor to program

- customisation: field programmable gate array (FPGA)

FPGA: Field Programmable Gate Array

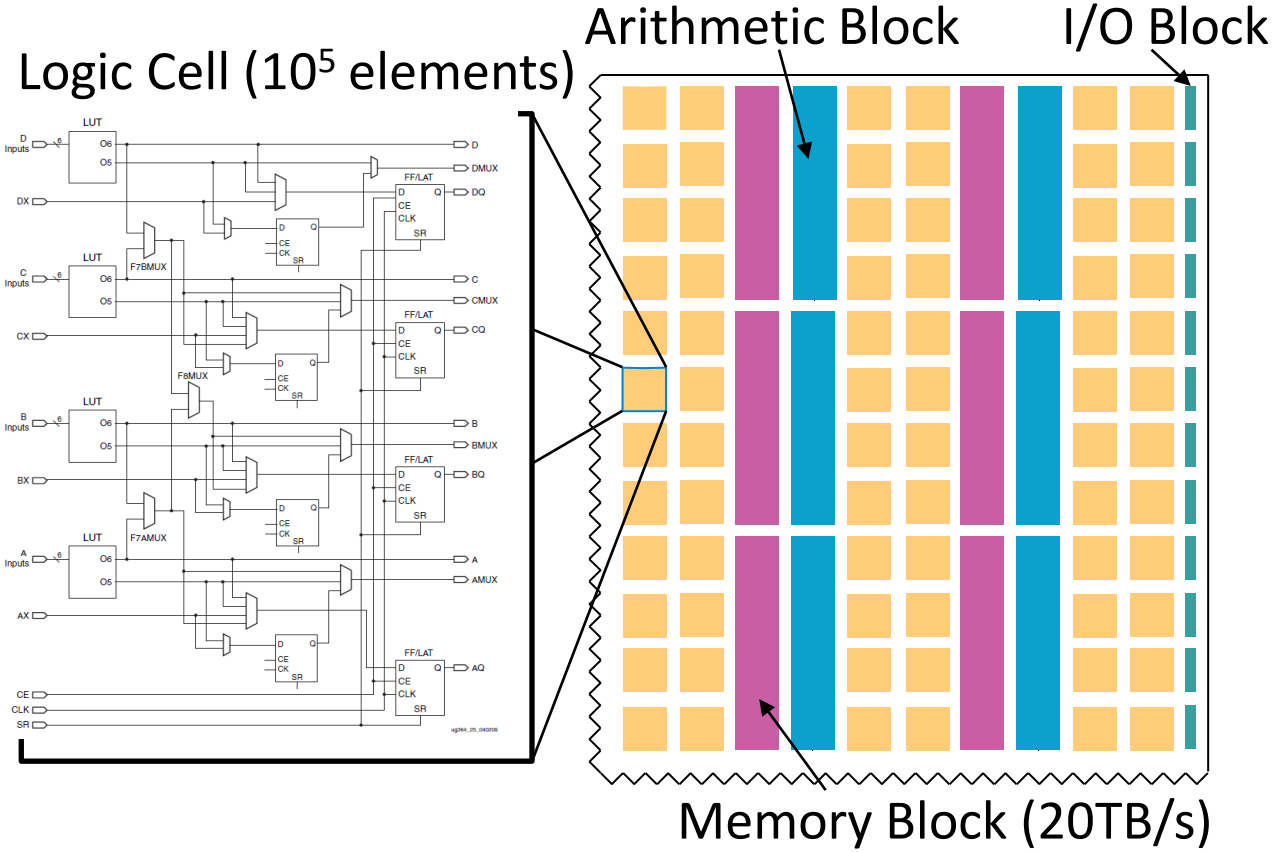
Xilinx Virtex-6 FPGA





source: Maxeler

FPGA: Field Programmable Gate Array



source: Maxeler

FPGA-based custom computing: now main-stream

www.top500.org/news/microsoft-goes-all-in-for-fpgas-to-build-out-cloud-based-ai/

Microsoft Goes All in for FPGAs to Build Out AI Cloud

Michael Feldman | September 27, 2016 08:42 CEST

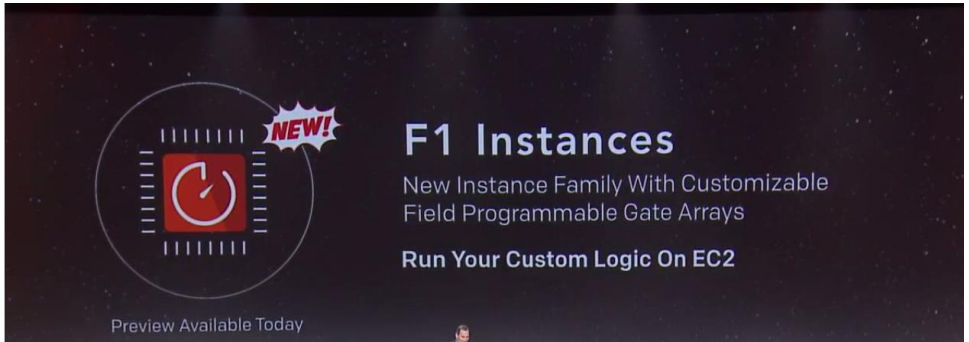
Software giant bets the (server) farm on reconfigurable computing

Microsoft has revealed that Altera FPGAs have been installed across every Azure cloud server, creating what the company is calling "the world's first AI supercomputer." T spans 15 countries and represents an aggregate performance of more thar announcement was made by Microsoft CEO Satya Nadella and engineer D the opening keynote at the Ignite Conference in Atlanta.

Amazon EC2 F1 Instances

Run Custom FPGAs in the AWS Cloud

Amazon EC2 F1 is a compute instance with field programmable gate arrays (FPGAs) that you can program to create custom hardware accelerations for your application.



aws.amazon.com/ec2/instance-types/f1/

Also AliCloud, HuaweiCloud,,,

Design challenges

- the path through the patient cube is random
- patient cube size: much bigger than FPGA on-chip memory
  - decompose big cube into smaller sub domains

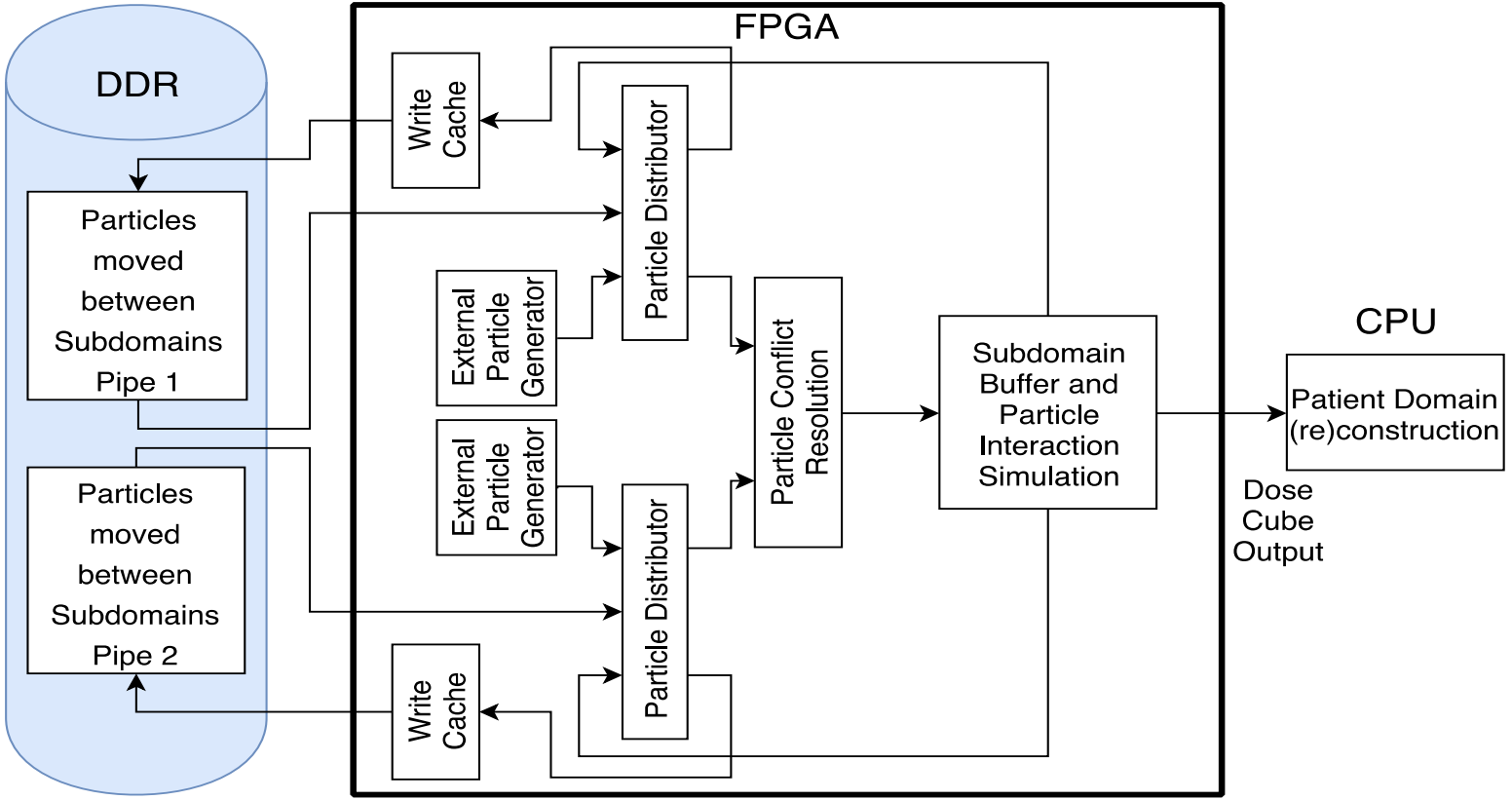
Design challenges

- the path through the patient cube is random
- patient cube size: much bigger than FPGA on-chip memory
  - decompose big cube into smaller sub domains
- particles are processed in a loop
  - final result: needed to start next iteration
  - data dependency between iterations: affect pipelined execution
  - reorder computations to allow pipeline parallelism

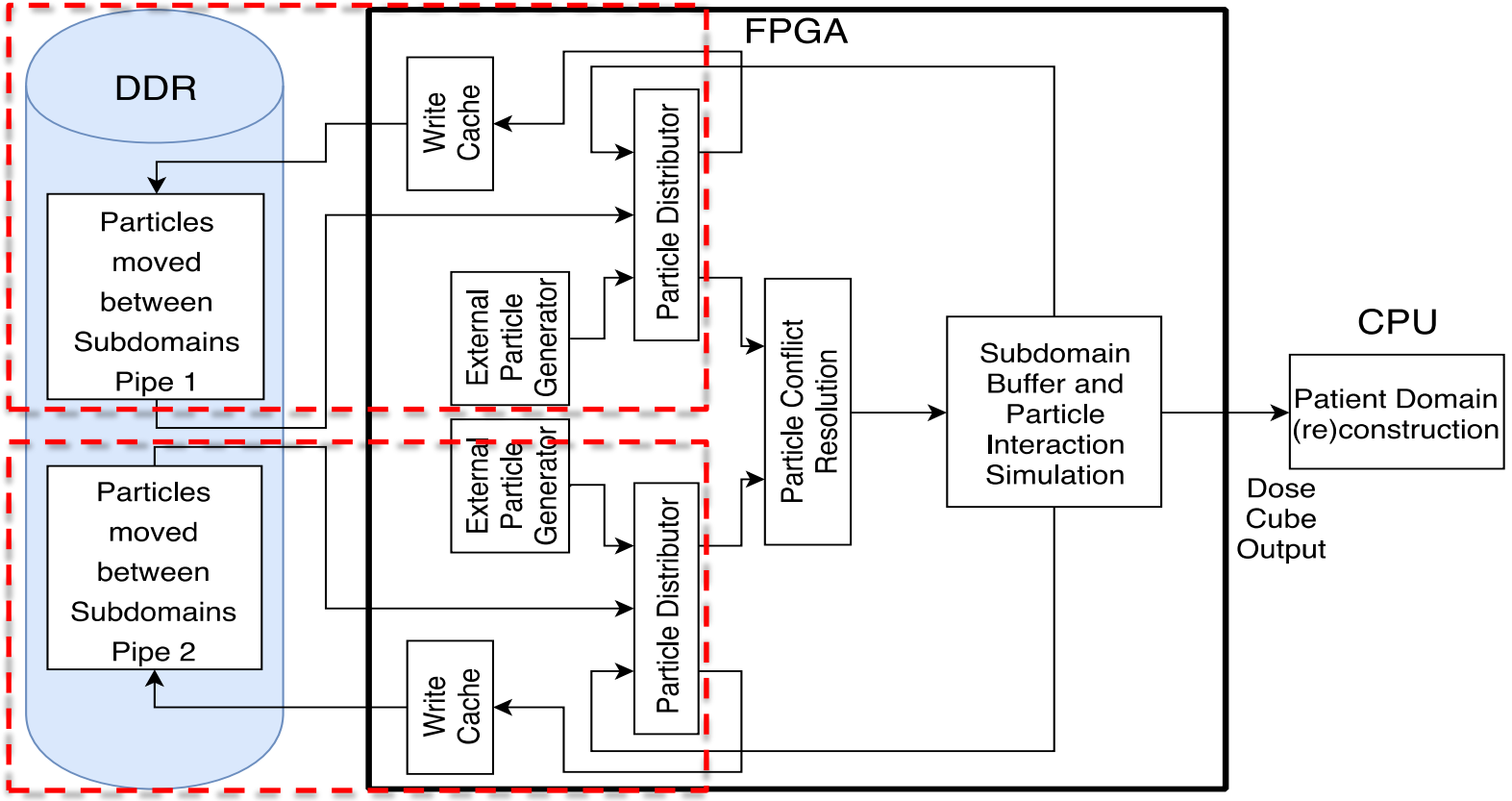
Programming model: data flow

- describe computation as a directed graph
- data get processed while traversing through the nodes
- Maxeler MaxCompiler: maps onto dataflow engines (DFEs)
- current generation DFE: MAX5C
  - based on VU9P FPGA from Xilinx
  - 7,000 multipliers
  - 40 MB on chip memory
  - 48 GB DDR4 memory
  - 50 GB/s memory bandwidth

Dose simulation architecture



Dose simulation architecture



Evaluation

- Maxeler MAX5C DFEs with MaxCompiler and Vivado
- CPU platform 2x Xeon E5-2643 v4: 6 cores each @ 3,4 GHz

Cards	FPGA Time [ms]	Total Time [ms]
1	3,267	3,435
2	1,173	1,342
3	810	988

- 4x speedup compared to CPU
- 8x speedup compared to GPU
  - random memory access + SIMD interactions problem

Future: accelerate Lhara simulation?

Simulation Codes for Stage 1 Tracking

Smilei

- Particle-In-Cell (PIC) code for plasma simulation.
- Generate distribution of particles.

BDSIM
Beam Delivery Simulation

- Uses Geant4 toolkit to simulate transport and particle-matter interactions.
- Propagates beam from Smilei through beam line.

GPT
General Particle Tracer

- 3D particle tracking with various 2D and 3D space charge models.
- Include space charge effects in distribution.

Alternative model of the electron cloud

- Full simulation of plasma and proton beam using the PIC code is computationally expensive
- **Aims:**
  - Investigate the parameter space for the plasma
  - Understand the origin of the images taken during the beam test
  - Reproduce the main features
- **Model** the instability:
  - Idealised cylindrical electron cloud which rotates
  - Generate 4D electric field map for the plasma
  - Track the pencil beams using BDSIM[5]

[5] https://doi.org/10.1016/j.cpc.2020.107200

Future: accelerate Lhara simulation and optimisation?

Simulation Codes for Stage 1 Tracking

Smilei

- Particle-In-Cell (PIC) code for plasma simulation.
- Generate distribution of particles.

BDSIM
Beam Delivery Simulation

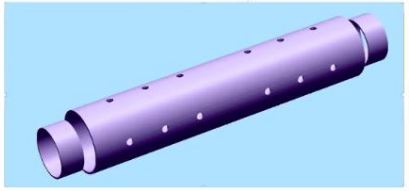
- Uses Geant4 toolkit to simulate transport and particle-matter interactions.
- Propagates beam from Smilei through beam line.

GPT
General Particle Tracer

- 3D particle tracking with various 2D and 3D space charge models.
- Include space charge effects in distribution.

Impact on new design of the lens

Old design



Simulation of the new lens

- electron loss near the central axis
- no instability develops on timescale of 10 μs

Next steps in the design

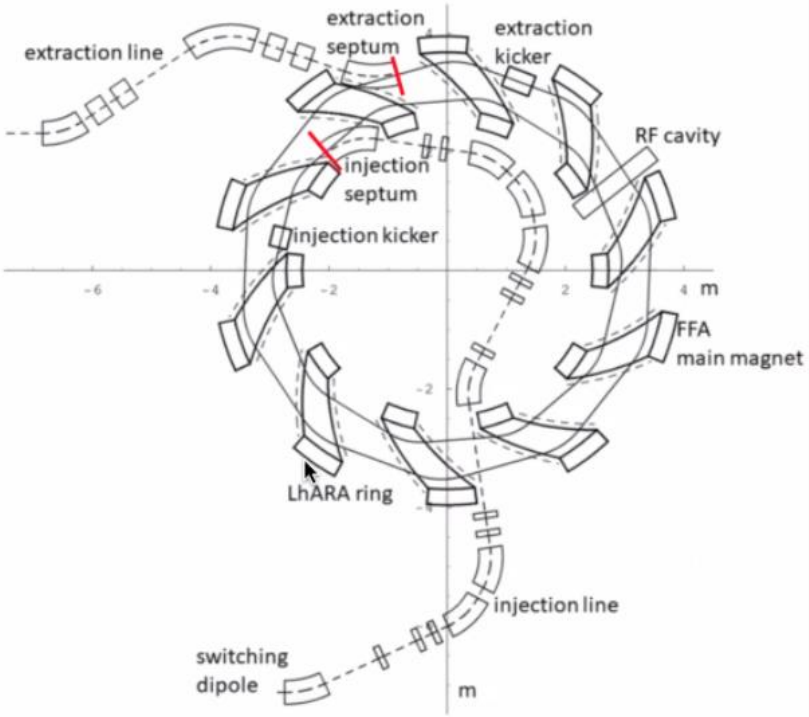
- active filling of the lens
- ensure $E \times B$ rotation of the plasma
- damping mechanism to ensure uniform filling

Future: accelerate Lhara development tool chain?

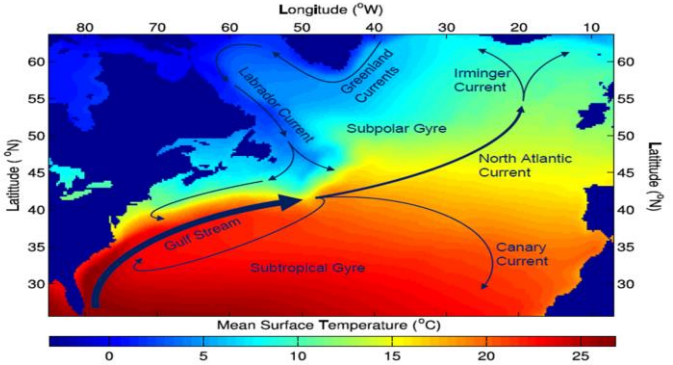
OPAL, BDSIM & GPT Toolchain

- (BDSIM,GPT) & OPAL on separate machines
  - No single self-contained code

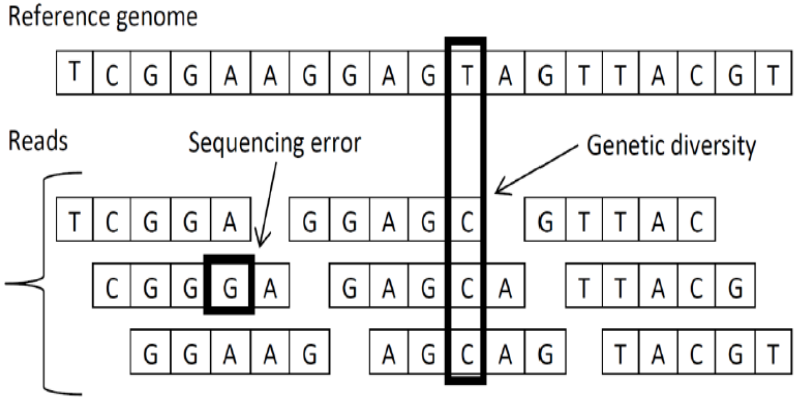
- Defined interface planes between codes & exchange particle coordinates



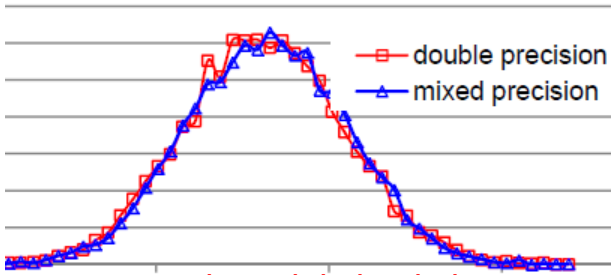
Other applications



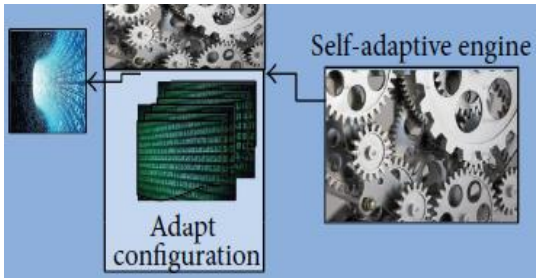
Climate modelling:
13 times faster, 23 times lower op/W
(with Oxford, Eur. Centre for M. Weather Forecasts)



Genomic data analysis:
88 times faster, 3 times less energy
(with Chinese University of Hong Kong)



Financial simulation:
163 times faster, 170 times less energy
(with J.P. Morgan, Bank of America, Jump Trading)



Avionics monitoring:
10 times faster
(with Airbus)

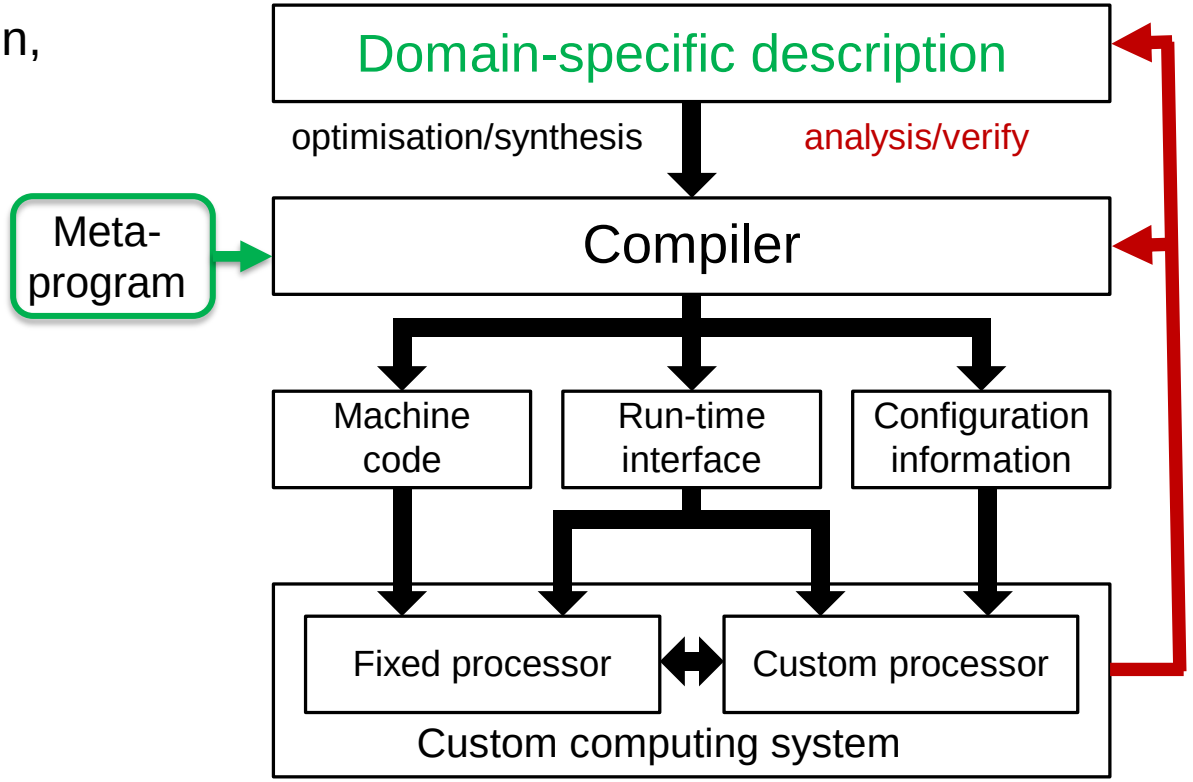
Future: design by domain specialists

design space exploration, goals and constraints

partition, compile, analysis, verify

system-specific programming interface

system-specific adaptation: clouds to edge devices



Vision: real-time image analysis + treatment planning

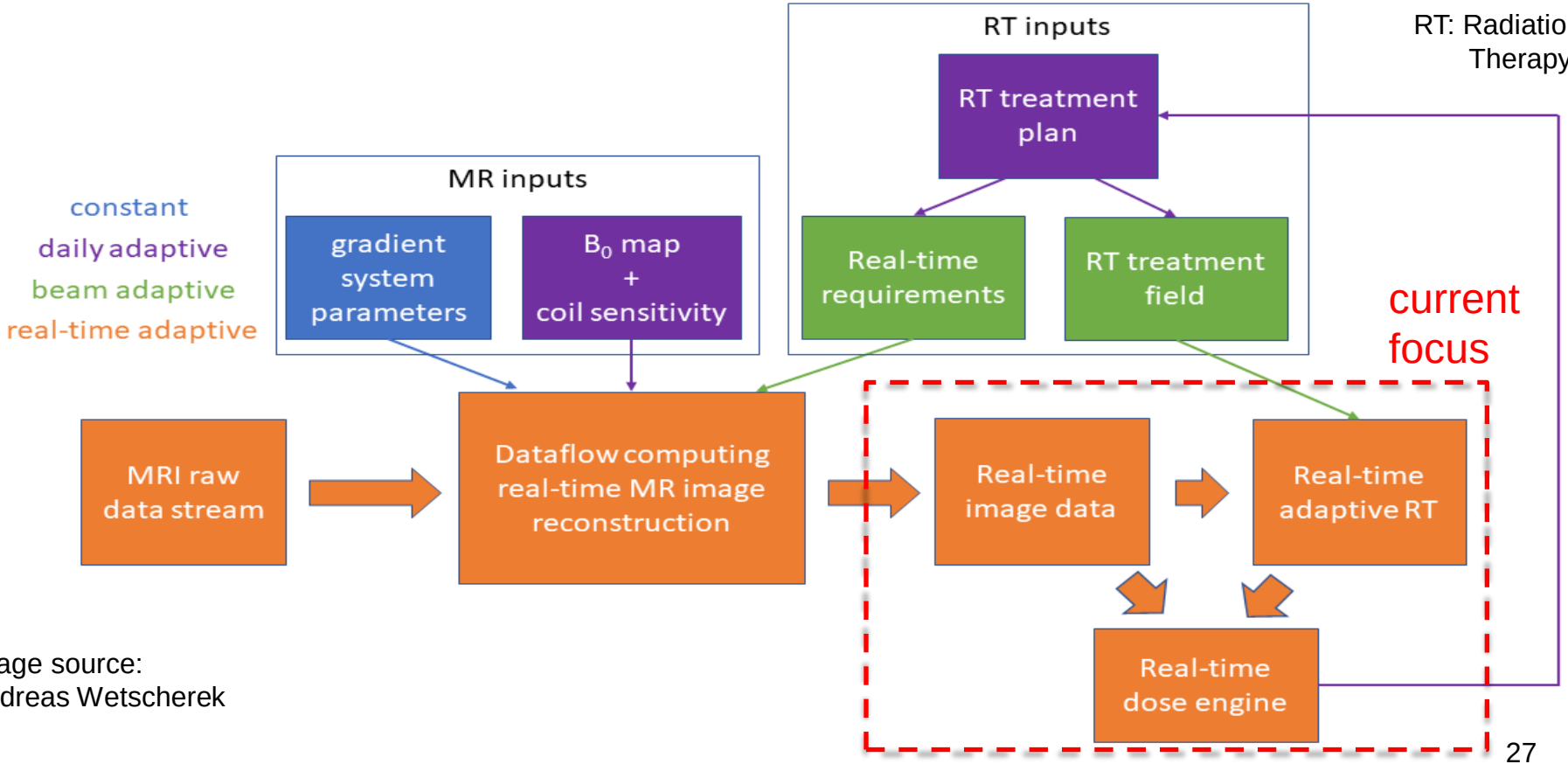


Image source:
Andreas Wetscherek

Vision: real-time image analysis + treatment planning

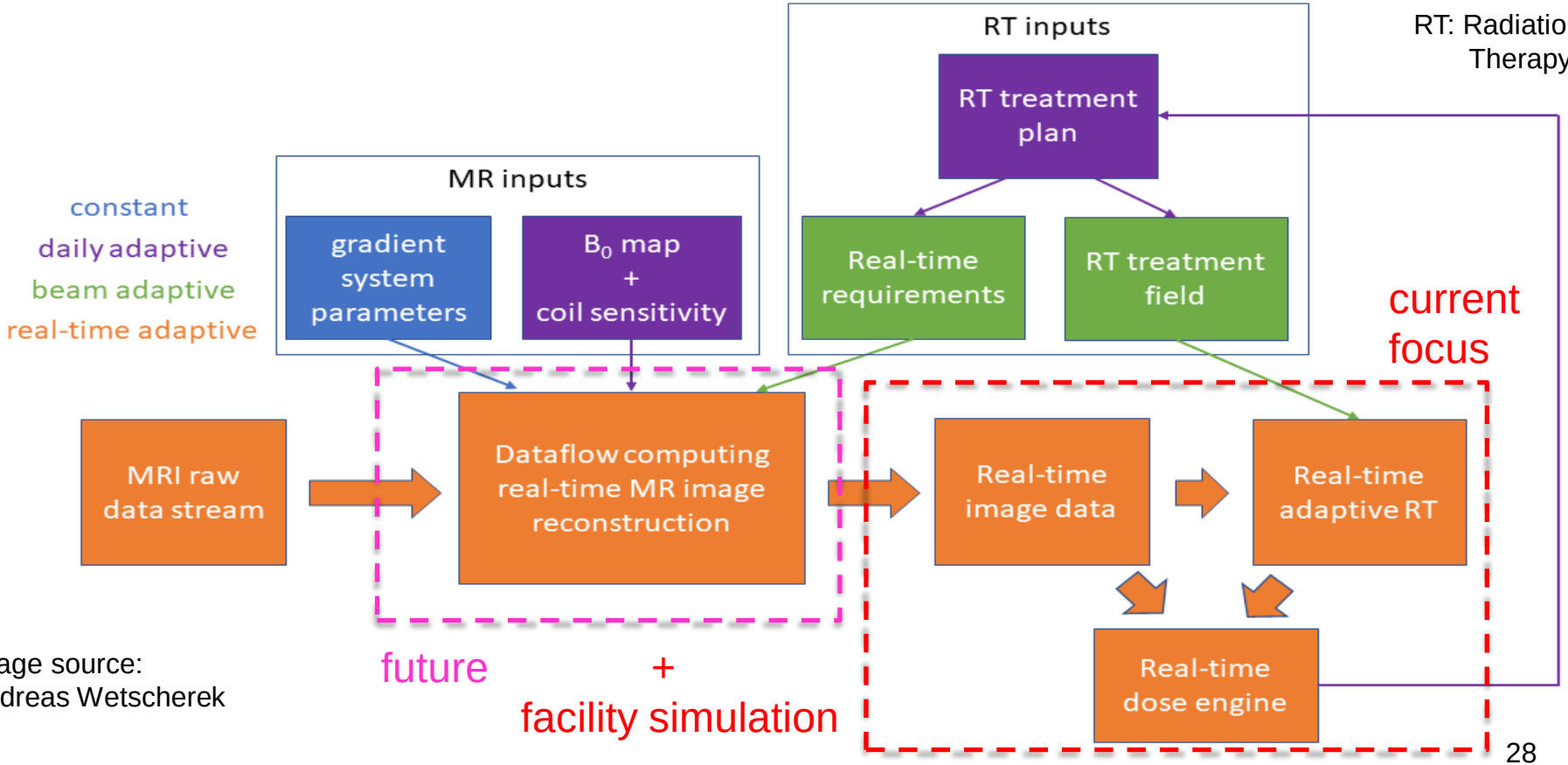


Image source:
Andreas Wetscherek