EFCL SUMMER SCHOOL

on Open Source IC Design and Computer Architectures

Zurich

June 3rd - 7h, 2024

ETH zürich

Table of Content

Program Overview

	Monday 3rd of June	Tuesday 4th of June	Wednesday 5th of June	Thursday 6th of June	Friday 7th of June
08:00 - 08:50	Registration				
08:50 - 09:00	Welcome *(ETF E1)*				
09:00 - 10:00	Keynote #1 *(ETF E1)*	Keynote #2 *(ETF E1)*	Keynote #3 *(ETF E1)*	Keynote #4 *(ETF E1)*	Keynote #5 *(ETF E1)*
10:00 - 10:30	Coffee Break *(ETZ Foyer)*	Coffee Break *(ETZ Foyer)*	Coffee Break *(ETZ Foyer)*	Coffee Break *(ETZ Foyer)*	Coffee Break *(ETZ Foyer)*
10:30 - 12:15	**Session #1** Track 1 *(ETZ E7)* Track 2 *(ETZ E8)* Track 3 *(ETZ E9)* Track 4 *(ETF E1)*	**Session #3** Track 1 *(ETZ E7 / D61.1)* Track 2 *(ETZ D96.1)* Track 3 *(ETZ E9)* Track 4 *(ETF E1)*	**Session #6** Track 1 *(ETZ E7 / D61.1)* Track 2 *(ETZ D96.1)* Track 3 *(ETZ D61.2)* Track 4 *(ETF E1)*	**Session #7** Track 1 *(ETZ E7 / D61.1)* Track 2 *(ETZ D96.1)* Track 3 *(ETZ E9)* Track 4 *(ETF E1)*	**Mixing event** *(ETF E1, ETZ E7, ETZ E8, ETZ E9, ETZ J64, ETZ D61.1, ETZ D61.2, ETZ D96.1)*
12:15 - 14:30	Lunch *(UniMensa)*	Lunch *(UniMensa)*	Lunch *(UniMensa)*	Lunch *(UniMensa)*	Lunch *(UniMensa)*
14:30 - 16:15	**Session #2** Track 1 *(ETZ E7 / D61.1)* Track 2 *(ETZ D96.1)* Track 3 *(ETZ E9)* Track 4 *(ETF E1)*	**Session #4** Track 1 *(ETZ E7 / D61.1)* Track 2 *(ETZ D96.1)* Track 3 *(ETZ D61.2)* Track 4 *(ETF E1)*	Social Event Visit of the Technorama	**Session #8** Track 1 *(ETZ E7 / D61.1)* Track 2 *(ETZ D96.1)* Track 3 *(ETZ D61.2)* Track 4 *(ETF E1)*	**Mixing event** *(ETF E1, ETZ E7, ETZ E8, ETZ E9, ETZ J64, ETZ D61.1, ETZ D61.2, ETZ D96.1)*
16:15 - 16:45	Coffee Break *(ETZ Foyer)*	Coffee Break *(ETZ Foyer)*		Coffee Break *(ETZ Foyer)*	Coffee Break *(ETZ Foyer)*
16:45 - 18:30	**Mixing event** *(ETF E1, ETZ E7, ETZ E8, ETZ E9, ETZ J64, ETZ D61.1, ETZ D61.2, ETZ D96.1)*	**Session #5** Track 1 *(ETZ E7 / D61.1)* Track 2 *(ETZ D96.1)* Track 3 *(ETZ D61.2)* Track 4 *(ETF E1)*		**Session #9** Track 1 *(ETZ E7 / D61.1)* Track 2 *(ETZ D96.1)* Track 3 *(ETZ D61.2)* Track 4 *(ETF E1)*	**Mixing Presentations** *(ETF E1)*
18:30 - 19:00	Apero *(ETZ foyer)*	Dinner *(Dozentenfoyer)*			Apero *(ETZ foyer)*
19:00 - 20:00					
20:00 - 21:00					
21:00 - 22:00					
22:00 - 23:00					

Keynote 1

Random thoughts after 60 years in the trenches

Yale Patt

Location: ETF E1
Date: 3th June 2024

After 60 years of teaching, I have acquired more than a few thoughts about what we do in computer architecture. Was computer architecture really dead for about 30 years, ...until recently, as some would have you believe? Is research in computer architecture important, or is it Scholarship we should be after in the university? When will useful Quantum computers ship?

Is Machine Learning missing a most important piece? Must research results have a serious baseline? Why does better branch prediction accuracy not usually come with similar IPC improvement? Does economics get in the way of solving some very important problems? My purpose in this talk to share my thoughts about some of these items. I will also give you my thoughts about what the microprocessor will look like ten years from now. ... and offer a few suggestions about your time here this week at ETH.

Yale Patt is a teacher at The University of Texas at Austin where he continues to enjoy teaching, research, and consulting more than 60 years after beginning his journey into computer technology. He earned obligatory degrees from reputable universities and has received more than enough awards for his research and teaching. For those who want it, more detail is on his home page: users.ece.utexas.edu/~patt.

Keynote 2

Open RISC-V Platforms in the era of Embodied Foundation Models

Luca Benini

Location: ETF E1
Date: 4th June 2024

The AI revolution is accelerating from its perception-focused origins to the generative era, where foundation models, trained on trillions of mostly unlabeled samples via self-supervision, produce multi-modal outputs such as text, sound, images. Foundation models are poised to disrupt multiple businesses. However, even more fundamental disruption will come when we will be able to «embody» these models in cars, robots, eyeglasses… To achieve this goal we need to tackle major challenges in energy efficiency, safety, security and real-time predictability of fine-tuned foundation models, while curtailing their sheer computational complexity. In this talk I will focus on designing hardware and systems for embodied AI, moving from perceptive to generative models and leveraging an open-platform approach, based on RISC-V processors and accelerators, to ensure long term sustainability, safety and security.

Prof. Dr. Luca Benini holds the chair of Digital Circuits and Systems at ETHZ and is Full Professor at the Universita di Bologna. He received a PhD from Stanford University. Prof. Dr. Benini's research interests are in energy-efficient parallel computing systems, smart sensing micro-systems and machine learning hardware. He has published more than 1000 peer-reviewed papers and five books. Prof. Dr. Benini has won numerous awards, including the 2016 IEEE CAS Mac Van Valkenburg award, the 2019 IEEE TCAD Donald O. Pederson Best Paper Award, and the ACM/IEEE A. Richard Newton Award in 2020. He is an ERC Advanced Grant winner, a Fellow of both the IEEE and the ACM, and is a member of the Acadamia Europea.

Keynote 3

From Software Programs to Digital Circuits

Lana Josipovic

Location: ETF E1
Date: 5th June 2024

High-Level Synthesis (HLS) tools enable programmers to automatically generate hardware designs from high-level software abstractions instead of writing tedious and time-consuming low-level hardware descriptions. However, today's HLS tools are still accessible only to expert users and for particular classes of applications; generating good-quality circuits still requires peculiar code restructuring and extensive experimentation with the tools. In this talk, I will discuss the challenges and limitations of current HLS approaches. I will outline an alternative HLS technique that overcomes these limitations and achieves high parallelism in general-purpose software applications. Finally, I will share my vision of future advancements in HLS and discuss the role of HLS in designing next-generation hardware applications.

Prof. Dr. Lana Josipović is an assistant professor in the Department of Information Technology and Electrical Engineering at ETH Zurich, where she leads the Digital Systems and Design Automation Group (https://dynamo.ethz.ch/). Her research aims to enable various programmers to benefit from digital hardware acceleration and explores synergies across compilers, programming languages, digital hardware design, and computer architecture.

Keynote 4

Memory-Centric Computing

Onur Mutlu

Location: ETF E1
Date: 6th June 2024

Computing is bottlenecked by data. Large amounts of application data overwhelm storage capability, communication capability, and computation capability of the modern machines we design today. As a result, many key applications› performance, efficiency, and scalability are bottlenecked by data movement. In this talk, we describe three major shortcomings of modern architectures in terms of 1) dealing with data, 2) taking advantage of the vast amounts of data, and 3) exploiting different semantic properties of application data. We argue that an intelligent architecture should be designed to handle data well. We posit that handling data well requires designing architectures based on three key principles: 1) data-centric, 2) data-driven, 3) data-aware.

We give several examples for how to exploit each of these principles to design a much more efficient and high performance computing system. We especially discuss recent research that aims to fundamentally reduce memory latency and energy, and practically enable computation close to data, with at least two promising directions: 1) processing using memory, which exploits analog operational properties of memory chips to perform massively-parallel operations in memory, with low-cost changes, 2) processing near memory, which integrates sophisticated additional processing capability in memory controllers, the logic layer of 3D-stacked memory technologies, or memory chips to enable high memory bandwidth and low memory latency to near-memory logic. We show both types of architectures can enable orders of magnitude improvements in performance and energy consumption of many important workloads, such as graph analytics, database systems, machine learning, video processing, climate modeling, genome analysis. We discuss how to enable adoption of such fundamentally more intelligent architectures, which we believe are key to efficiency, performance, and sustainability. We conclude with some research opportunities in and guiding principles for future computing architecture and system designs.

Onur Mutlu is a Professor of Computer Science at ETH Zurich. He is also a Visiting Professor at Stanford University and a faculty member at Carnegie Mellon University, where he previously held the Strecker Early Career Professorship. His current broader research interests are in computer architecture, systems, hardware security, and bioinformatics. A variety of techniques he, along with his group and collaborators, has invented over the years have influenced industry and have been employed in commercial microprocessors and memory/storage systems. He obtained his PhD and MS in ECE from the University of Texas at Austin and BS degrees in Computer Engineering and Psychology from the University of Michigan, Ann Arbor. He started the Computer Architecture Group at Microsoft Research (2006-2009), and held various product and research positions at Intel Corporation, Advanced Micro Devices, VMware, and Google. He received various honors for his research, including the Persistent Impact Prize of the Non-Volatile Memory Systems Workshop, the Intel Outstanding Researcher Award, the IEEE High Performance Computer Architecture Test of Time Award, the IEEE Computer Society Edward J. McCluskey Technical Achievement Award, ACM SIGARCH Maurice Wilkes Award and a healthy number of best paper or “Top Pick” paper recognitions at various computer systems, architecture, and security venues. He is an ACM Fellow, IEEE Fellow, and an elected member of the Academy of Europe.

His computer architecture and digital logic design course lectures and materials are freely available on YouTube (https://www.youtube.com/OnurMutluLectures), and his research group makes a wide variety of software and hardware artifacts freely available online (https://safari.ethz.ch/). For more information, please see his webpage at https://people.inf.ethz.ch/omutlu/.

Keynote 5

Accelerating Design Innovation

Andrew Kahng

Location: ETF E1
Date: 7th June 2024

Over the past week, this Summer School has provided an exciting introduction to future directions for hardware systems. And, it has demonstrated how open source can lower barriers and democratize both education and innovation in IC design and computer architectures. This talk will focus on why open-source design automation (i.e., EDA) technology must be considered an essential element – and accelerant – of design innovation. The intrinsic benefits of open source (availability, flexibility, scalability, transparency and reproducibility, cost, low friction, etc.) are well known, and become magnified as regions seek to develop self-sustaining research and innovation ecosystems. Additional value stems from the following.

(1) Commercial EDA tools exist only where there are established markets. Hence, open-source EDA must support early "pathfinding" explorations, e.g., of system-technology co-optimizations, 3D/heterogeneous integration, or multiphysics design closure methodologies.

(2) Realizing the promise of AI and machine learning across hardware system designs, design methods, and design tools will require open-source EDA to unblock data generation, public sharing of datasets and benchmarks, availability of foundation models, and more.

(3) Open source is also the basis for algorithm and optimization innovation toward an "EDA 2.0" that will one day achieve results in less time (multithreading, GPU) and better results in the same time (cloud-native, sampling), likely in concert with AI/ML methods that steer and orchestrate the design process. The talk will conclude with thoughts on how a global community might foster and sustain open-source EDA technology to obtain these benefits.

Andrew B. Kahng is Distinguished Professor of CSE and ECE and holder of the endowed chair in high-performance computing at UC San Diego. He was visiting scientist at Cadence (1995-97) and founder/CTO at Blaze DFM (2004-06). He is coauthor of 3 books and over 500 journal and conference papers, holds 35 issued U.S. patents, and is a fellow of ACM and IEEE. He was the 2019 Ho-Am Prize laureate in Engineering. He has served as general chair of DAC, ISPD and other conferences, and from 2000-2016 served as international chair/co-chair of the International Technology Roadmap for Semiconductors (ITRS) Design and System Drivers working groups. He was the principal investigator of the U.S. DARPA "OpenROAD" project (https://theopenroadproject.org/) from June 2018 to December 2023, and until August 2023 served as principal investigator and director of "TILOS" (https://tilos.ai/), a U.S. NSF AI Research Institute.

Track 1

Digital IC Design with Open Source EDA Tools

Track 1 offers hands on tutorials to showcase how to design Digital Integrated Circuits using OpenRoad design tools using IHP 130nm technology on a Linux-capable 64bit RISC-V system.

You will learn

- An introduction to digital design flow (synthesis, placement, routing)
- What is the state of the art in open source PDK and libraries for designing a modern IC
- Synthesize an actual open-source design (Cheshire) using a CVA6 core and peripherals
- Back-end design using OpenRoad
- Pieces that are still missing, or lacking currently in open-source design flows
- The cost and opportunities of bringing an open-source design into reality

Location

The first session will be held in room ETZ E7.
The following sessions will be either in ETZ E7 (lecture room) or in ETZ D61.1 (computer room, for the hands on).

Agenda

Session 1: Introduction, goals of the track, people involved, quick summary of IC Design flow, IHP and open source PDKs
Duration: 2 hours
Type: lecture

Session 2: Using the OpenRoad UI, navigating designs
Duration: 2 hours
Type: hands-on

Session 3: SystemVerilog/RTL primer. RTL simulation
Duration: 2 hours
Type: lecture + hands-on

Session 4: Synthesis flow and timing constraints
Duration: 2 hours
Type: hands-on

Session 5: Floorplanning, Macro placement, power planning
Duration: 2 hours
Type: hands-on

Session 6: Issues for Back-end design, Cost of IC Design
Duration: 2 hours
Type: lecture

Session 7: OpenRoad backend flow
Duration: 2 hours
Type: hands-on

Session 8: Completing the design flow, DRC/LVS, odds and ends
Duration: 2 hours
Type: hands-on

Session 9: Closing/State of the Art Open Source tools, what is still missing
Duration: 2 hours
Type: lecture

Track 2

Customizing RISC-V Based Microcontrollers

Track 2 offers hands on tutorials to demonstrate how to add your own extensions to an existing 32bit RISC-V microcontroller system, running on an FPGA.

You will learn

- An introduction to microcontroller architectures (RISC-V, Peripherals, Memories)
- Open source microcontroller architectures available through the PULP eco-system
- FPGA design flow for FPGAs, mapping the RTL code to FPGA
- Getting your code run on the system, cross compiling, RTL simulation, emulation
- Alternatives to expand the capabilities, accelerators, instruction set extensions
- Cost benefit analysis of extensions

Location

The first session will be held in room ETZ E8.
The following sessions will be either in ETZ E8 (lecture room) or in ETZ D96.1 (computer room, for the hands on).

Agenda

This Track aims at bringing you up to speed with the development of your own modified flavor of RISC-V based technology – specifically, based on PULPissimo. We focus in particular on the design of extended computation units both inside the core (ISA extensions) and outside it (cooperative or loosely-coupled accelerators), and on delivering all the tools necessary for a productive research environment using PULPissimo as a starting point.

PULP platform & PULPissimo microcontroller architecture (4h Mon)

- Lecture: Introduction to the PULPissimo microcontroller architecture
- Hands-on:
  - Introduction to the PULPissimo simulation environment
  - Plugging an APB timer into the PULPissimo MCU
  - Introduction to the software environment and development of a FIR filter

Extending RISC-V cores and the RISC-V LLVM compiler (6h Tue)

- Lecture: In-depth RISC-V ISA, microarchitecture of CV32E40X, and CV-XIF specification
- Hands-on: Extending the RISC-V ISA with new instructions for FIR Filter acceleration
- Lecture: LLVM RISC-V compiler and its extension
- Hands-on: Extending LLVM for our FIR instructions, and testing them on the SW FIR filter

Integrating cooperative HW Processing Engines / HWPEs (4h Wed-Thu)

- Lecture: Why should an ISA extension not be enough? PULP cooperative HWPEs and loosely-coupled accelerators
- Hands-on:
  - Guided design of a FIR filter HWPE
  - Integration of the FIR HWPE in the PULPissimo architecture
  - Evaluating acceleration in the PULPissimo environment

Testing extended PULPissimo on FPGA (4h Thu)

- Lecture: PULPissimo FPGA emulation flow and methodology
- Hands-on: Deployment flow on FPGA: synthesis + place & route + bitstream generation
- Hands-on: Testing the design on FPGA with OpenOCD

Track 3

Embedded AI for Biosignal Processing

Track 3 offers hands on tutorials to work on porting embedded AI applications for biosignal processing to state of the art heterogeneous IoT processors from Greenwaves Technologies (GAP8/GAP9).

You will learn

- An introduction to shared-memory many-core architectures
- Issues in explicit memory management, tiling, DMA transfers
- Working with 16-8bit data formats to fit data into memory, quantization
- Programming the devices
- Interfacing sensors, collecting data
- Running your programs, debugging

Location

The first session will be held in room ETZ E9.
The following sessions will be either in ETZ E9 (lecture room) or in ETZ D61.2 (computer room, for the hands on).

Agenda

Session 1: Motivation + SW-Optimization 1 (Generic DNN Optimization concepts, Efficient Layers)

- Duration: 2 hours
- Type: lecture

Session 2: Biosignal processing applications, Introduction to GAP9 (PULP architecture etc) and Acquisition Platform

- Duration: 2 hours
- Type: lecture

Session 3: SW-Optimization 2 (NAS, Pruning, Quantization)

- Duration: 2 hours
- Type: lecture

Session 4: DNN Creation and optimization (NAS, Pruning) on pre-collected dataset

- Duration: 2 hours
- Type: hands-on

Session 5: Personalized Dataset Acquisition, Visualization

- Duration: 2 hours
- Type: hands-on

Session 6: Finetuning of the optimized DNNs on personalized data; Quantization

- Duration: 2 hours
- Type: hands-on

Session 7: Compilers/Deployment (AI compilers, TVM, Zigzag, MATCH)

- Duration: 2 hours
- Type: lecture

Session 8: Compilation of DNN models and programming of BioGAP, test of DNNs found in Session #4 on the board, using random inputs

- Duration: 2 hours
- Type: hands-on

Session 9: Streaming data from BioGAP to PC and real-time visualization of predicted gestures

- Duration: 2 hours
- Type: hands-on

Track 4

Computer Architecture and Memory Systems

Track 4 offers a series of lectures that focuses on building fundamentally better, data-centric, data-driven, and data-aware architectures in computer systems.

You will learn

- An introduction to intelligent architectures that are NOT overwhelmed by data
- How to architect fundamentally secure/reliable/safe systems
- How to architect fundamentally energy-efficient systems
- How to architect fundamentally low-latency and predictable architectures
- How to design accelerators for modern data-intensive applications, e.g., AI/ML (including LLMs), Genomics, Medical & Health Systems, Graph Analytics, Data Analytics

Location

Track 4 will be held in room ETF E1

Agenda

Basics of Modern Memory Systems: Caching, Memory, Storage

- Duration: ~4 hours

Memory reliability, security, safety

- Duration: ~4 hours

Accelerating ML, Genomics, and data-intensive workloads: SIMD/GPU/Systolic Arrays

- Duration: ~4 hours

Memory-Centric Computing and Storage-Centric Computing

- Duration: ~4 hours

Memory System for ML and ML for Memory System

- Duration: ~2 hours (one lecture)

Mixing Event

This event gives you the opportunity to interact with participants from other tracks. You will be assigned a Mixer Group number when you register.

Each Mixer Group is made up of participants from different tracks and Mixer Leaders. Each Mixer Group is also characterised by a ‹theme›.
The aim of the event is to discuss your theme: research directions, start-up opportunities, how you can use what you have learned in this Summer School for your theme, what you are excited about, what you think are the issues/problems to realise your vision.

On **Monday afternoon**, there will be an introductory lecture for the Mixer Events. Here you will meet your team leaders, who will introduce you to your group›s theme and give you more details about the goals you have set.
Please meet your team partners and team leaders at these locations during the Monday afternoon Mixer Event slots:

- Astro Logic Engineers: Room ETZ J64.1
- BioPizza: Room ETF E1
- Capybara: Room ETZ D61.1
- Mozzarella Light: Room ETZ E7
- The Academic Avengers: Room ETZ E8
- The Void: Room ETZ D96
- Synedrion: Room ETZ D61.2
- Gene in a Bottle: Room ETZ E9

For **social events** (including the Wednesday afternoon excursion) your Mixer Leader will be your point of reference. They will guide you throughout the event. Feel free to contact them if you need help.

On **Friday**, you will work with your Mixer Group partners to prepare a short presentation and a poster about the ‘theme’ of your Mixer Group.

Please visit the introductory lecture to Mixer Events on Monday afternoon to receive all the details.

Lunch, Aperos, Social Dinner and Social Event

The **Welcome Apero** (Monday afternoon) and the **Closing Apero** (Friday afternoon) will be held in the ETZ Foyer.

The social dinner will be held in the Dozentenfoyer (see “How to reach the venues” at the end of this booklet) on Tuesday, starting at 18:30.

Lunch will be provided daily at the UniMensa (see “How to reach the venues” at the end of this booklet). You have received a lunch voucher for each day of the school.

Social event

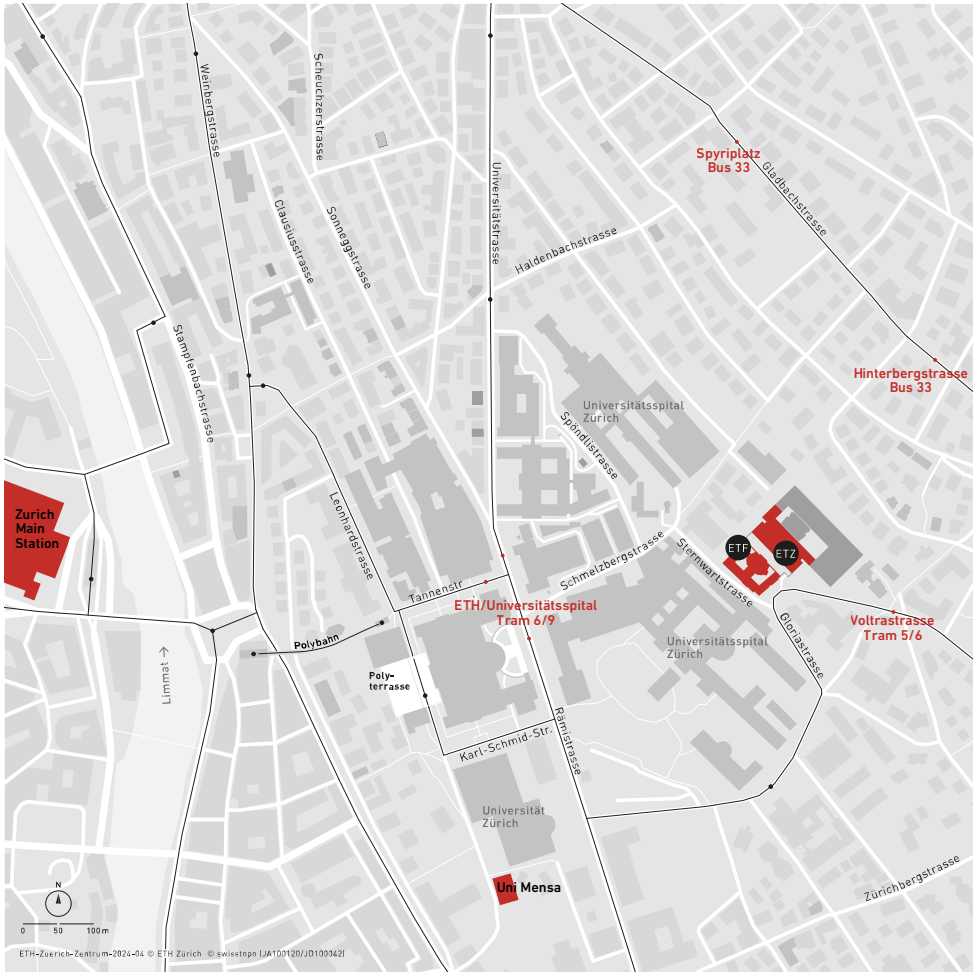
On Wednesday afternoon, we will visit the Swiss Science Center Technorama, in Winterthur. This interactive science center is renowned for its hands-on exhibits and engaging demonstrations that make science come alive. You will have the chance to explore over 500 exhibits spanning various fields of science and technology, sparking your curiosity and encouraging experimentation.

We will depart together from the UniMensa, after lunch, at 14:30. We will take the S11 train at 15:05 from the Zurich Stadelhofen train station. Participants will be divided in sub-groups for the trip and the visit according to the Mixing Event number (see “Mixing event” section of the booklet). Please follow your Mixing Event Leader during the social event.

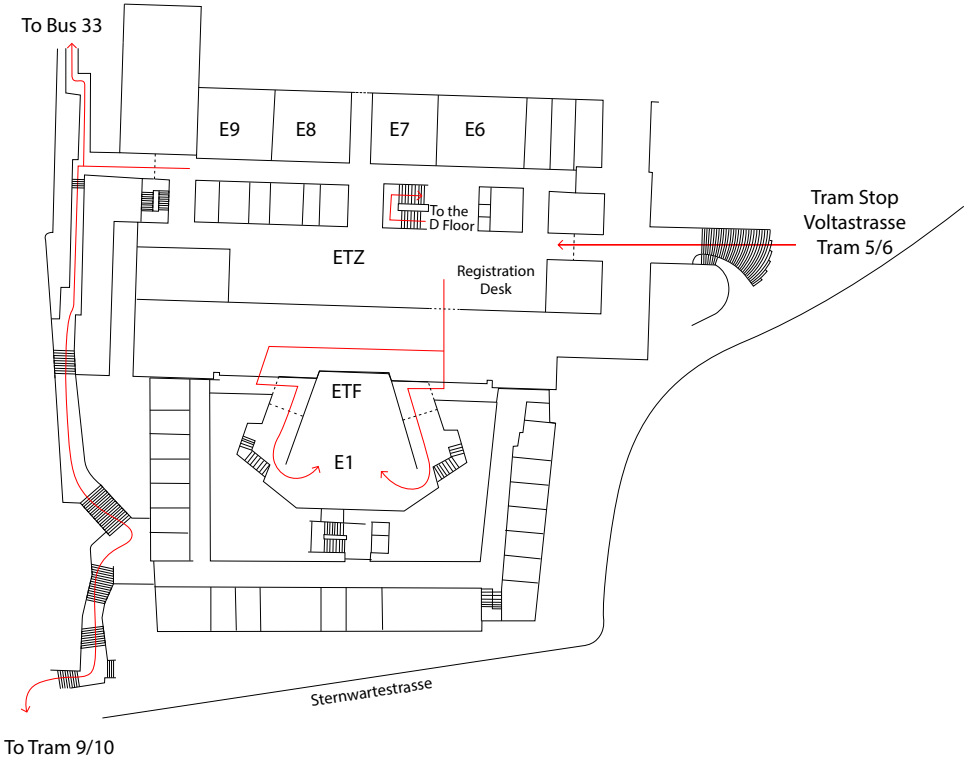
How to reach the venues

From Zurich Main Station to the ETZ Building

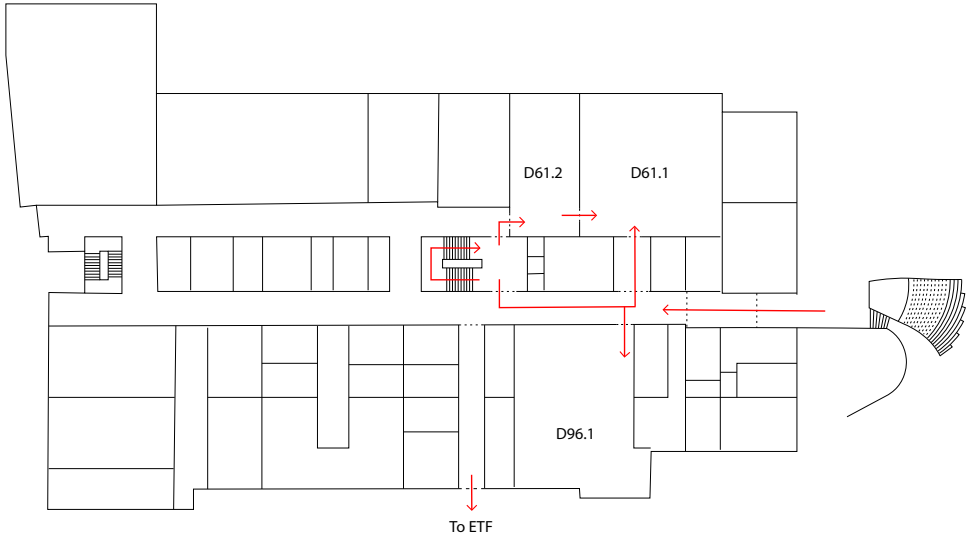
Follow the signs at the Zurich Main Station to get to the tram stop «Bahnhofstrasse». Take a tram number 6 towards the destination «Zoo». Leave at the fifth stop called «Voltastrasse». The tram ride will take about 10 minutes. Walk back a few steps to get to the main entrance of the ETZ building. Go up the large staircase to get to the ETZ foyer. You will find the registration desk on the left.



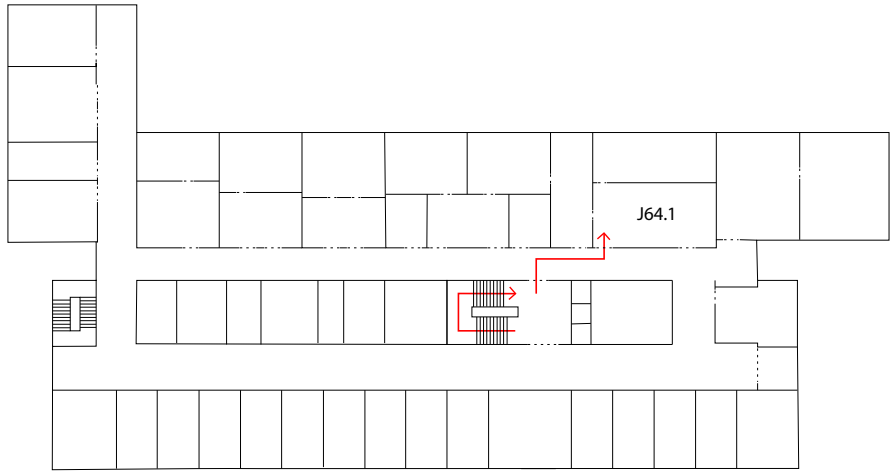
ETZ & ETF - E Floor Plan



ETZ - D Floor Plan

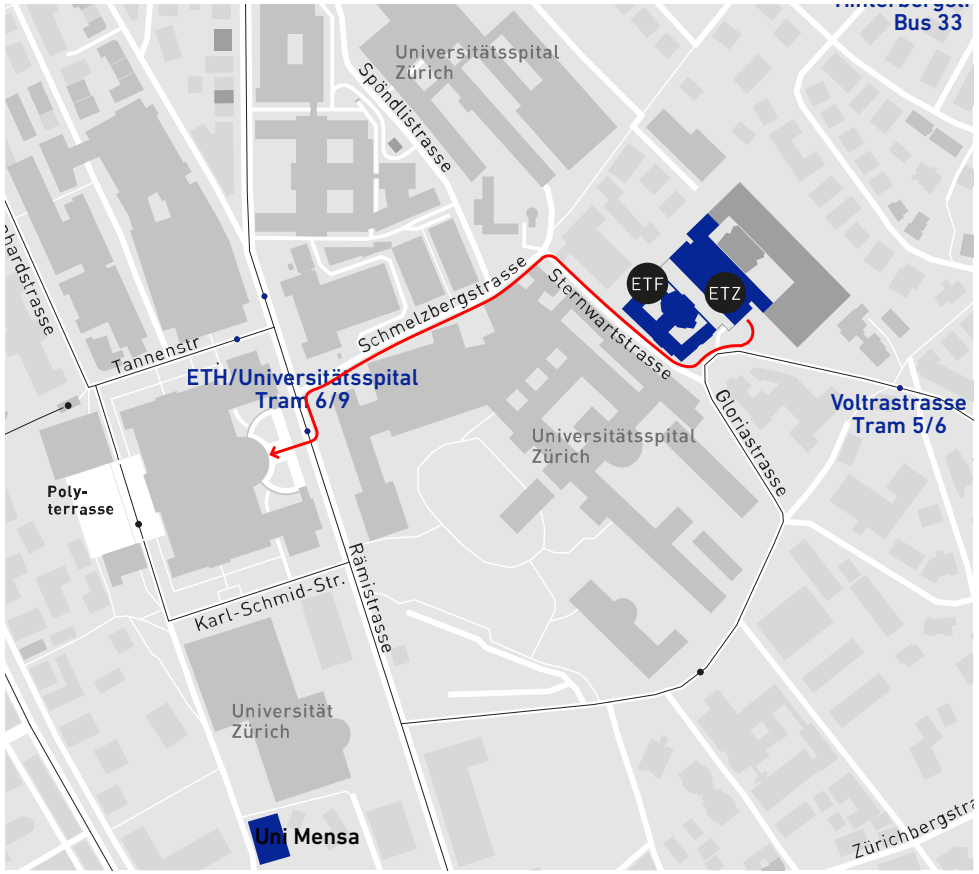


ETZ - J Floor Plan



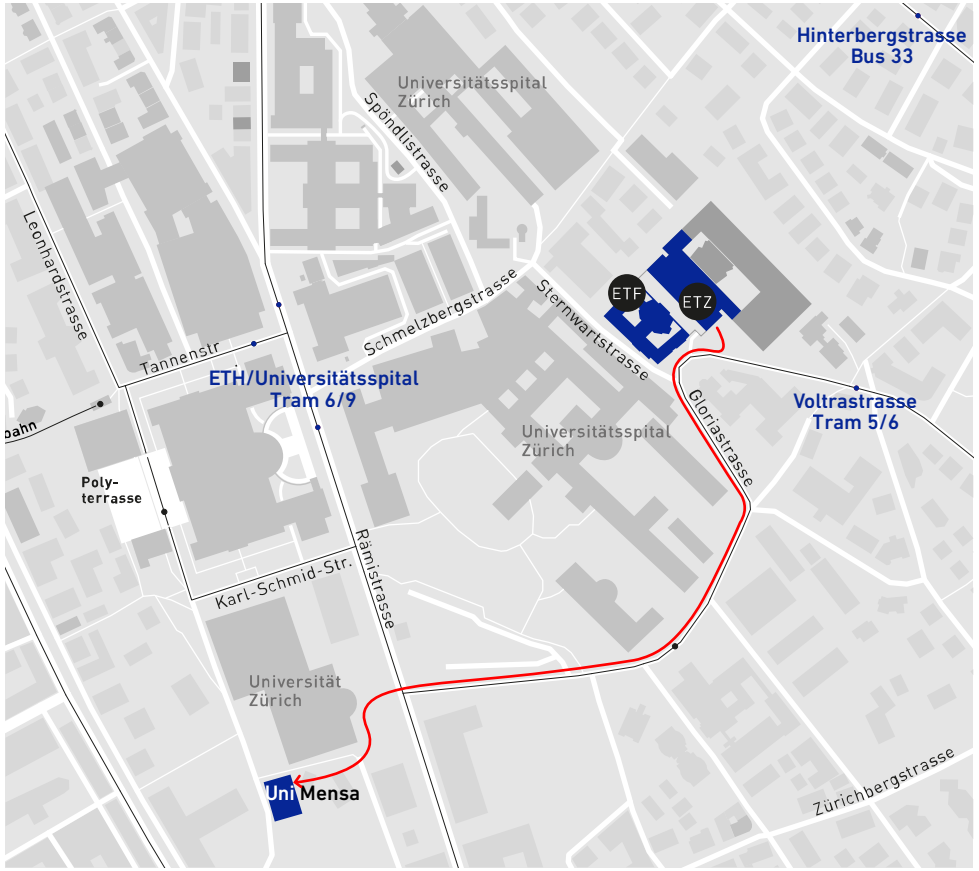
Dozentenfoyer

Exit the ETZ building through the main entrance on Gloriastrasse. Turn right on Gloriastrasse and head towards Sternwartstrasse. Turn right onto Sternwartstrasse. Then, turn left onto Schmelzbergstrasse and follow it all the way down. Turn left again onto Rämistrasse. Take the crosswalk and enter the ETH Zurich main building (Hauptgebäude, HG) at Rämistrasse 101. Once inside, take the elevator to the Dozentenfoyer on the top floor.



UniMensa

Exit the ETZ building through the main entrance on Gloriastrasse. Turn right on Gloriastrasse and continue down the street. Cross Rämistrasse and walk around the University Main building. The entrance to the mensa is down the stairs.



Questions ?

Send us an email at: summer-school@efcl.ethz.ch

An overview of the event is also available online:
https://efcl.ethz.ch/efcl-summer-school.html

ETH Future Computing Laboratory (EFCL)
info@efcl.ethz.ch
Andreasstrasse 5
8050 Zurich
Switzerland