# Preferred citation style

Axhausen, K.W. (2007) MATSIM-T: Aims, approach and implementation, presentation at the Department for International Development and Collaboration, Hiroshima University, Hagashihiroshima, July 2007.

# MATSIM-T: Aims, approach and implementation

KW Axhausen

IVT
ETH
Zürich

July 2007

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

# Overview

- Structure and team
- Task and solution methods
- MATSIM aims

- Description of the scenario, population and its travel demand
- Progress on shortest path calculations
- Traffic flow model
- Scheduling and its utility function
- Improving the convergence
- System architecture

- Outlook and next steps

# Structure

Software:

- Open-source project under GNU public licence

Coordination:

- Kai Nagel, TU Berlin

Data:

- Public sources, where available
- Private sources, when needed or as occasion arises

# Current team

Strategy:

- Kai Nagel, TU Berlin
- Kay Axhausen, ETH Zürich
- Fabrice Marchal, LET, Lyon

Coordination of the implementation and project management:

- Michael Balmer, ETH Zürich
  - Marcel Rieser, TU Berlin

# Current team: Implementation

- Michael Balmer, ETH
- David Charypar, ETH
- Francesco Ciari, ETH
- Jeremy Hackney, ETH
- Andreas Horni, ETH
- Nicolas Lefebvre, ETH
- Michael Löchl, ETH
- Fabrice Marchal, LET
- Konrad Meister, ETH
- Kai Nagel, TU Berlin
- Marcel Rieser, TU Berlin
- Nadine Schüssler, ETH
- David Stripgen, TU Berlin

# Current funding sources

- Basic research support for the chairs

- (competitive) ETH research fund
- Swiss National Fund
- German Research Society
- EU Framework funding
- VW Foundation
- Swiss Commission for Technology and Information (KTI) (datapuls, Lucerne)

# Task and solution methods

# Understanding scheduling

- Budget constraints
- Capability constraints

- Generalised costs of the schedule

  - Generalised cost of travel
  - Generalised cost of activity participation

    - Risk and comfort-adjusted weighted sums of time, expenditure and social content

# Degrees of freedom of activity scheduling

- Number and type of activities
- Sequence of activities
  - Start and duration of activity
  - Composition of the group undertaking the activity
  - Location of the activity
    - Connection between sequential locations
      - Location of access and egress from the mean of transport
      - Vehicle/means of transport
      - Route/service
      - Group travelling together

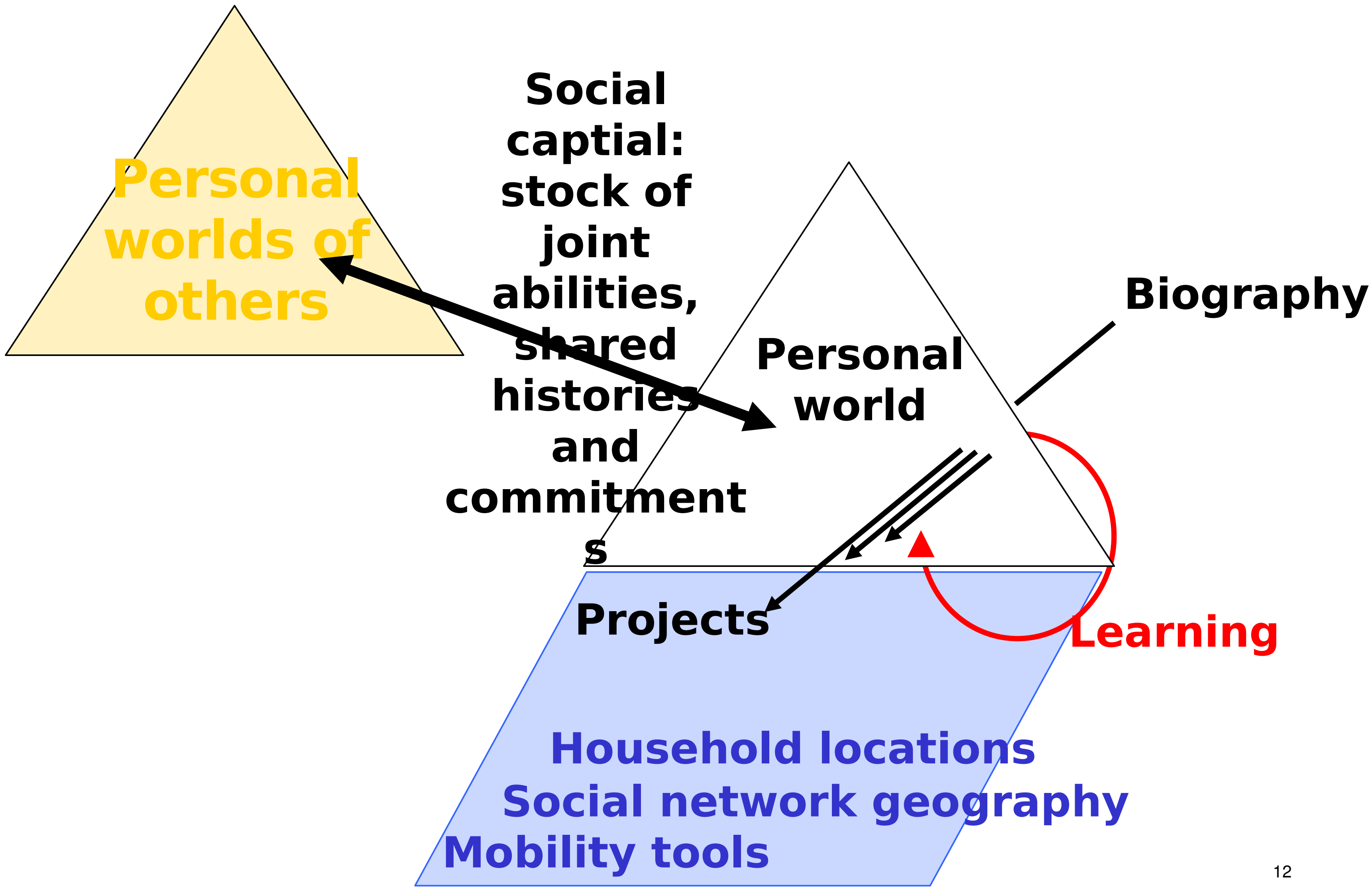
# Understanding supply

Slot: A path in the time-space environment, which allows moving or activity performance

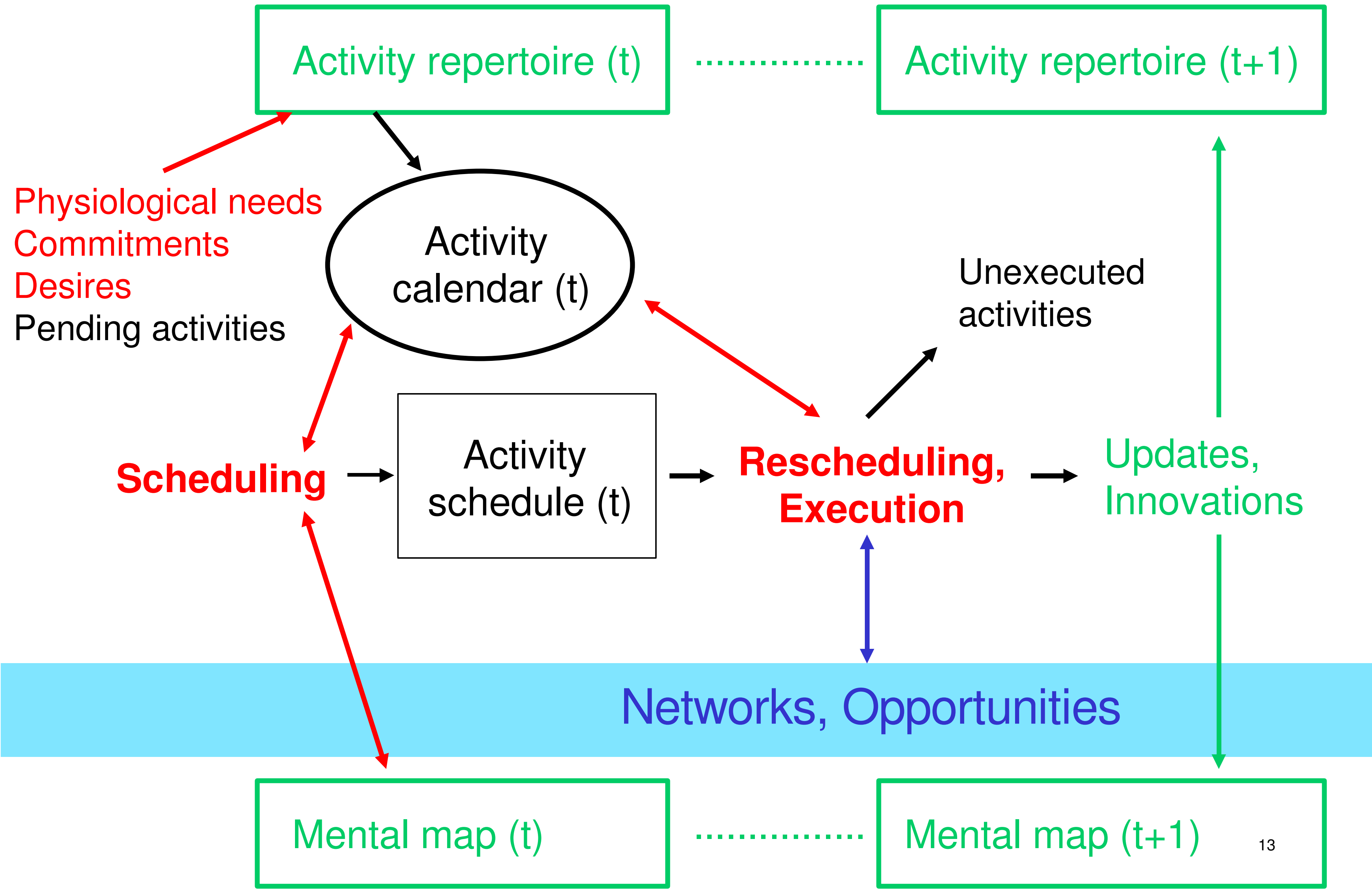
- Regulated slots (e.g. table in a restaurant, reserved seat in a theatre, gate position of a plane, green light at a junction)
- Emergent slots (e.g. trajectory of a car on a motorway, players in a pub-soccer tournament)

Waiting time ~ Reserve capacity = Capacity – Demand for slots

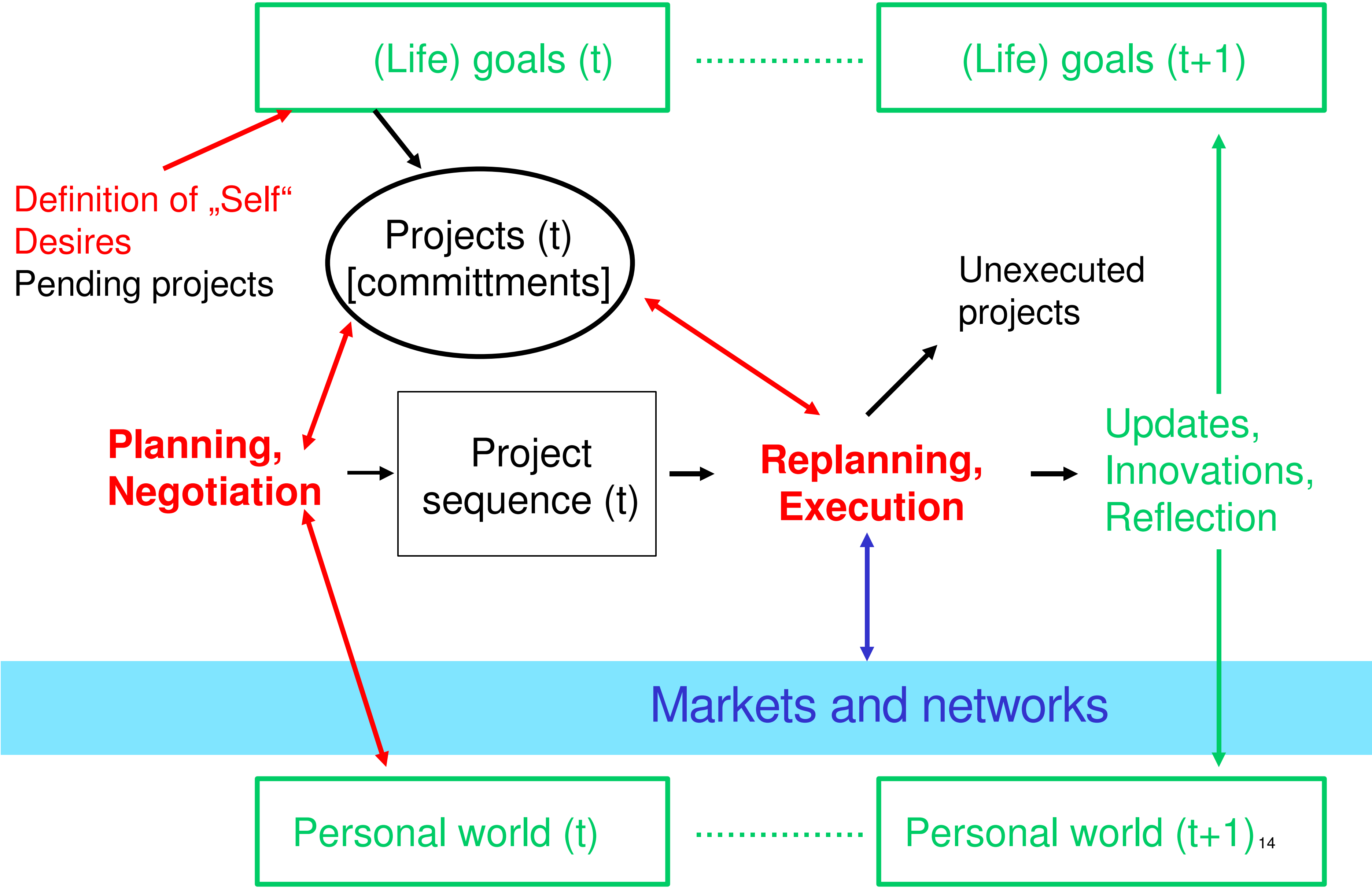
# What we would like to do ?



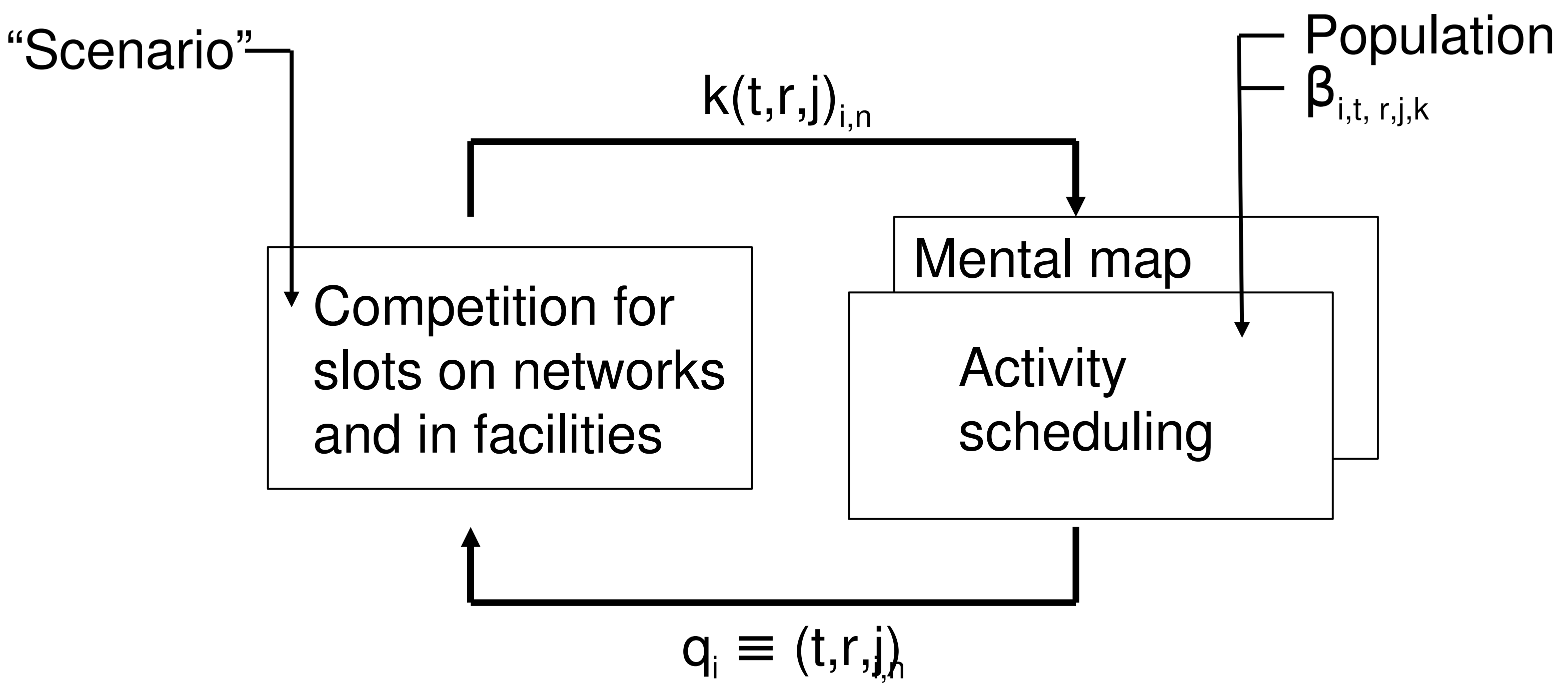
# What would we like to do ? Personal daily dynamics



# What would we like to do ? Personal long-term dynamics

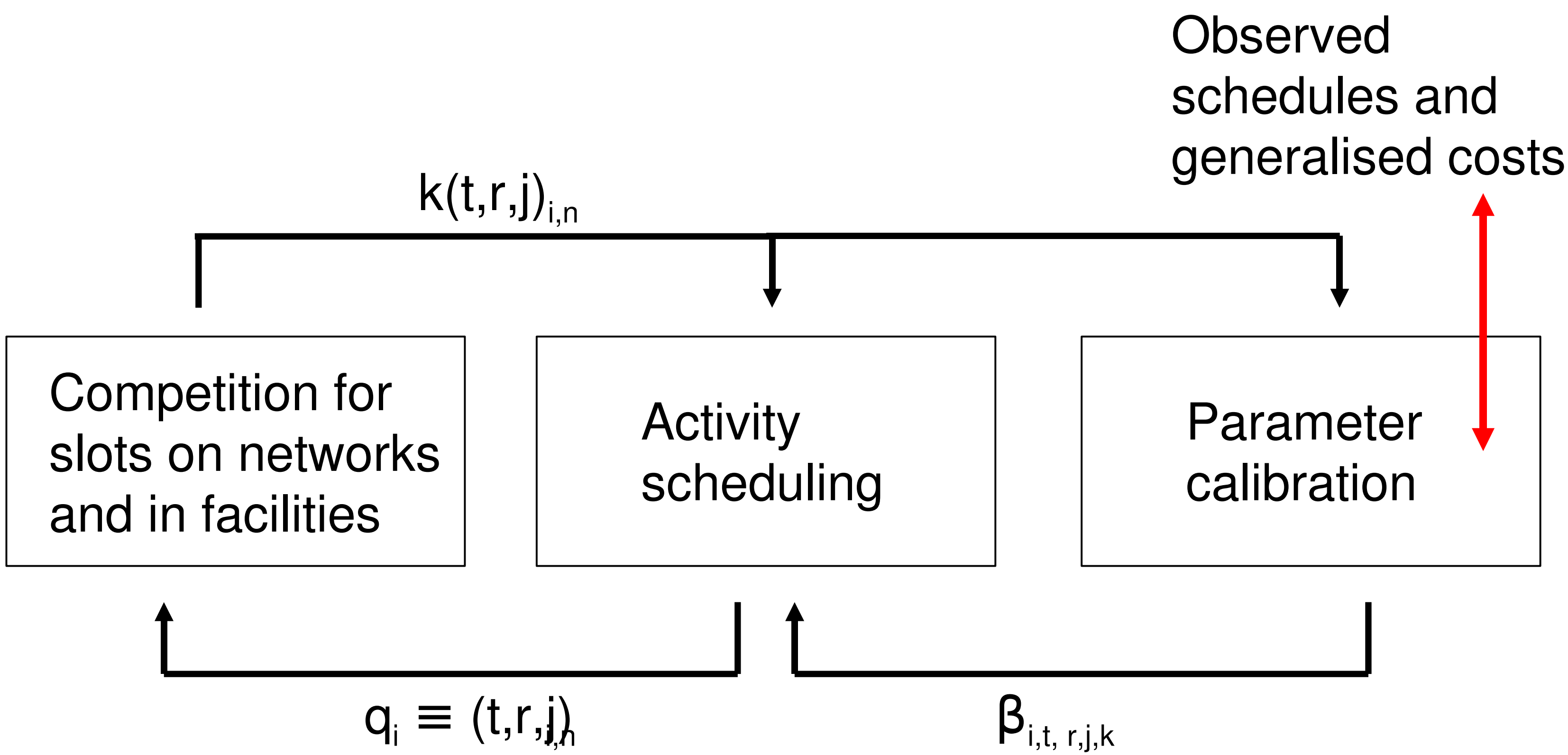


# What do we (generally) do ?

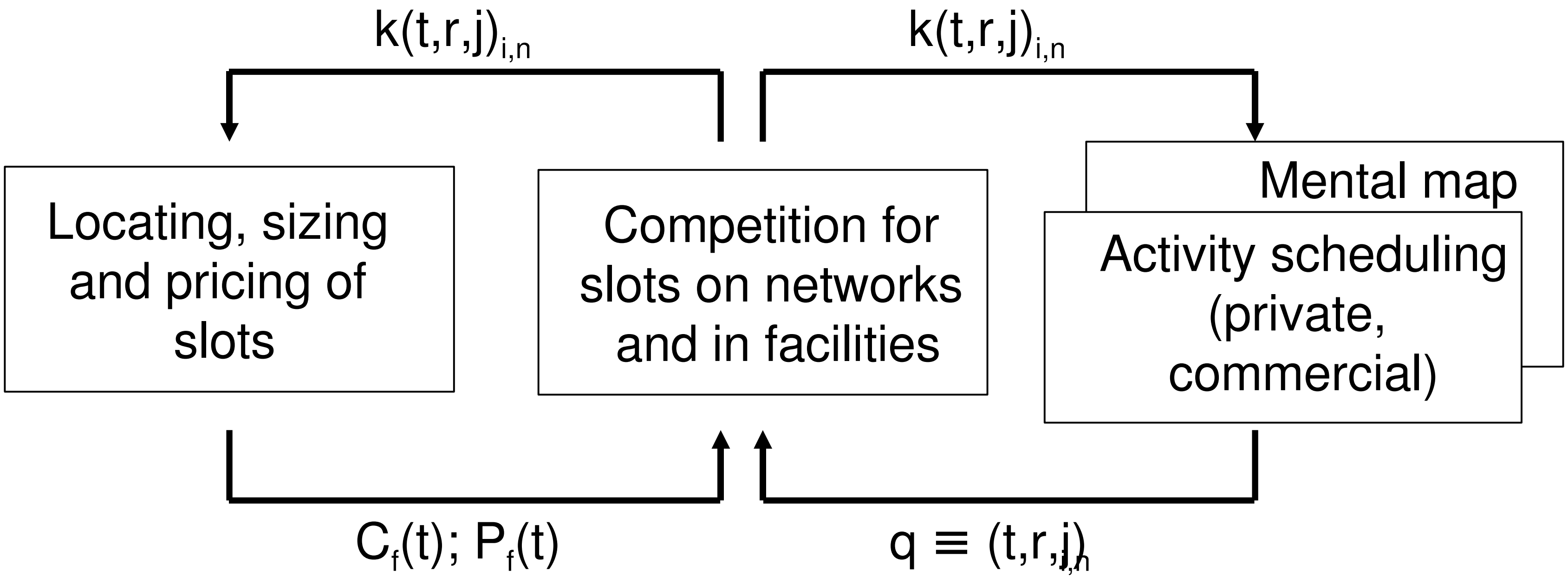



Demand q are the i[th]movements of person p from the current location at time t on route (connection) r to location j. The resulting generalised costs k are used to adjust the schedules and to change the capacities C and prices P of facilities f

# What should we do ?



# What would we like to do ?



# Classification criteria

- Steady state (equilibrium) ?
- Aggregate demands ?
- Complete and perfect knowledge ?
- Optimised schedules ?
- Degrees of freedom and detail of scheduling
- Modelling of capacity restrictions (movement, activities) ?

# MATSIM-T aims (1): Steady-state version

- Steady state within 12 hours on a small multi-CPU machine
- 7.5 mio agents, parcels, navigation networks (Switzerland)
- Shared time-of-day dependent generalised costs of travel and activity participation
- Optimised scheduling
- Continous time resolution; space: parcels; social networks
- Queuing for slots for movement and activities

# MATSIM-T aims (2): Path-dependent version

- Path-dependent development; precise estimates within 12h on a small multi-CPU machine

- Large scale scenario

- Agent-specific, learned time-of-day dependent generalised cost of travel and activity participation

- Optimised scheduling at multiple decision points
- Continuous time resolution; space: parcels; social networks

- Queuing for slots for movement and activities

# Current state (with a focus on ETH work)

# Scenario: Facilities for 140’000 hectares

```
<facilities name="Swiss National Enterprise Census">
    <facility id="101" x="606300" y="281549">
        <activity type="shop">
        <capacity value="50"/>
        <opentime day="wkday" start_time="8:00:00" end_time="19:00:00" />
        <opentime day="sat" start_time="8:00:00" end_time="16:00:00" />
        </activity>
        <activity type="work">
        <capacity value="5" />
        <opentime day="wkday" start_time="8:00:00" end_time="19:00:00" />
        <opentime day="sat" start_time="8:00:00" end_time="16:00:00" />
        </activity>
    </facility>
</facilities>
```

# Scenario: Facilities – Data sources

Census of Workplace tables by hectare for:

| | Precision for Firms | Precision for Employment |
|---|---|---|
| Firm and employees per Sector 2/3 | low | precise |
| Firms and employees in classes per NOGA-2 digit class | middle | middle |
| Firms per NOGA-4 digit class | precise | N.A. |

# Scenario: Facilities – Data sources

Example:

| NOGA Code | Description | Firms | FT-equivalents |
|---|---|---|---|
| 10-45 | (sector 2) | 6 | 200 |
| 19 | Leather and shoe production | 2 | 0-9 |
| 20 | Chemical industry | 3 | 50-249 |
| 19.10 | Leather production | yes | |
| 19.20 | Leather goods | no | |
| 19.30 | Shoe production | no | |

# Scenario: Facilities – current allocation

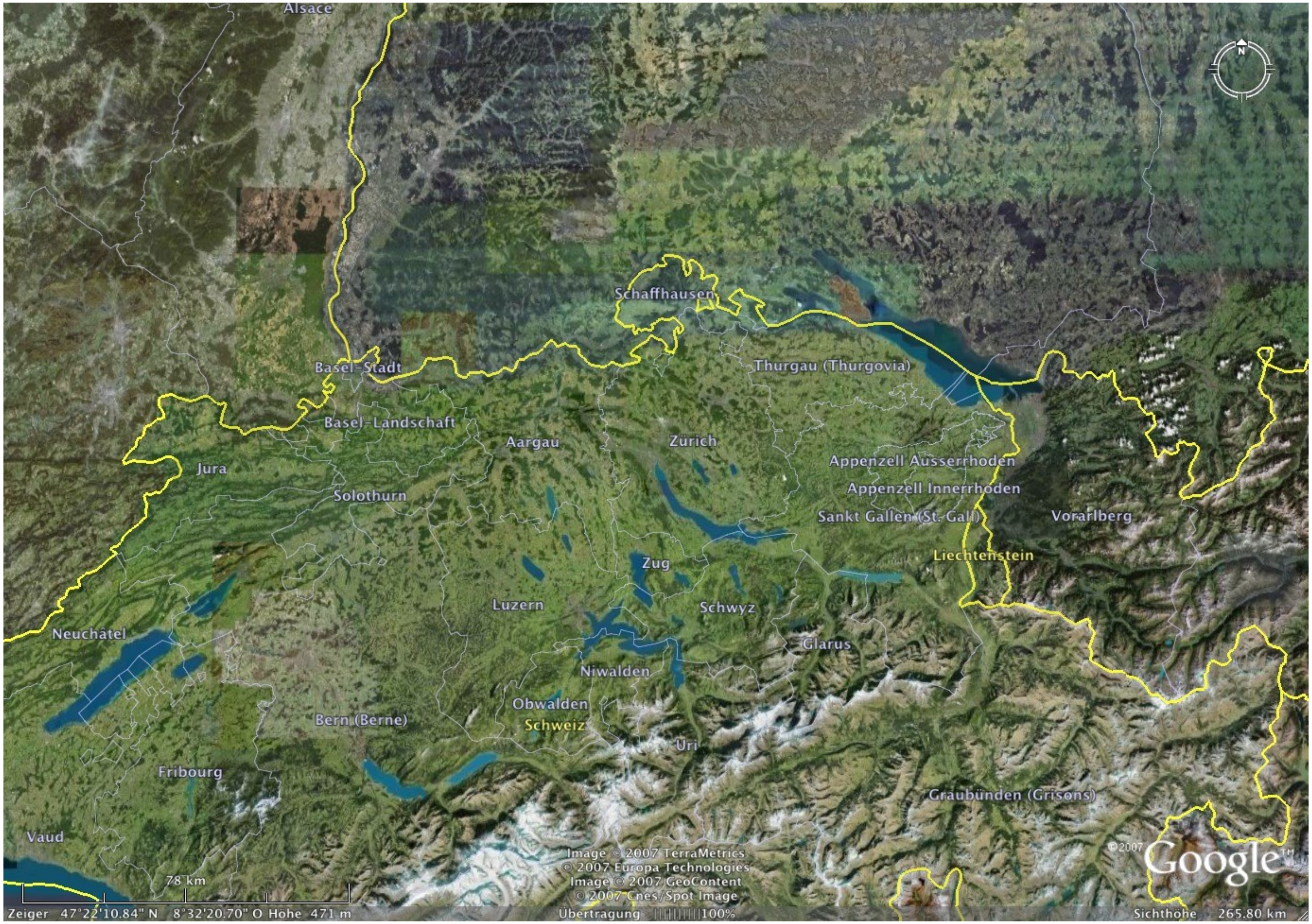
- Read census of employment for each hectare

- Generate the required number of facilities

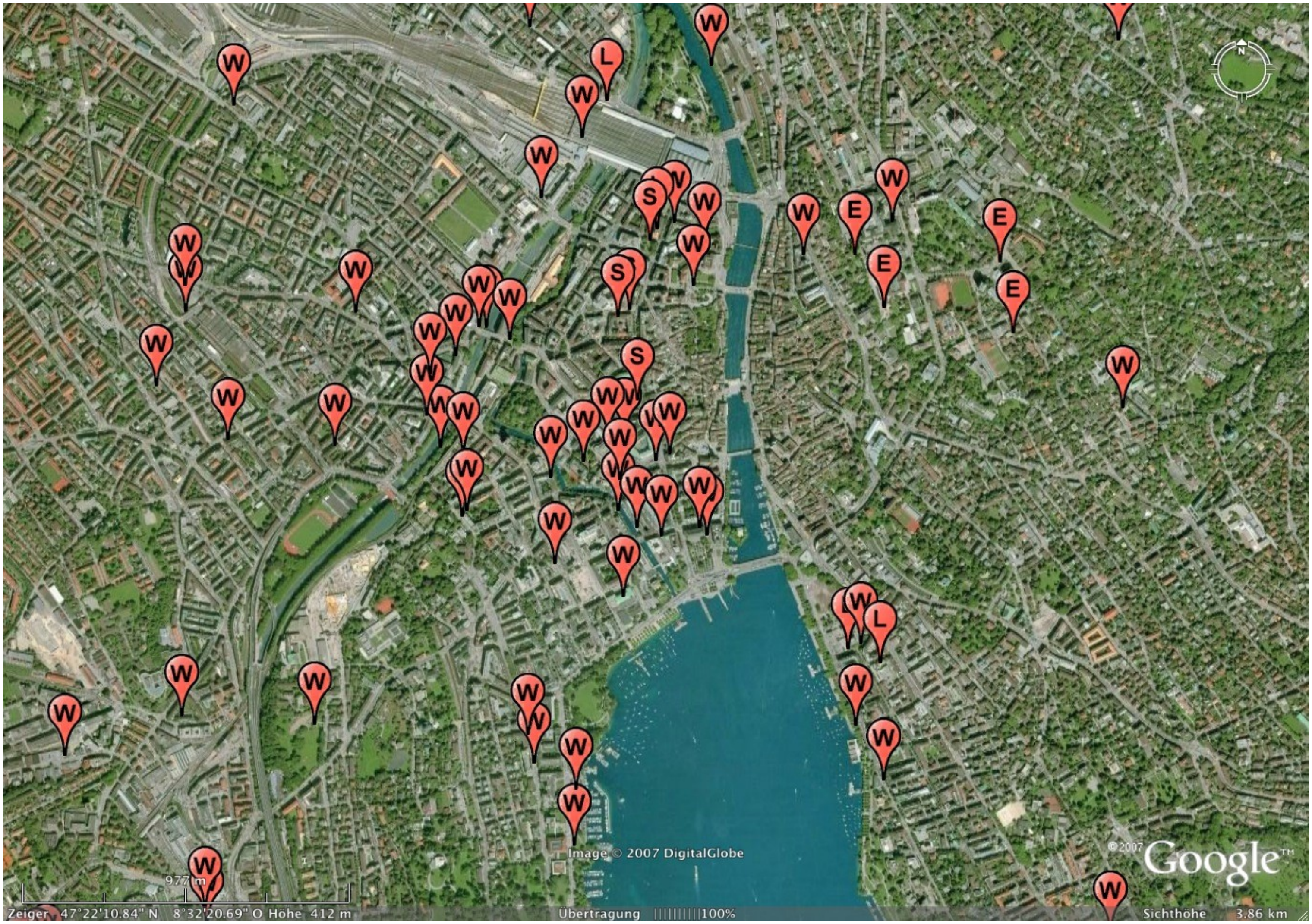
- Add activitiy work
- Set number of work place to minimum of class
- Distribute remainder proportionally to class size

- Add activity of use
- Add standard opening hours
- Randomly distribute on the hectare and attach to nearest link

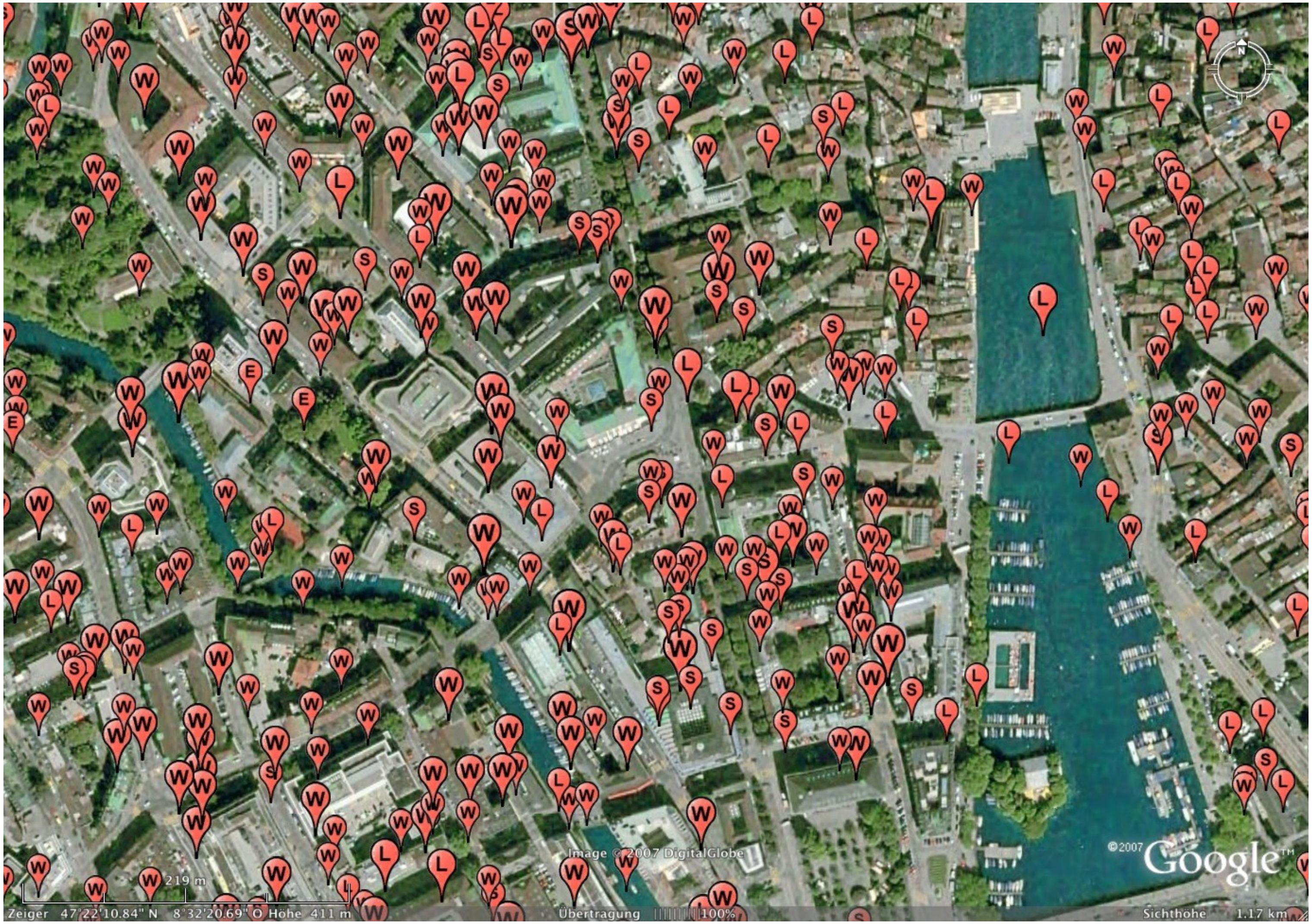
# Scenarios: Facilities - results



# Scenario: Facilities - results



# Scenario: Facilities - results



# Scenario: Network

Simplification:

- Not connected links and subnetworks
- Remove nodes between unchanging link types
- Dead ends cut off

Results:

- Navteq (882’120 links) (25% loss of links/nodes)
- Teleatlas (1’288’757 links) (currently not in use)

- National transport model network

# Scenario: Population

Alternatives:

- Artificial sample generation from Census marginals
- Census
- Private census with additional imputations (datapuls, Lucerne) plus household “formation”

# Scenario: Mobility tools

Approach:

- MNL of mobility tool packages (Beige)

- Socio-demographics
- Location type
- Travel times to main centre by road and public transport

Data:

- National travel survey (MZ 2000)

# Scenario: Demand

Data

- MZ 2000
- MZ 2005

Approach:

- Selection via conditional probability distribution from Chains * Person types frequencies
- ca. 50 activity chains for MZ 2000 (>95% of all cases)
- extended with activity durations given by MZ 2000

➔ 450 different activity chains (weighted)

# Scenario: Demand - destinations

Home:

- Random location in hectare

Work/school

- Random allocation from census commuter matrices
- Disaggregation to facilities inside municipality

Other locations

- “Neighbourhood search” from given home/work/school locations (a variation of Gravity Model)

# Scenario: Mode choice

Approach:

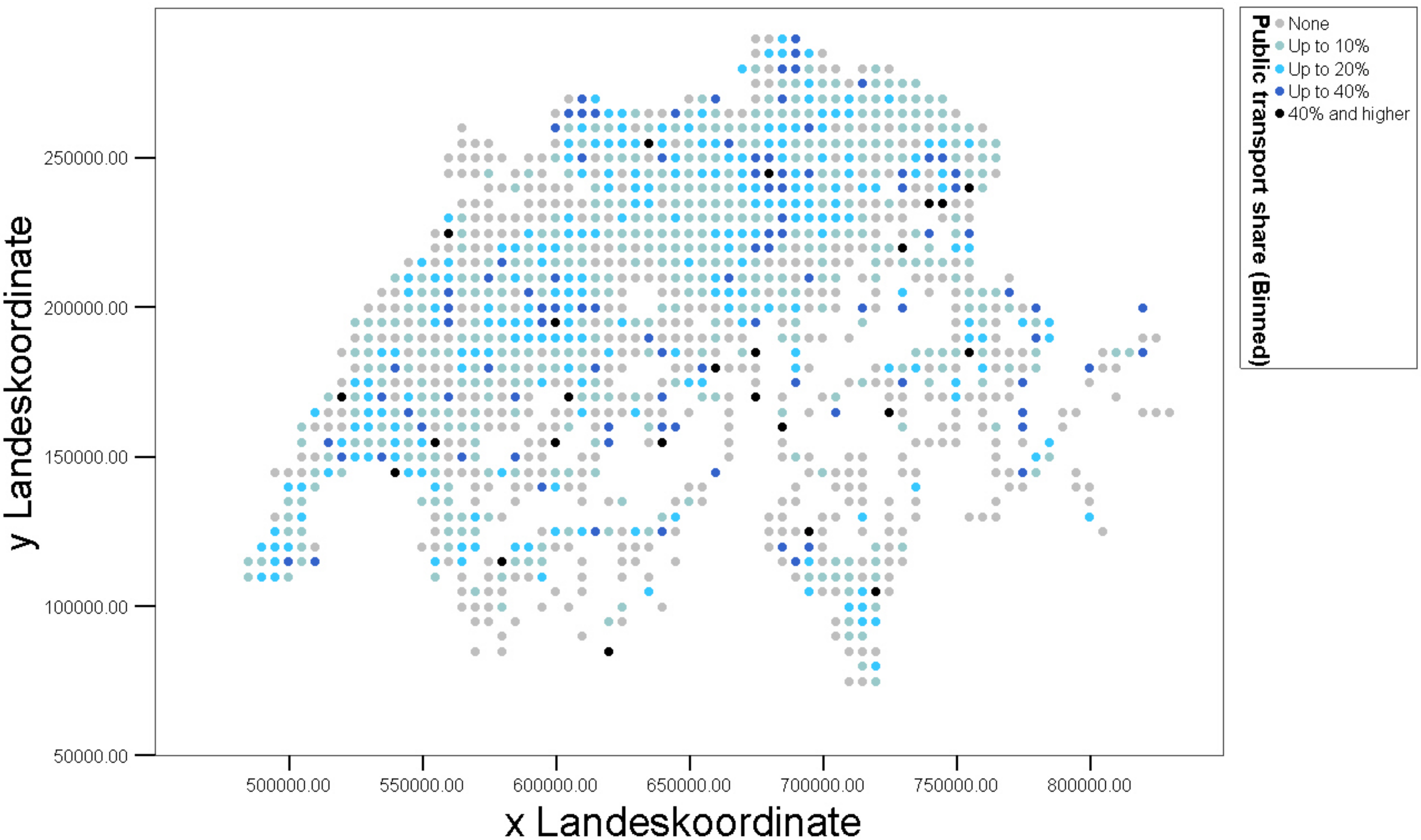
Fixed mode choice at the tour level (subtours are identified) as a function of

- Driving licence
- Mobility tool ownership
- Distance
- Age * season ticket

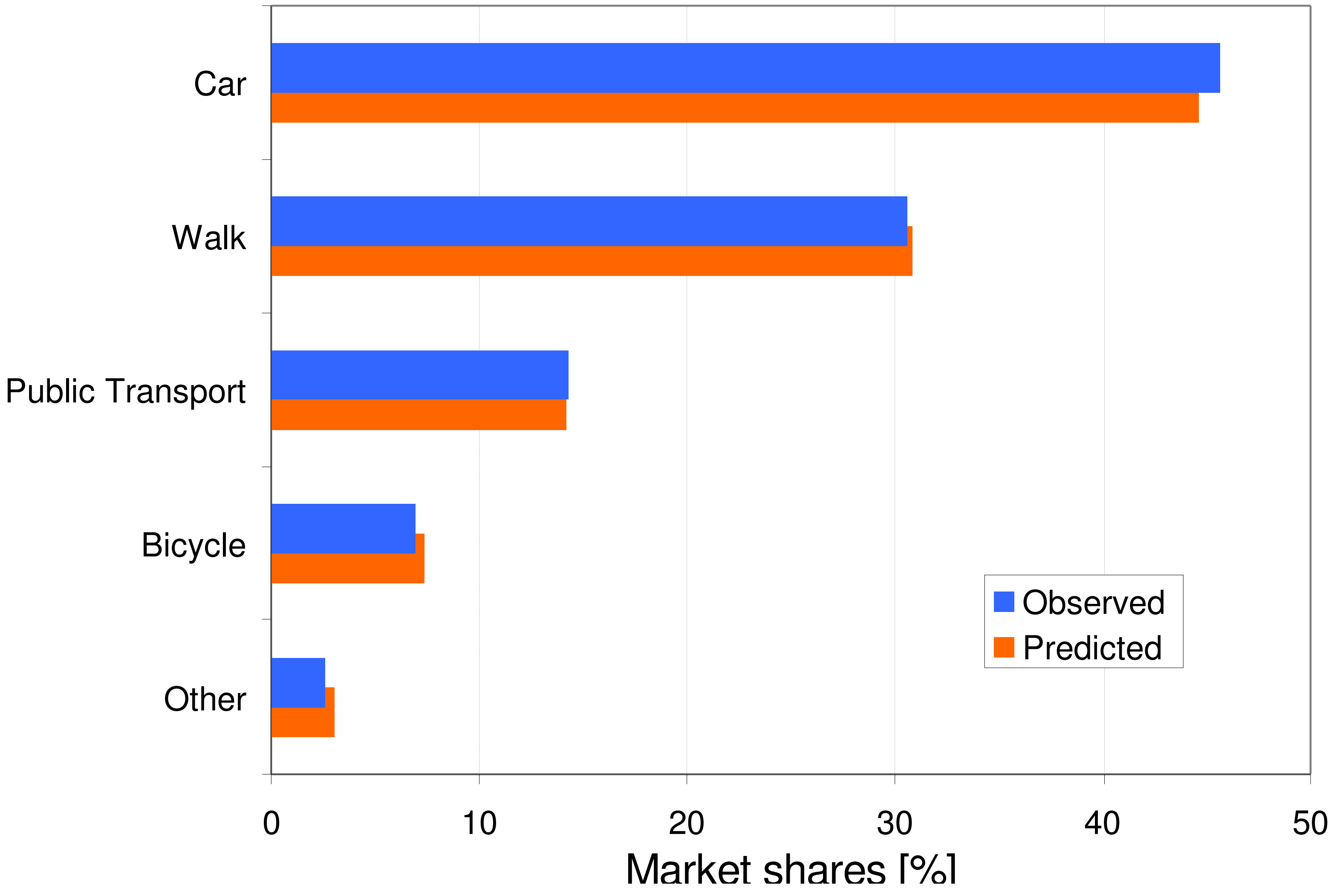
Data:

- MZ 2005

# Scenario: Mode choice – Observed mode public transport



# Scenario: Tour mode choice - Results



# Interim summary: Demand generation

- *Number and type of activities*
- *Sequence of activities*
  - Start and duration of activity
  - Composition of the group undertaking the activity
  - *Location of the activity*
    - Connection between sequential locations
      - Location of access and egress from the mean of transport
      - *Vehicle/means of transport*
      - Route/service
      - Group travelling together

# Progress on shortest – path calculation

Possibilities:

- **Bounding boxes**: Find out whether certain nodes can at all be on a shortest path and if not, do not consider them.
- **Multi-level approach**: Add shortcuts to the network where possible to bypass several edges at a time when routing.
- **Bi-directed search**: Start routing at the end and at the start node at the same time.
- **Goal-directed search**: Change the way the nodes are ranked such that nodes which are less likely to be on the shortest path are also less likely to be visited. The most popular algorithm that uses this technique is A*.

# Speed-up (1)

Basic algorithm: A*

Step 1: Improve data structure for list of candidate nodes (7% reduction)

Step 2: Improve handling of the “visited flag” (~ length of route)

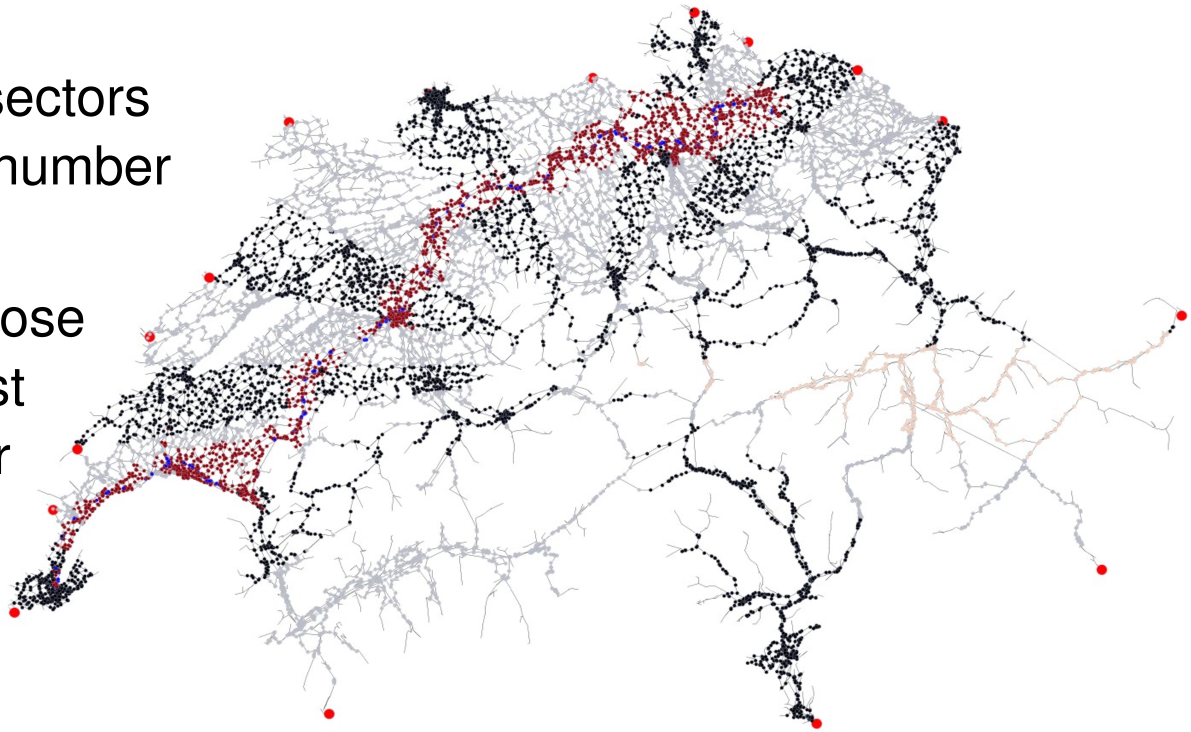
Step 3: Detect dead ends (1/4 of nodes in the NavTeq network) (50% reduction)

Step 4: Use euclidian-distance to destination to rank candidates (50-80% reduction)

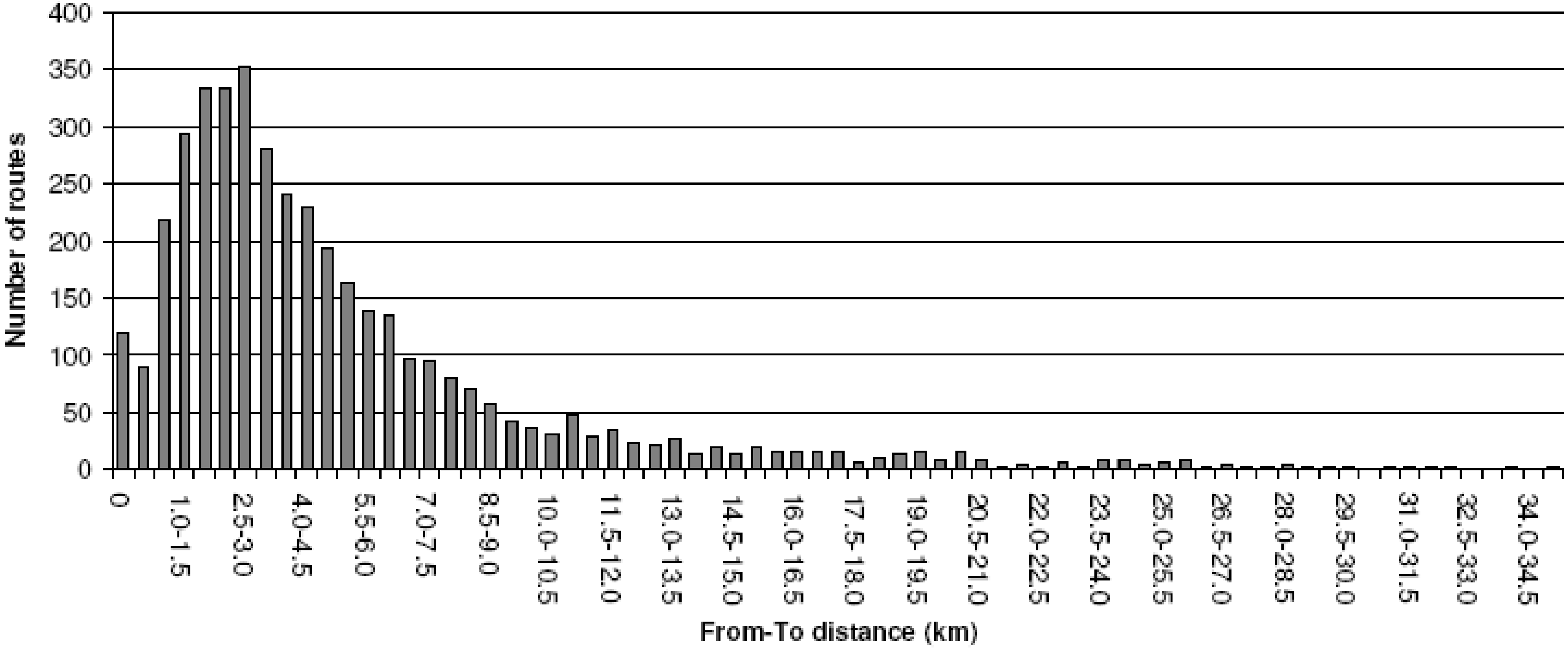
Step 5: Use intermediate landmarks (80-90% reduction)

# Speed-up: Landmark selection

- Divide network into sectors containing an equal number of nodes
- For each sector, choose a node that is farthest away from the center
- Check that the landmarks are not too close to each other. If so, narrow one sector and choose a new landmark within the sector



# Speed up of shortest-path calculations




Free flow conditions: 97-99% reductions

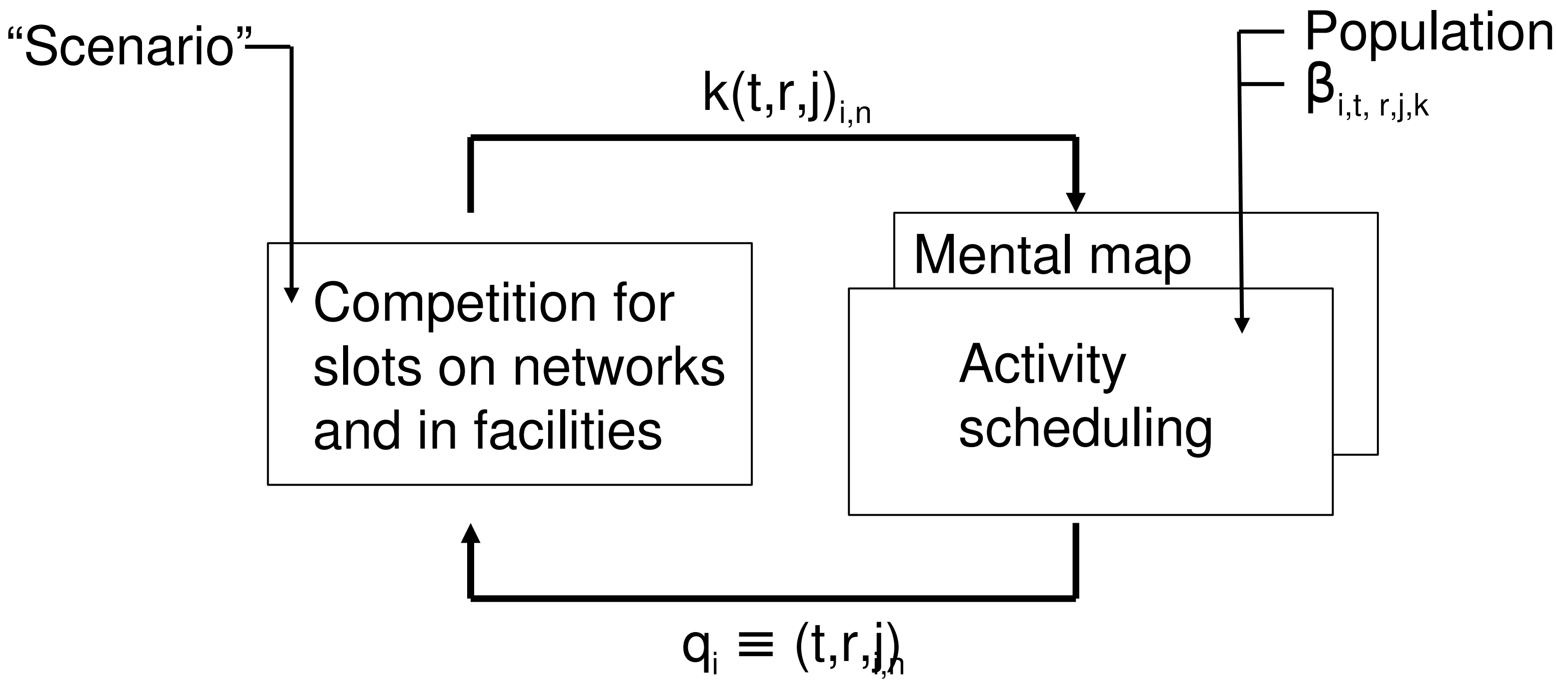
Loaded conditions: 95% reductions

on navigation networks

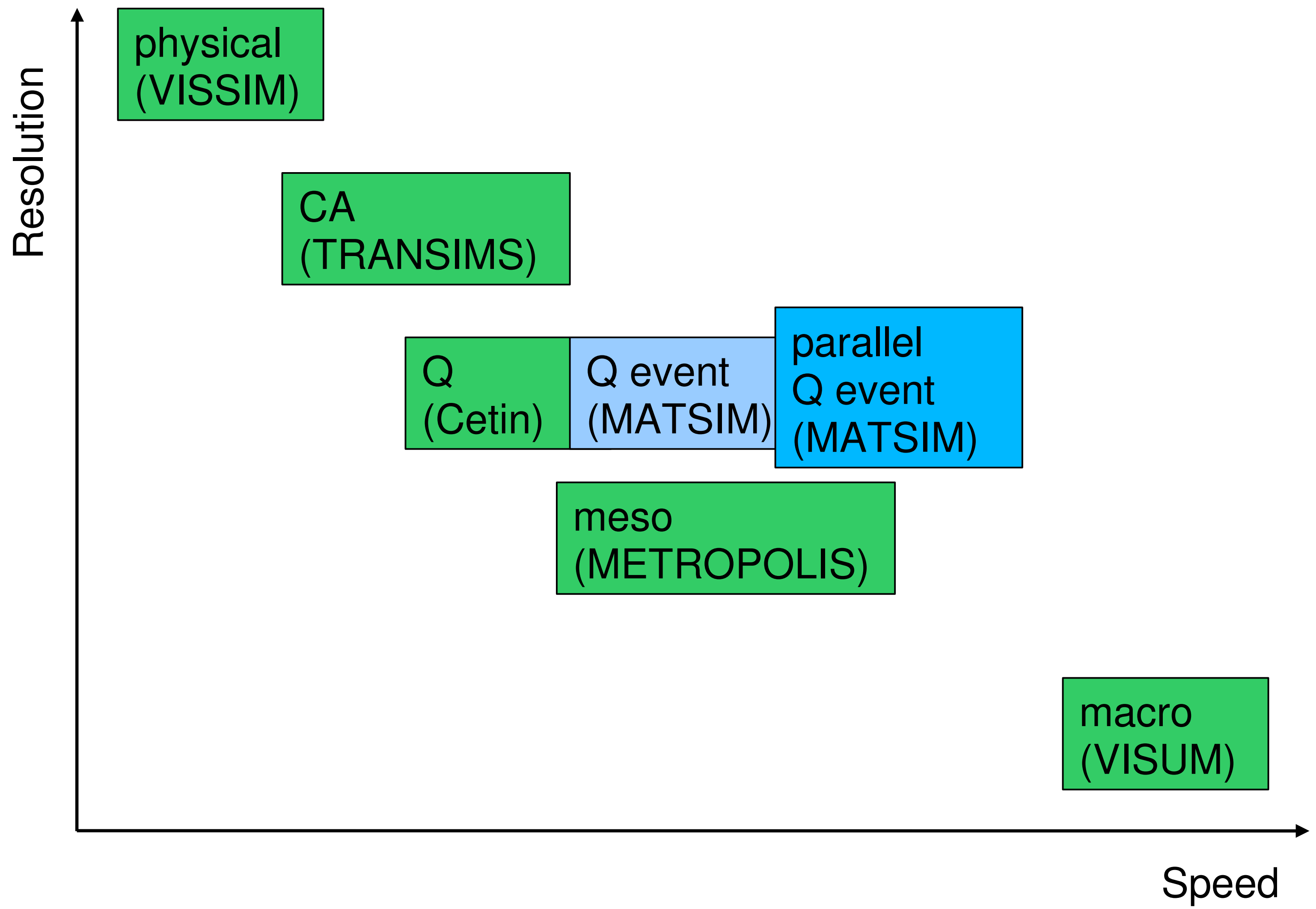
# Interim summary: Plan


```xml
<person id="22018">
    <plan score="157.72" selected="yes">
        <act type="h" x100="703600" y100="236900" link="5757"
                                                    end_time="07:35:04" />
        <leg num="0" mode="car" dep_time="07:35:04" trav_time="00:16:31">
            <route>1900 1899 1897</route>
        </leg>
        <act type="w" x100="702500" y100="236400" link="5749" dur="08:12:05" />
        <leg num="1" mode="car" dep_time="16:03:40" trav_time="01:10:22">
            <route>1899 1848 1925 1924 1923 1922 1068</route>
        </leg>
        <act type="l" x100="681450" y100="246550" link="2140" dur="01:20:00" />
        <leg num="2" mode="car" dep_time="" trav_time="00:34:35">
            <route>1067 1136 1137 1921 1922 1923 1925 1848 1899</route>
        </leg>
        <act type="h" x100="703600" y100="236900" link="5757" />
    </plan>
</person>
```

# Traffic flow model



# Approach

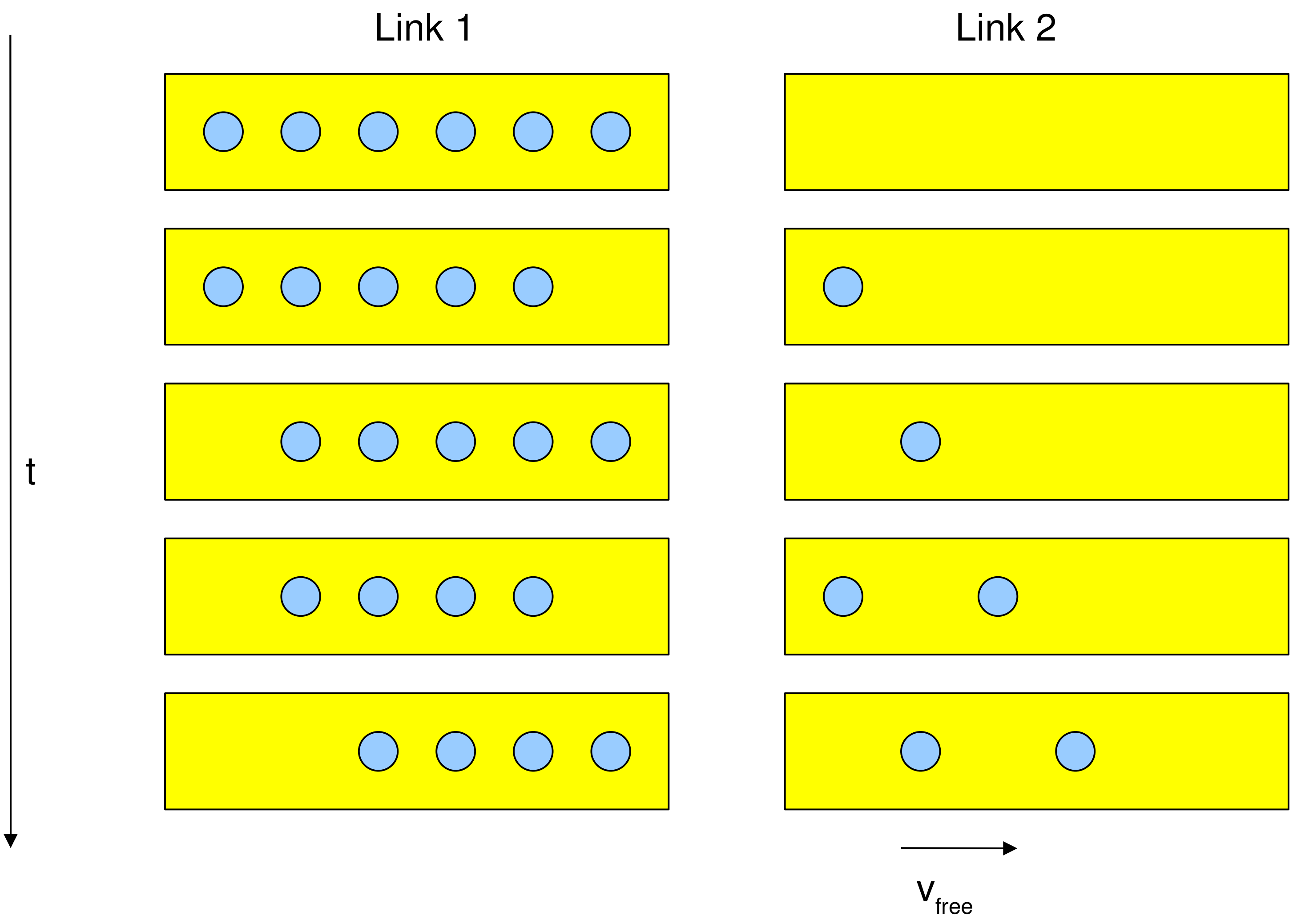


Resolution
physical (VISSIM)
CA (TRANSIMS)
Q (Cetin)
Q event (MATSIM)
parallel Q event (MATSIM)
meso (METROPOLIS)
macro (VISUM)
Speed

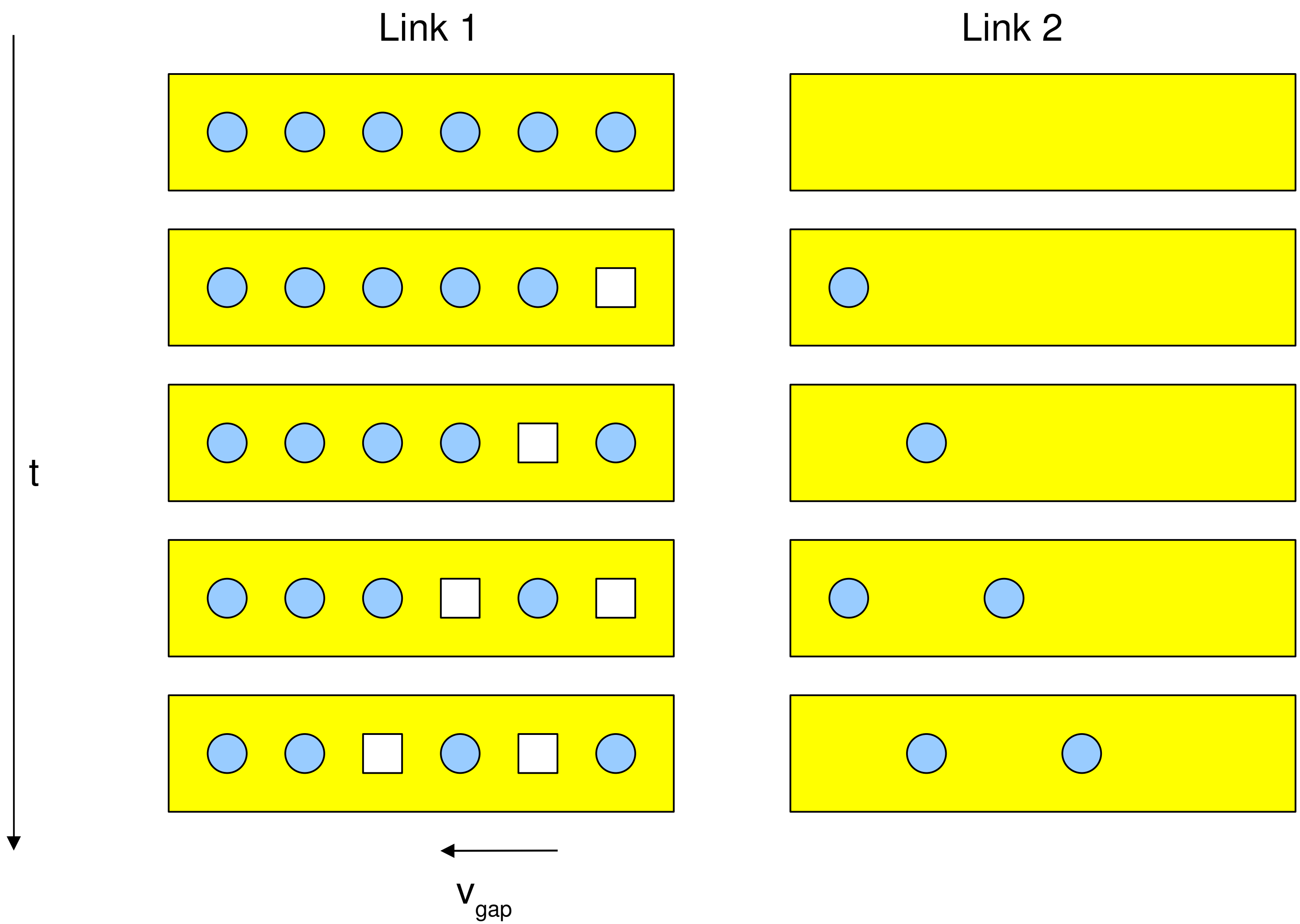
# Parallel q-event driven simulation with gaps

- Approach
- Fundamental diagram (on a ring motorway)
- Domain decomposition
- Test

# Q-event: Approach without gaps



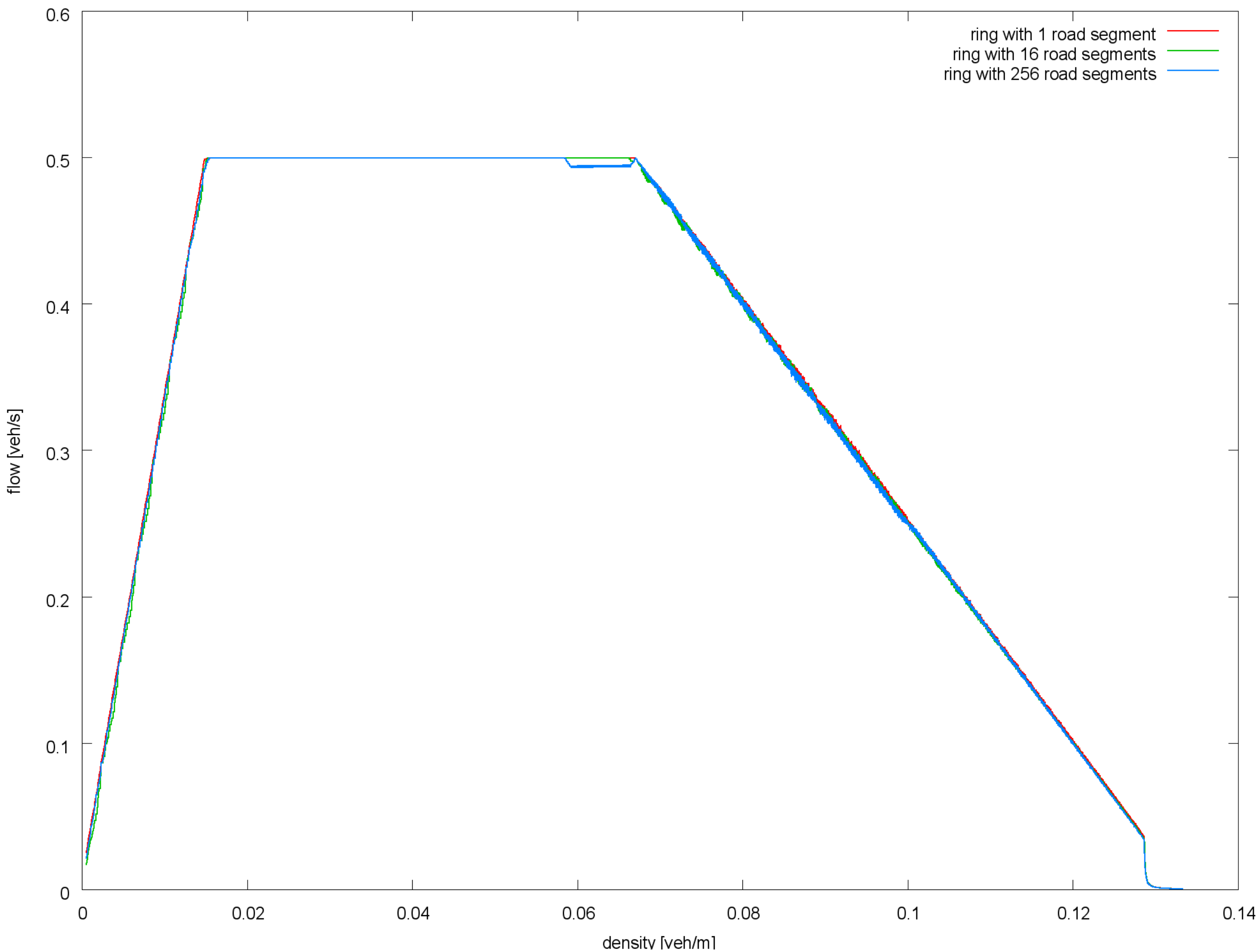
# Q-event: Approach with gaps



# Q-event: Implementation details

- Squeezing to avoid grid-lock
- Inflow capacity = 110% of outflow capacity (1800 veh/h* lanes)
- Vehicles are served in order of arrival at the junctions
- C++ with binary data interface to MATSim-T

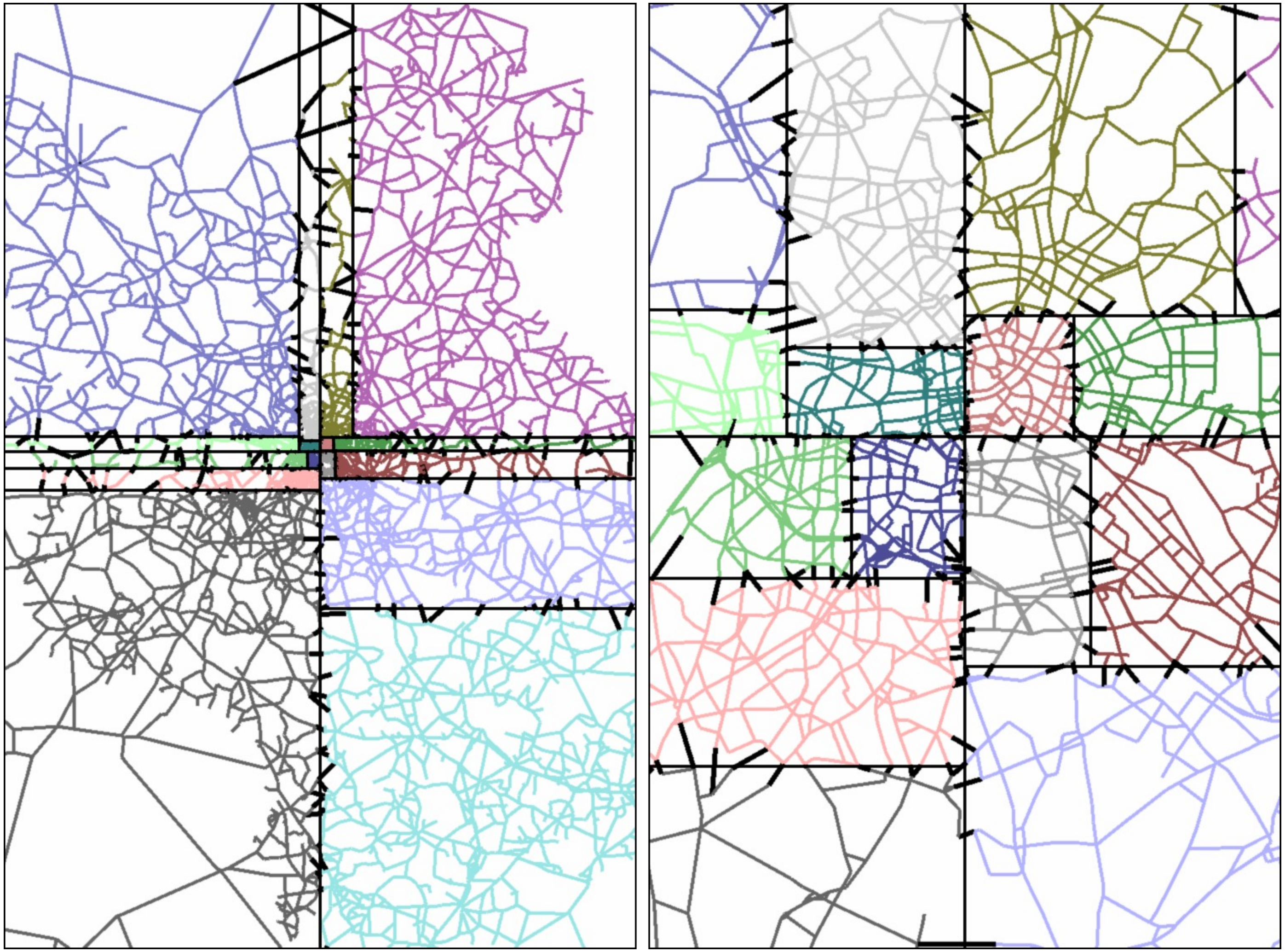
# Q-event: Fundamental diagram



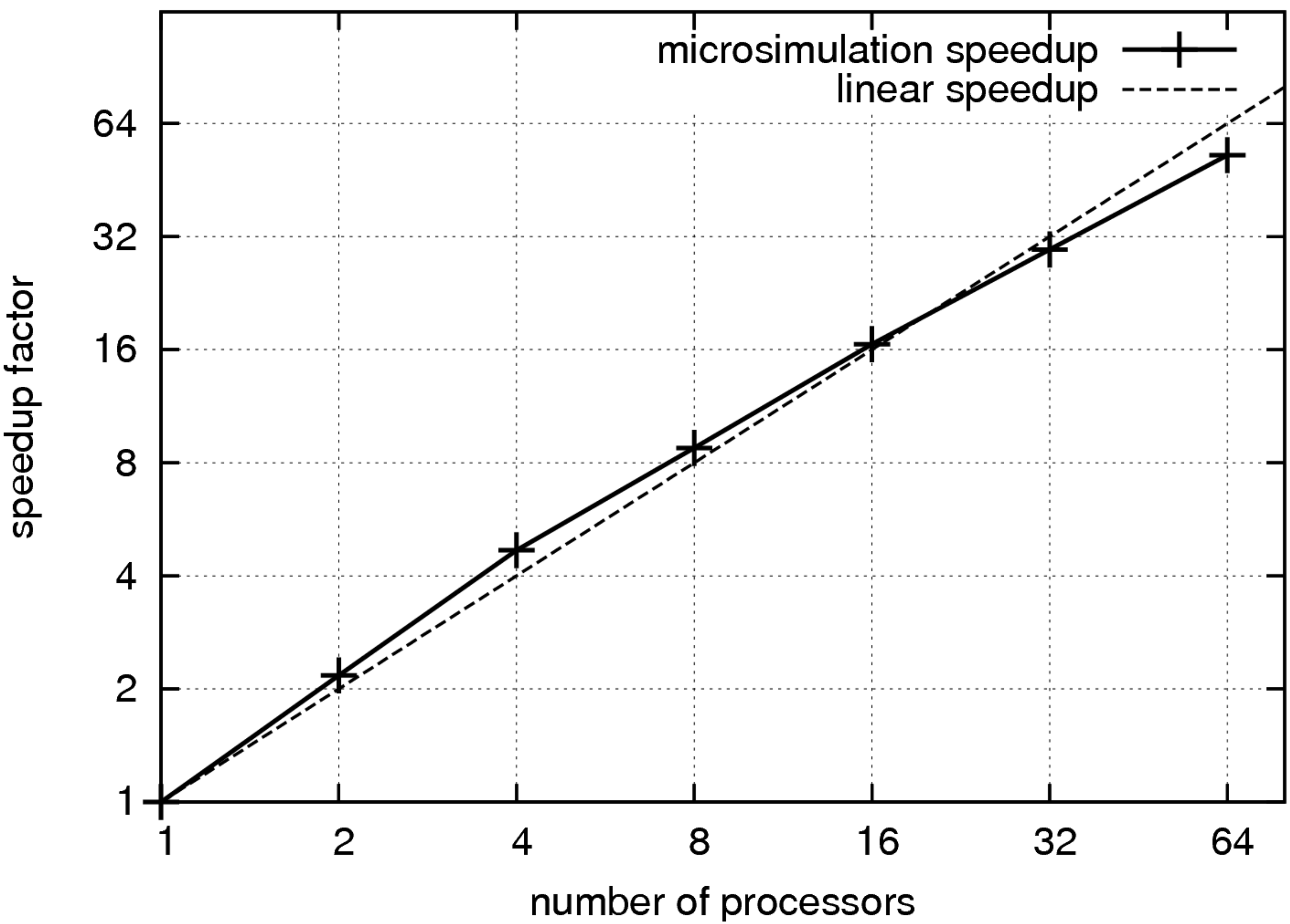
# Q-event: Test setup

- Road network of the federal states of Germany Berlin and Brandenburg
- 11.6k nodes, 27.7k links
- 7.05M person days
- Average number of trips per agent: 2.02
- Average length of a trip: 17.5 links
- Total: 249M road segments to be traveled
- 77 min on a single dual-core CPU (1.6GH; 256 GB RAM)

# Q-event: Integrated domain decomposition



# Q-event: Parallelisation



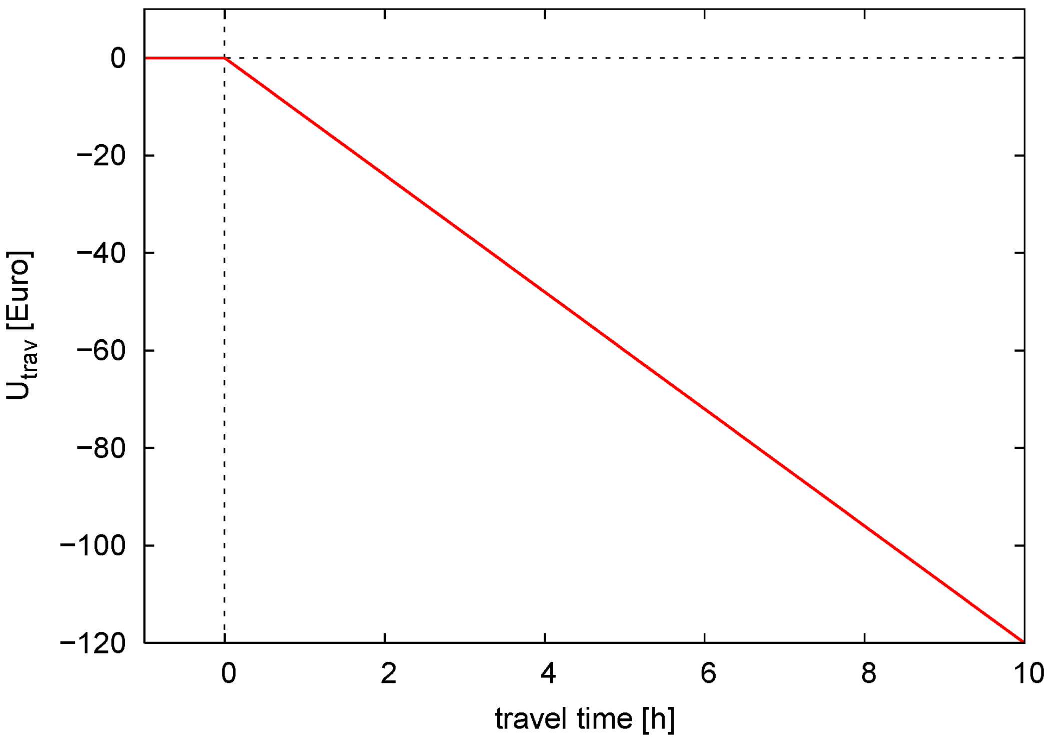
# Scheduling and its utility function

# Utility function: Individual schedules

$$U_{plan} = \sum_{i=1}^{n} U_{act,i} + \sum_{i=2}^{n} U_{trav,i-1,i}$$

$$U_{act,i} = U_{dur,i} + U_{wait,i} + U_{late.ar,i} + U_{early.dp,i} + U_{short.dur,i}$$

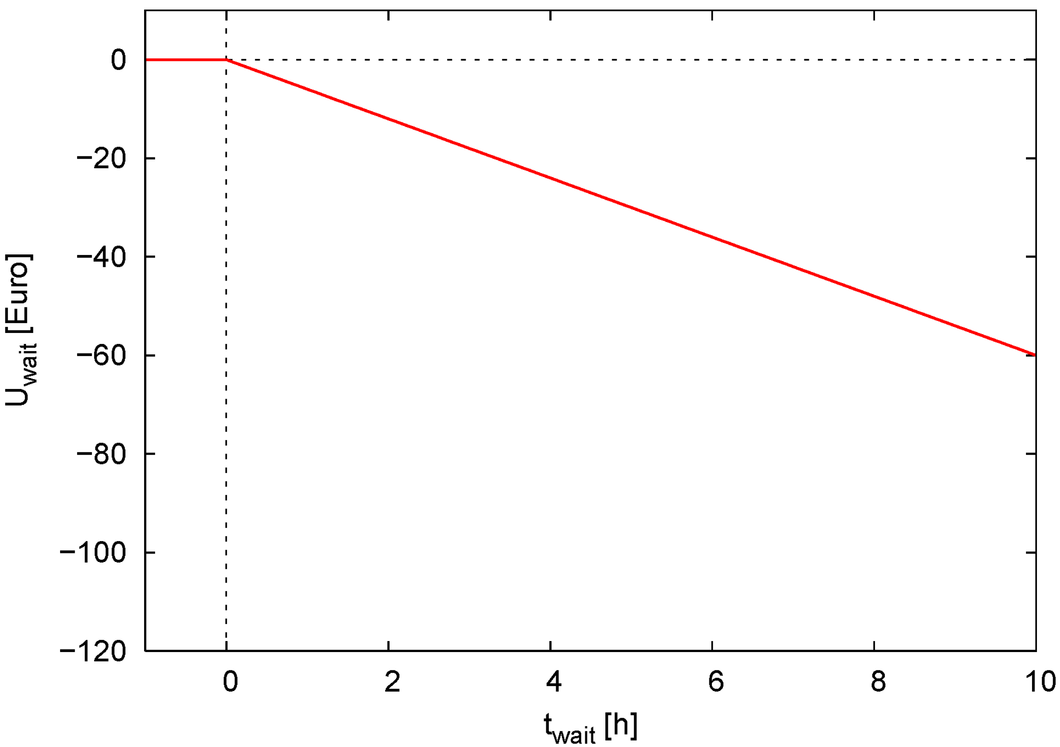
# Utility function: Travel




$$\beta_{trav} = -12\ Euro/h$$

$$U_{trav,i-1,i} = \begin{cases} \beta_{trav} \cdot t_{trav,i-1,i} & if\ t_{trav,i-1,i} \geq 0 \\ 0 & else \end{cases}$$

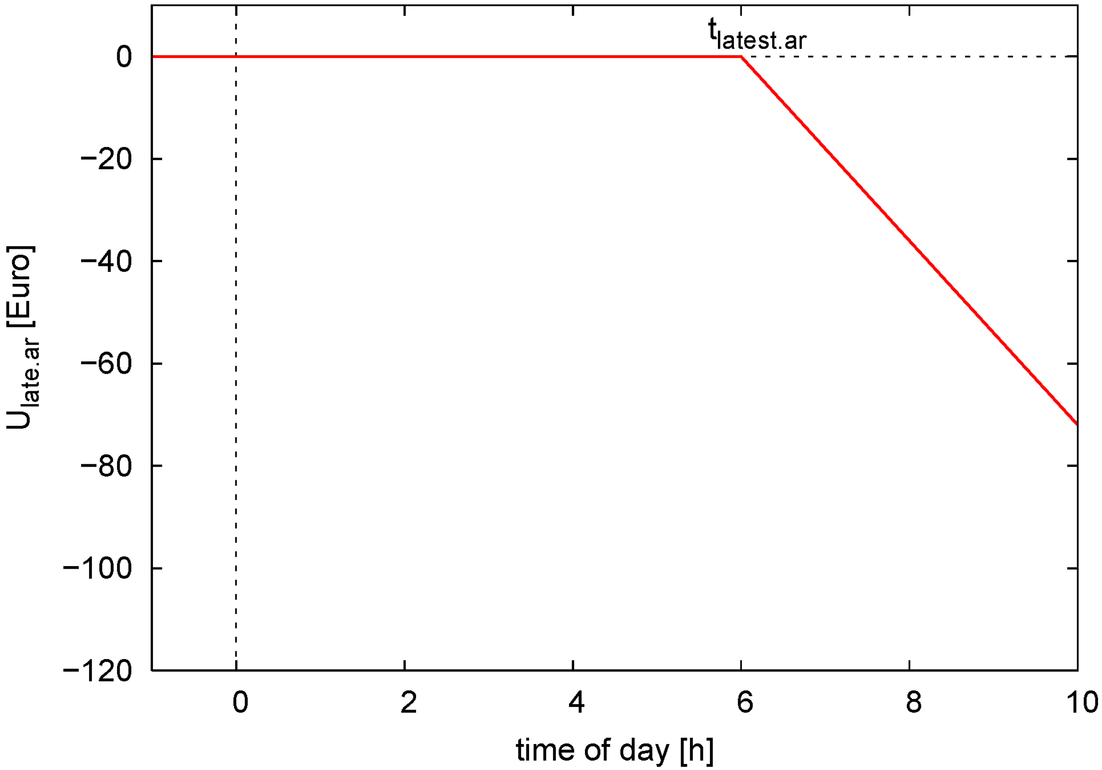
# Utility function: Waiting




$$\beta_{wait} = -6 \ Euro/h$$

$$U_{wait,i} = \begin{cases} \beta_{wait} \cdot t_{wait,i} & if \ t_{wait,i} \geq 0 \\ 0 & else \end{cases}$$

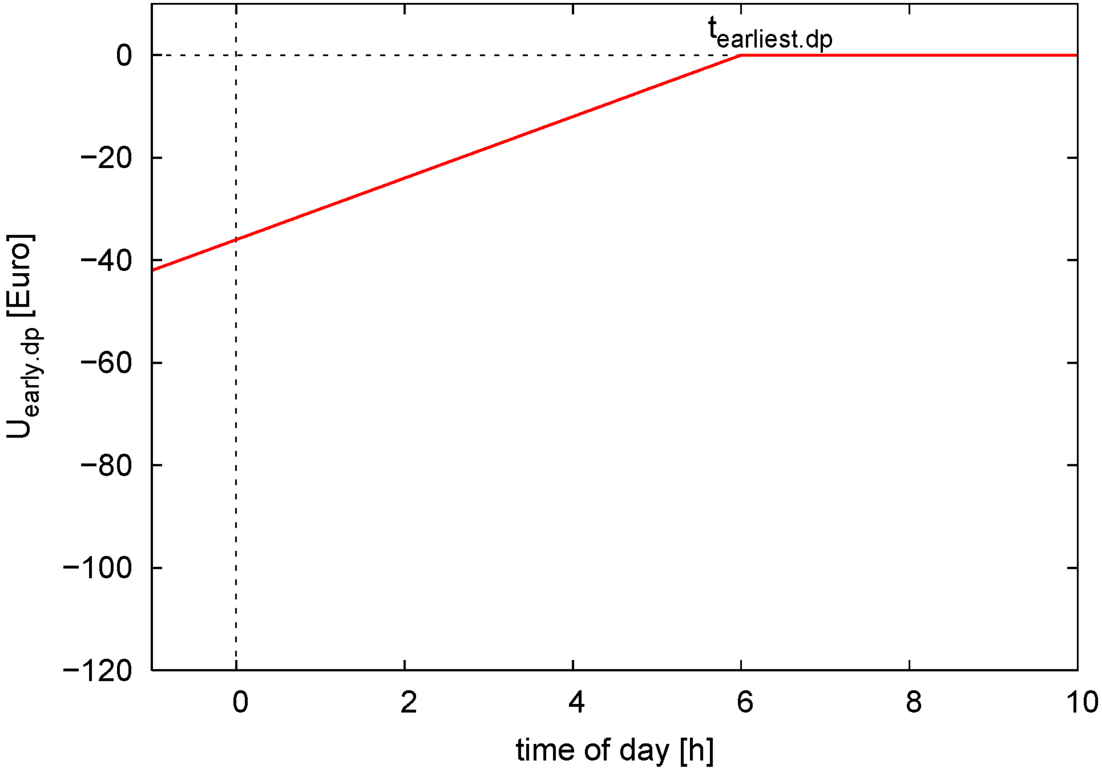
# Utility function: Late arrival




$$\beta_{late.ar} = -18 \ Euro/h$$

$$U_{late.ar,i} = \begin{cases} \beta_{late.ar} \cdot \left( t_{start,i} - t_{latest.ar,i} \right) & if \ t_{start,i} \geq t_{latest.ar,i} \\ 0 & else \end{cases}$$

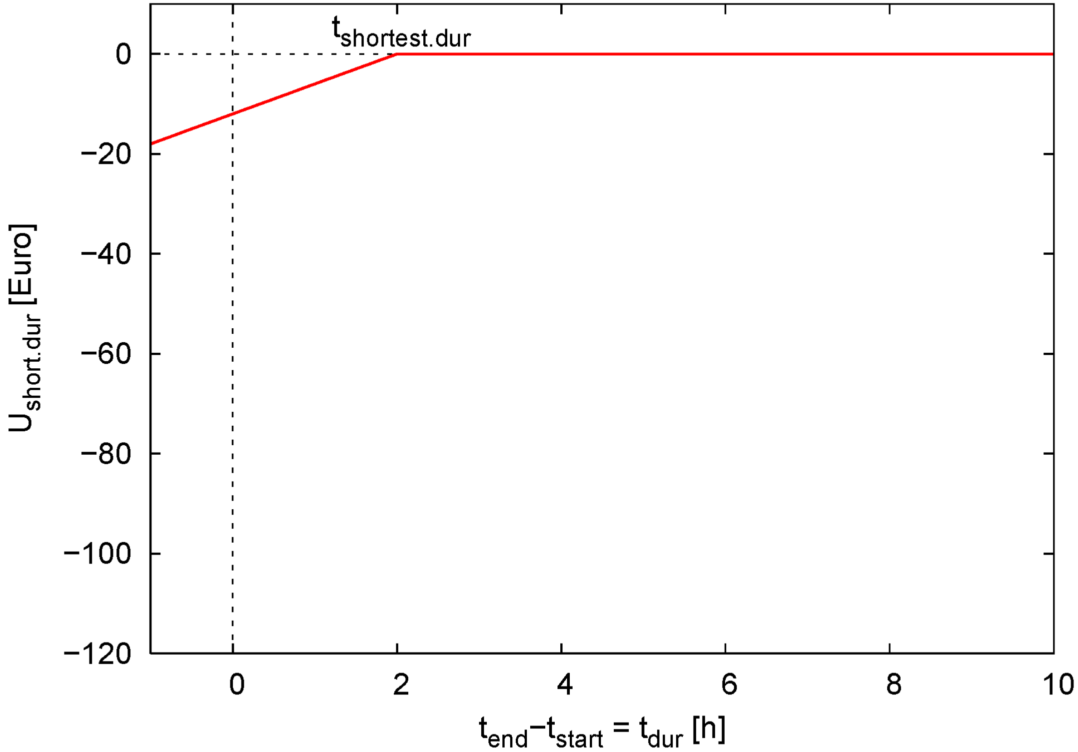
# Utility function: Departing early




$$\beta_{early.dp} = -6 \ Euro/h$$

$$U_{early.dp,i} = \begin{cases} \beta_{early.dp} \cdot \left( t_{earliest.dp,i} - t_{end,i} \right) & if \ t_{end,i} \leq t_{earliest.dp,i} \\ 0 & else \end{cases}$$

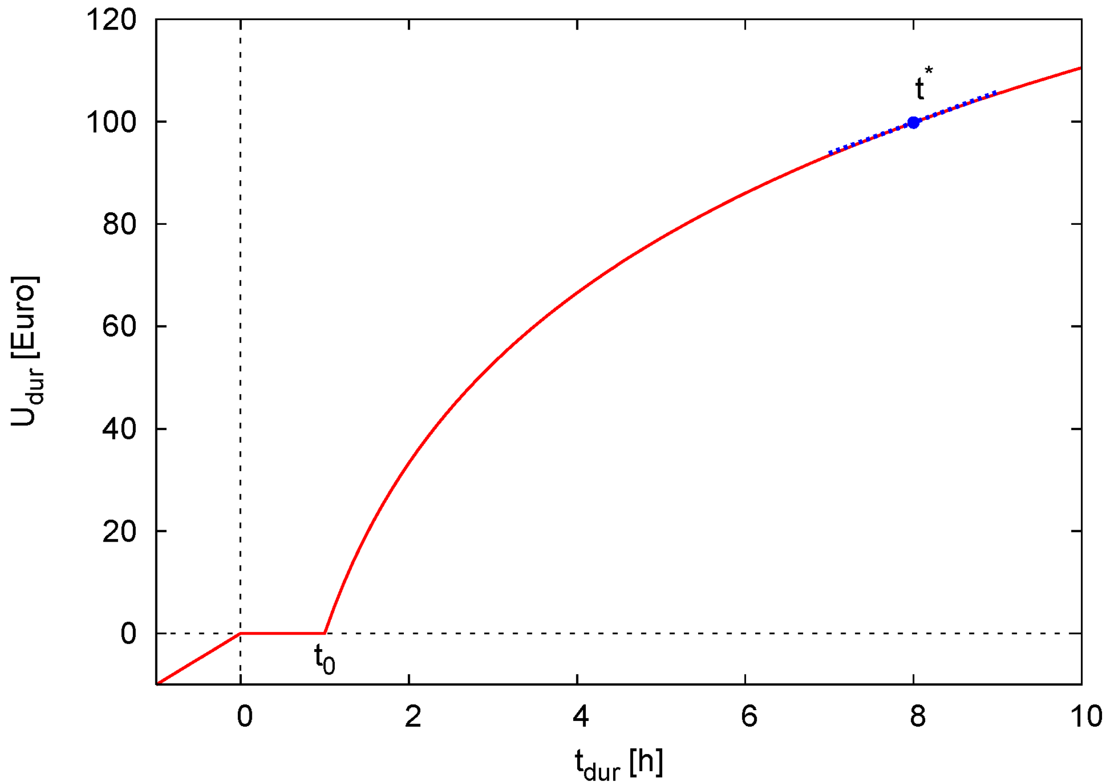
# Utility function: Duration too short




$$\beta_{short.dur} = -6\ Euro/h$$

$$U_{short.dur,i} = \begin{cases} \beta_{short.dur} \cdot \left( t_{shortest.dur,i} - \left( t_{end,i} - t_{start,i} \right) \right) & if\ \ t_{end,i} - t_{start,i} \leq t_{shortest.dur,i} \\ 0 & else \end{cases}$$

# Utility function: Activity performance

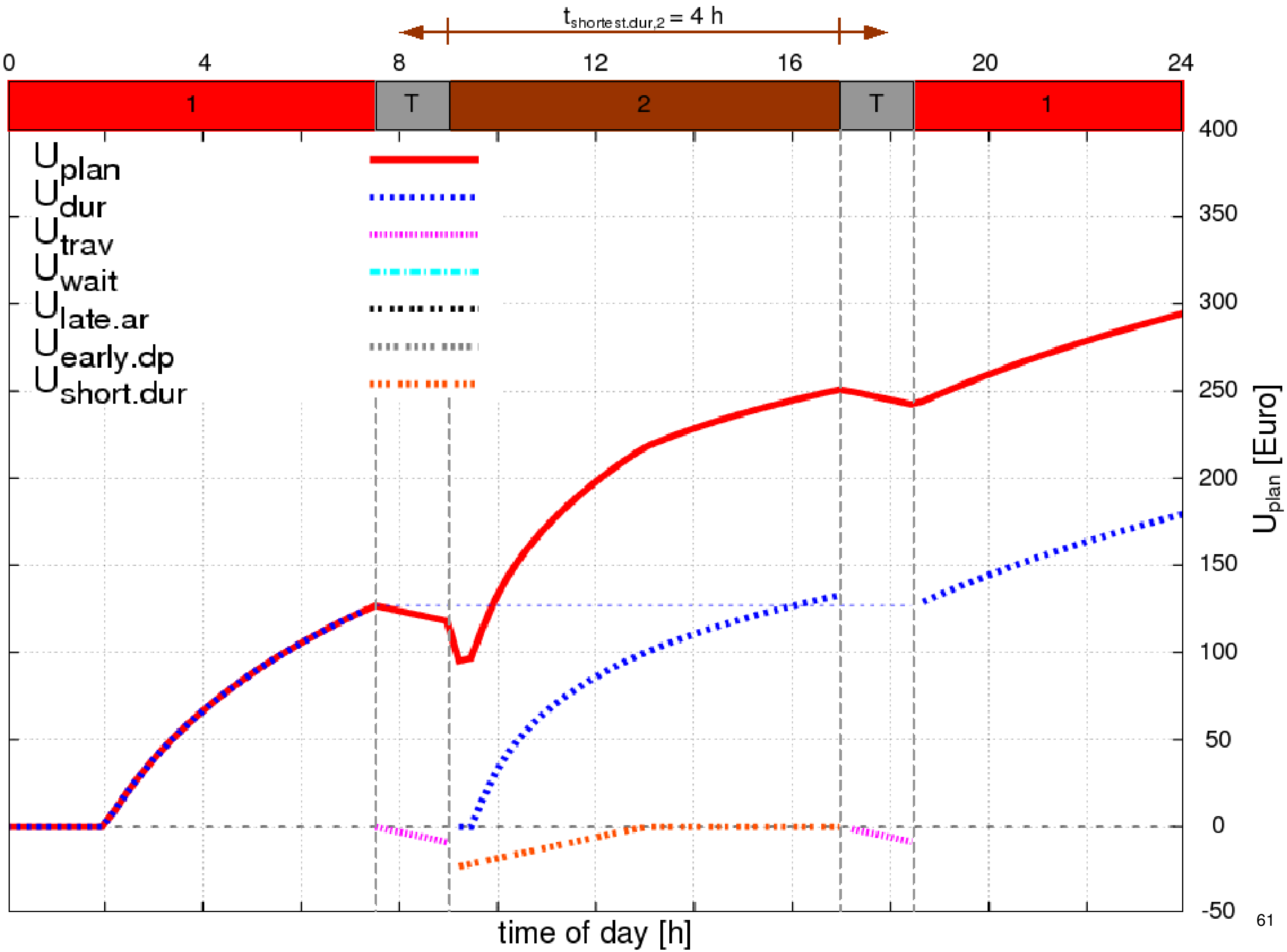



$\alpha$ 墹 $= 10\ Euro/h$

$\beta_{dur}$ 墹 $= 6\ Euro/h$

$$U_{dur,i} = \begin{cases} \beta_{dur} \cdot t^{¿} \cdot \ln\left(t_{dur,i} / t_{0,i}\right) & if \ \ t_{0,i} \leq t_{dur,i} \\ 0 & if \ \ 0 \leq t_{dur,i} < t_{0,i} \\ \alpha \cdot t_{dur,i} & else \ \left(\alpha > 0\right) \end{cases}$$

# Utility function: Home-Work-Home example



# Utility function: Revising the parameters

$\beta_{dur}$ = 6 €/h, $\beta_{wait}$ = 0 €/h,
$\beta_{trav}$ = –6 €/h, $\beta_{late.ar}$ = –18 €/h
$\beta_{early.dp}$ = 0 €/h, $\beta_{short.dur}$ = 0 €/h,

$\beta_{neg.dur}$ = –18 €/h.

Vickrey model ratio is 1:2:3:

$$\beta_{wait} : \beta_{trav} : \beta_{late.ar} = 0 : -6 : -18$$

Considering the opportunity costs of *not* performing an activity while waiting or travelling, one has to subtract $\beta_{dur}$:

$$\beta_{wait,eff} : \beta_{trav,eff} : \beta_{late.ar,eff} = -6 : -12 : -18$$

# Scheduling: Current Planomat(s)

Version 1:

- GA optimiser of durations and starting times
- Retains time-of-day profile of generalised costs

Version 2:

- CMA-ES (Covariance Matrix Adaptation Evolutionary strategy) of durations and starting times
- Retains time-of-day profile of generalised costs

# Scheduling: Once and future Planomat

- *Number and type of activities* out of an agenda
- *Sequence of activities*
  - *Start and duration of activity*
  - *Composition of the* household group undertaking the activity
  - *Location of the activity*
    - *Connection between sequential locations*
      - *Location of access and egress from the mean of transport*
      - *Vehicle/means of transport*
      - Route/service
      - *Group travelling together*

# Improving the convergence

# Initial approach

Method:

- Plan everybody at Iteration 1
- Replan for a fixed share

Convergence:

- On mean performance

# Improved approach

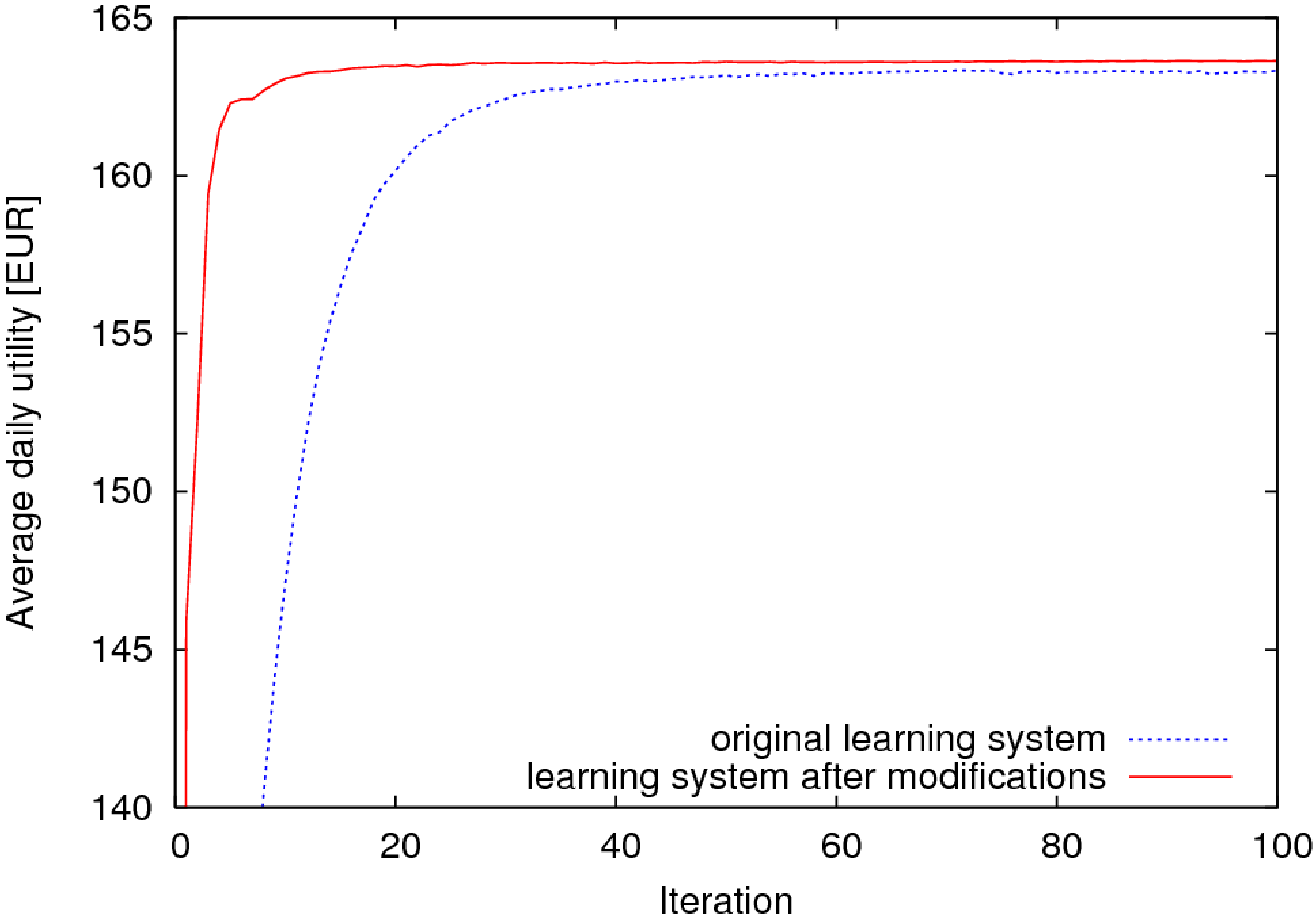
Method:

- Plan for everyone at Iteration 1
- Replan for a fixed, but decreasing share over the number of iterations

Measurement:

- Aggregates

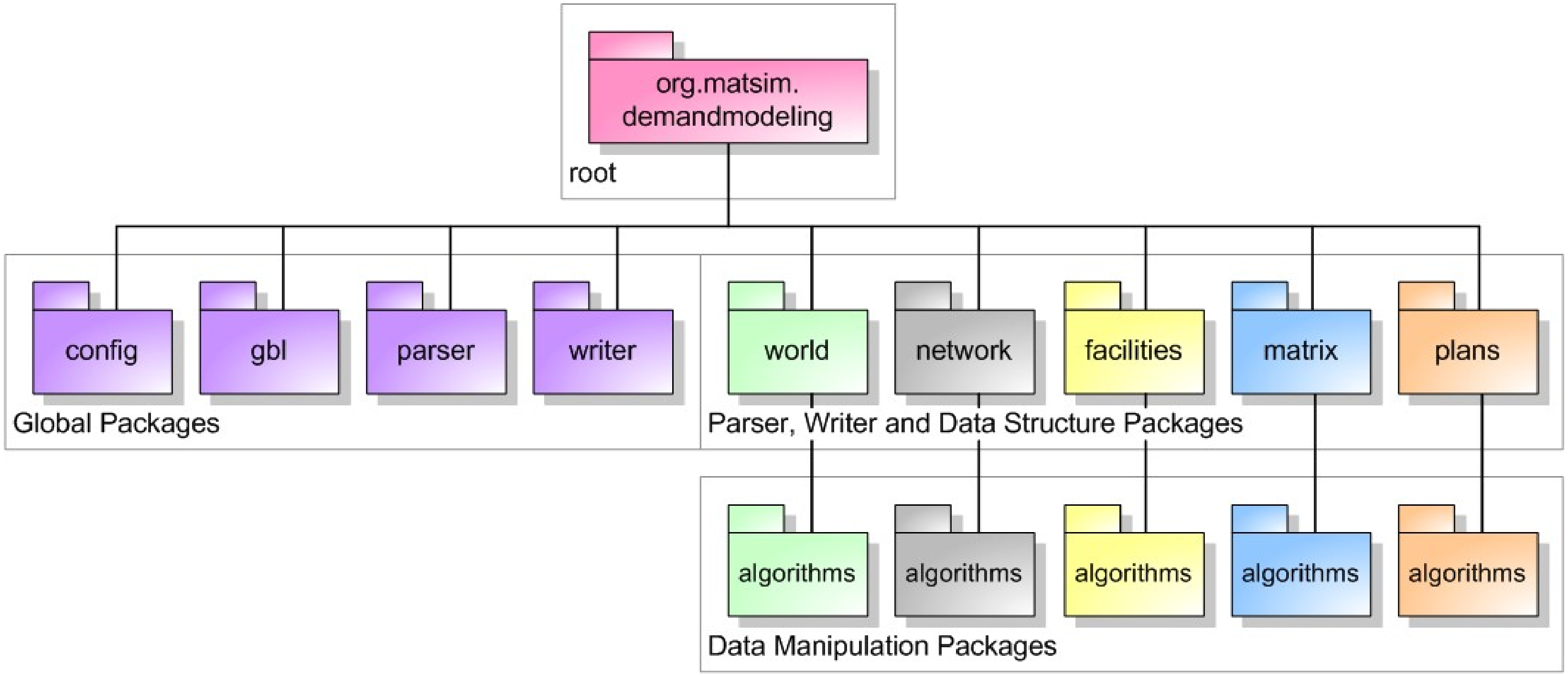
# Impact of improved approach



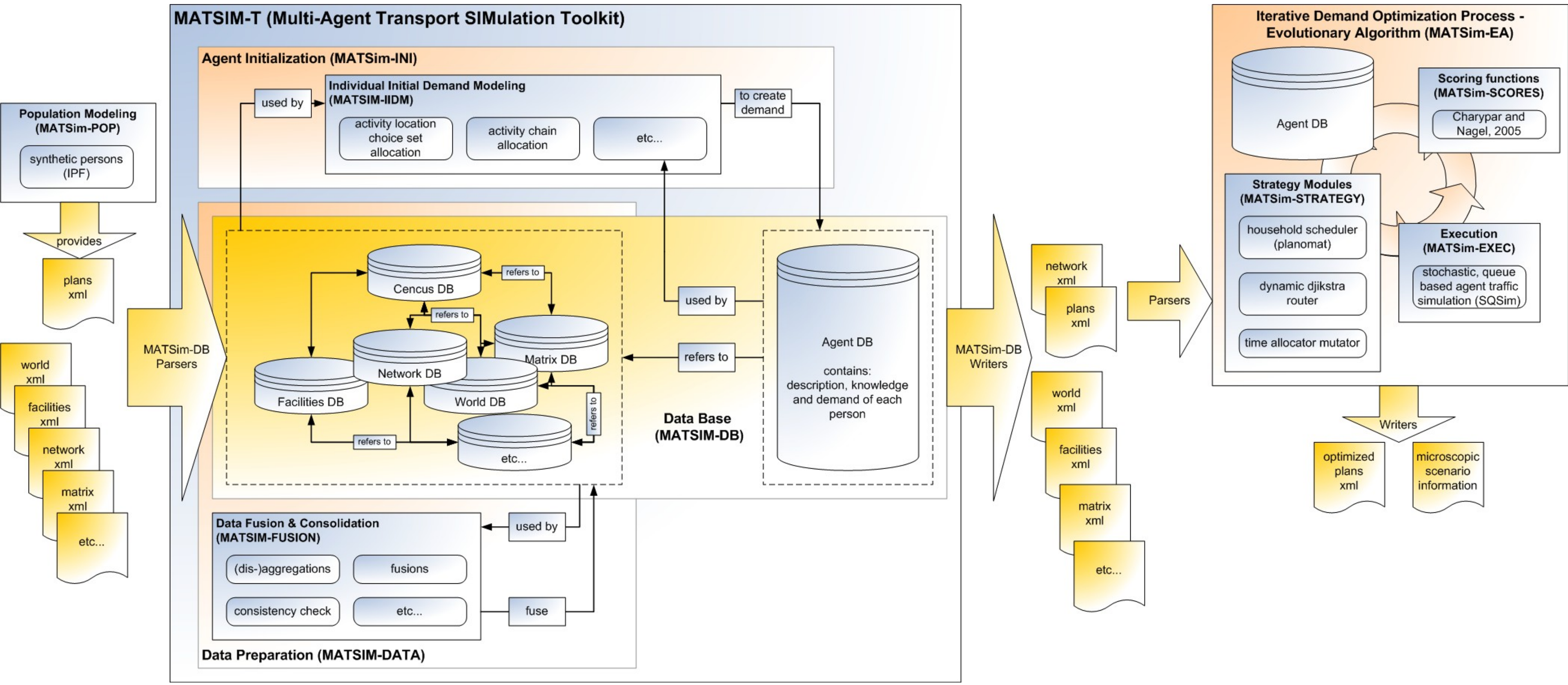
# System architecture

- XML standards
- Data handling tools (aggregation/disaggregation)
- Data base
- Iteration handling and control of convergence
- Models

# Architecture: JAVA - packages



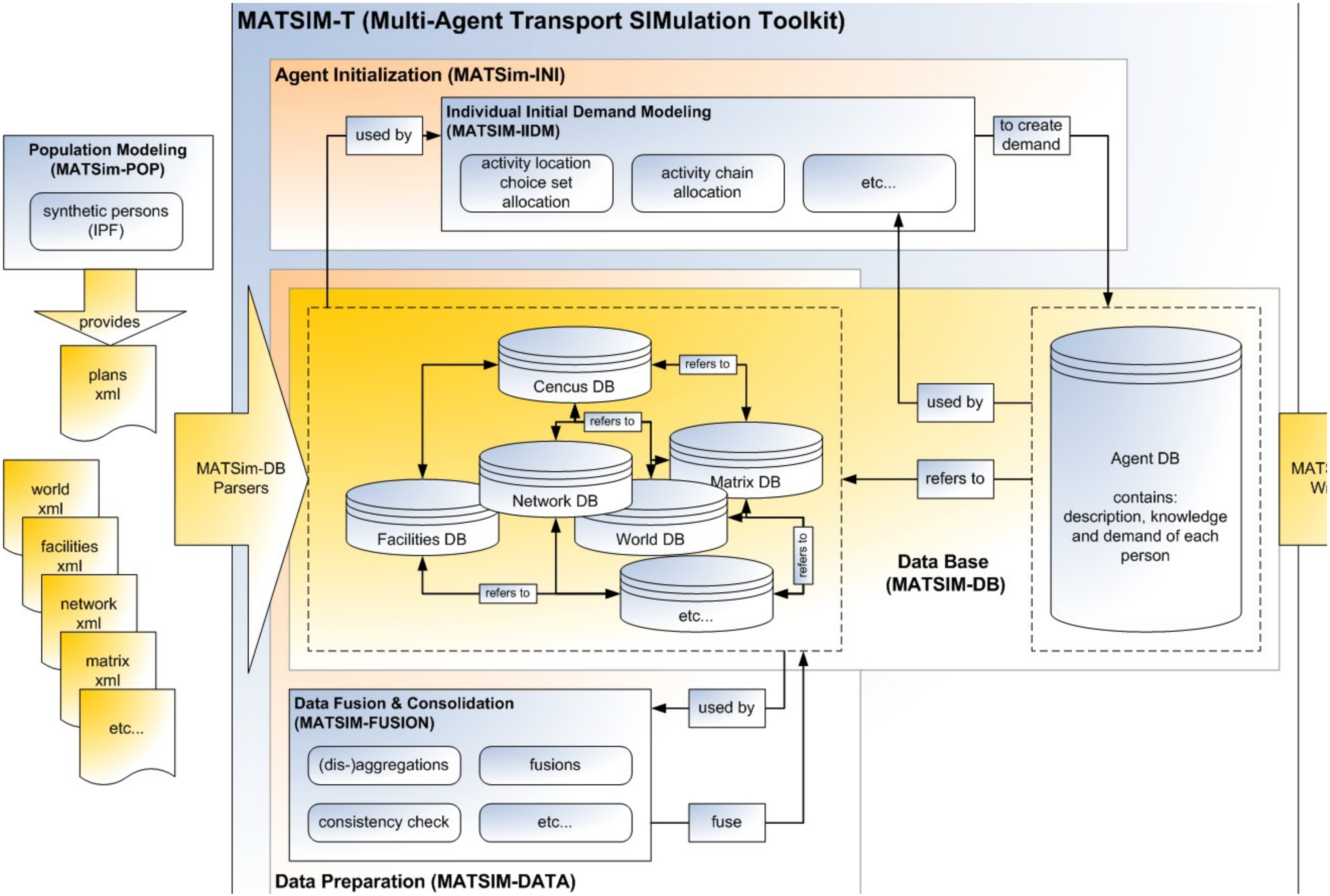
# Architecture: Data flow



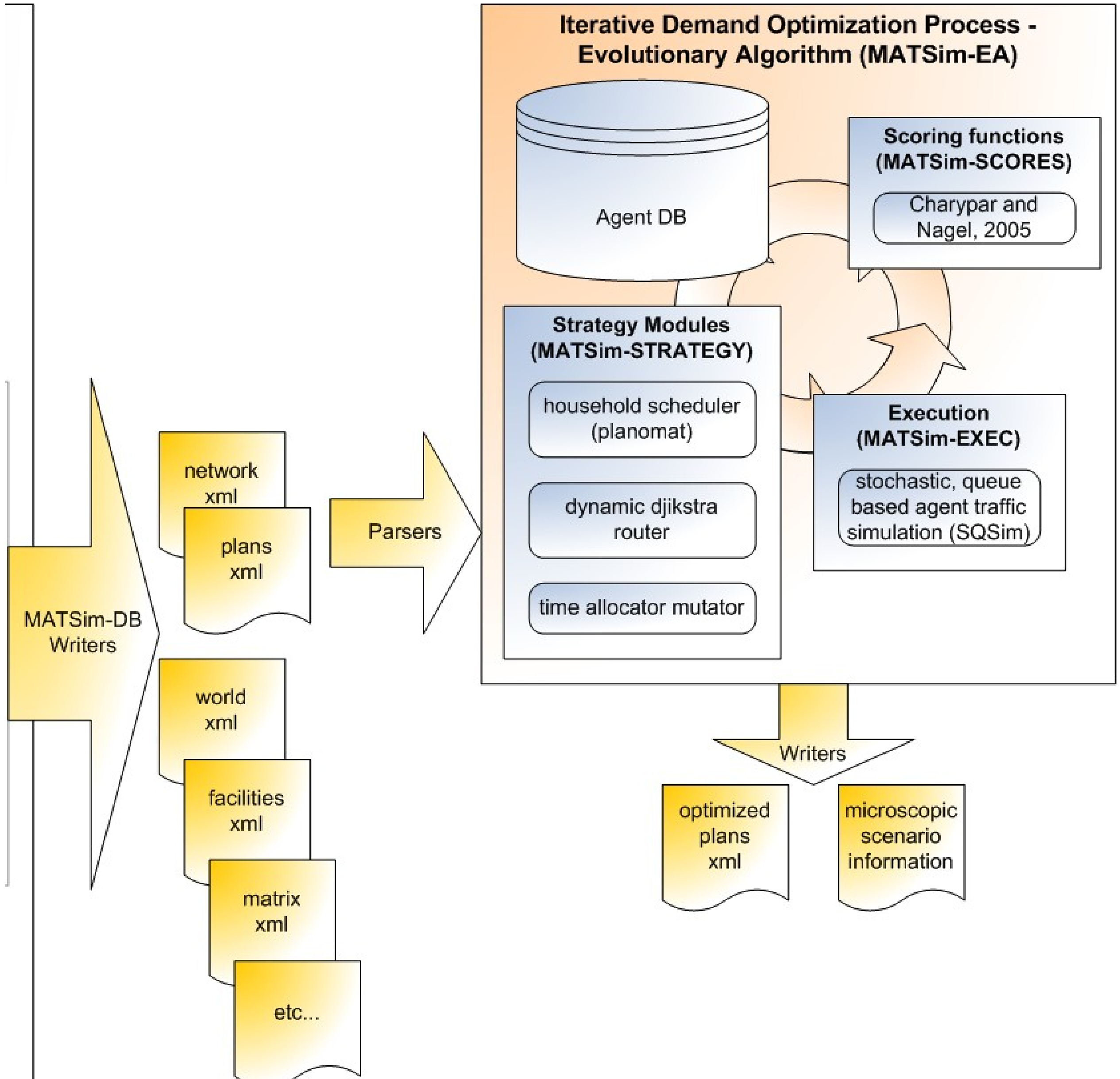

JAVA

C++

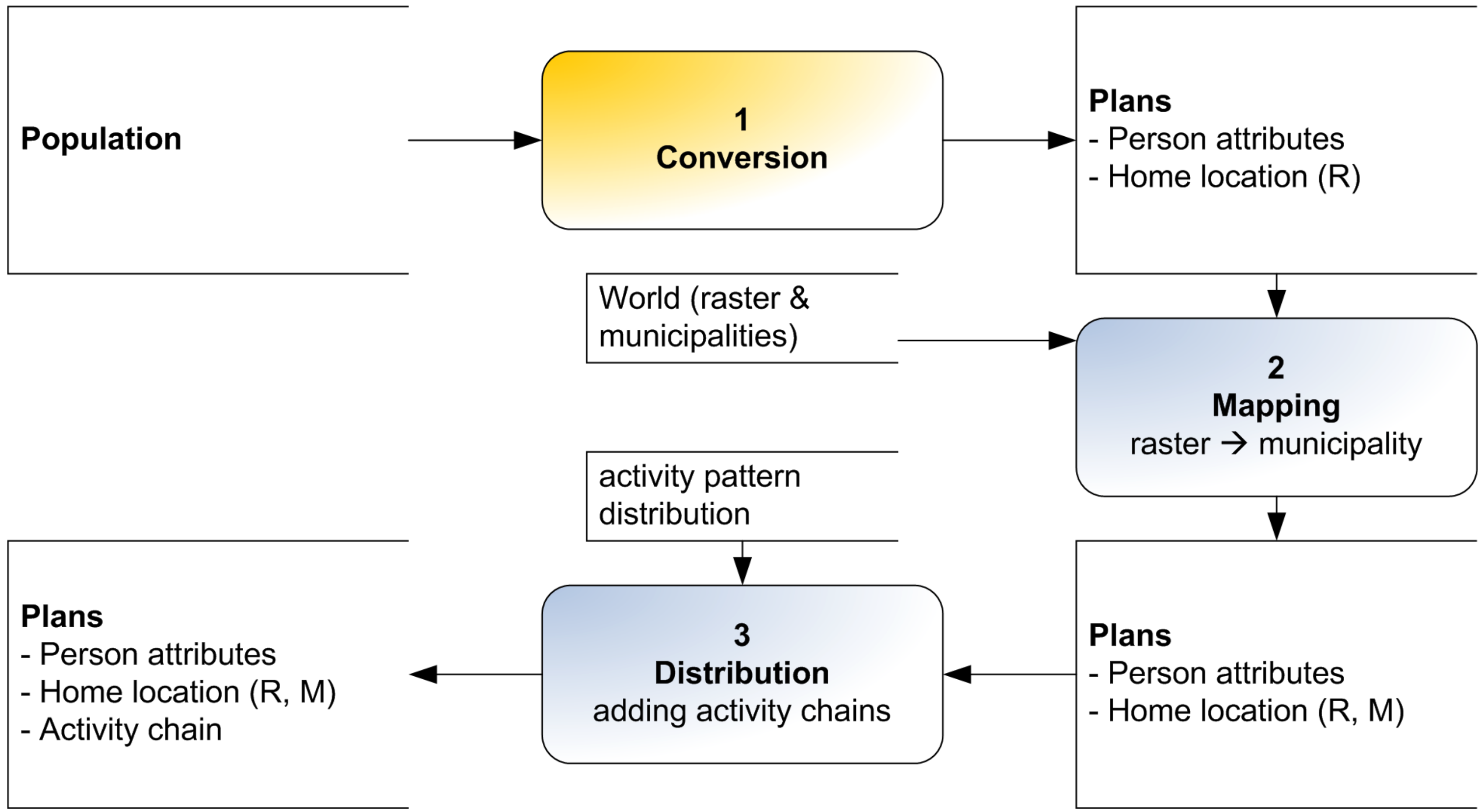
# Architecture: Data flow



# Architecture: Data flow

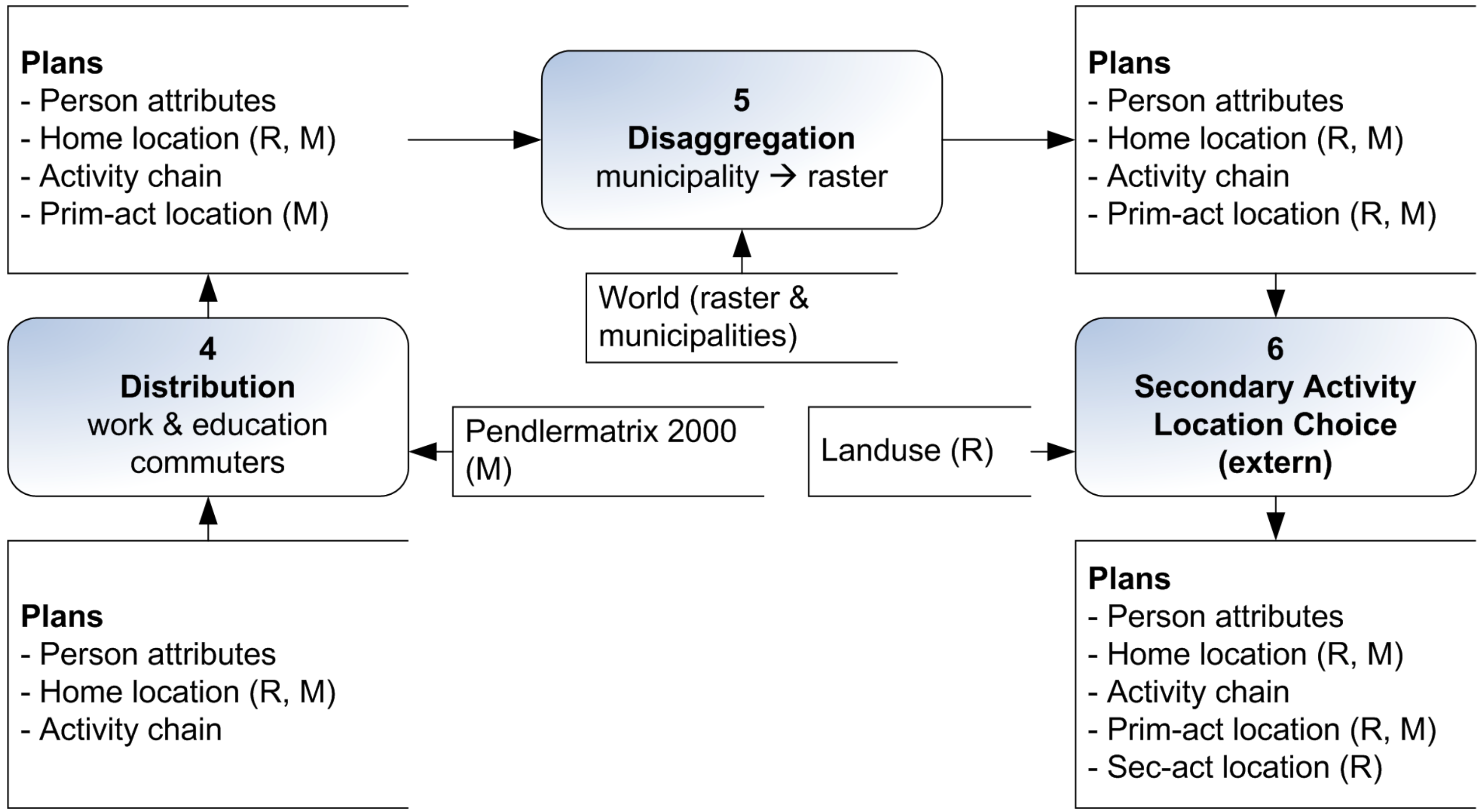


# What the tools allow: Kt. Zürich (1)

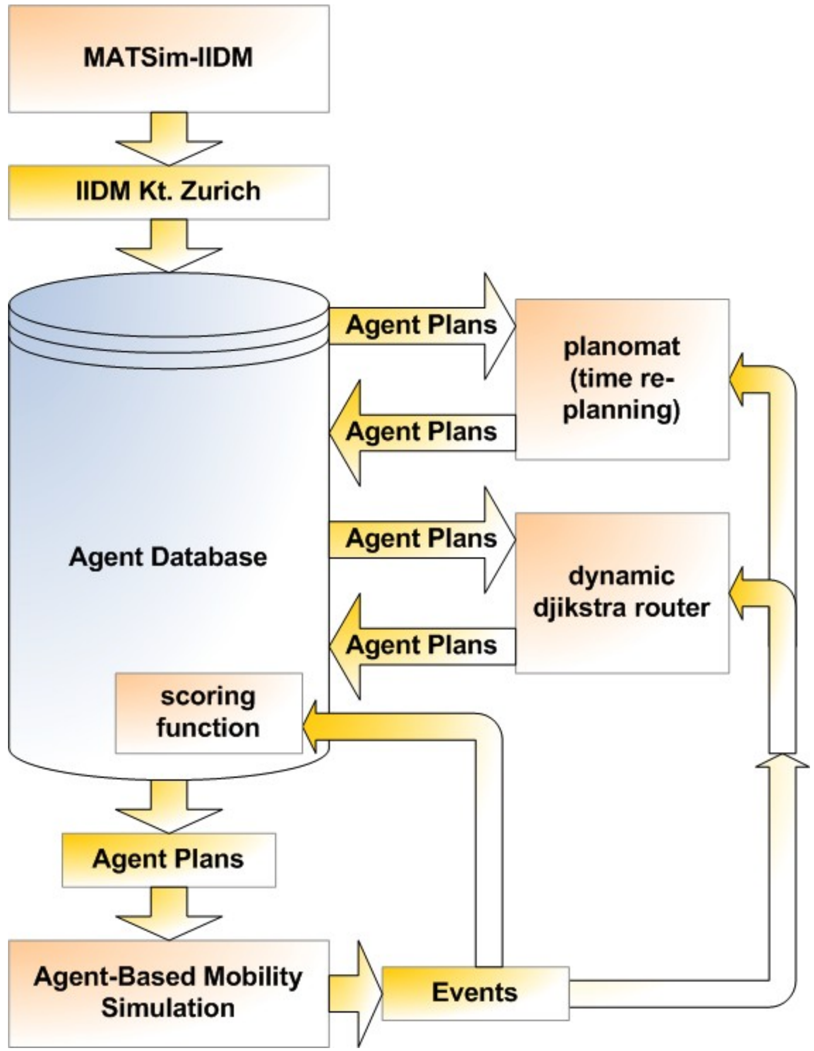



R: Hectare; M: Municipality

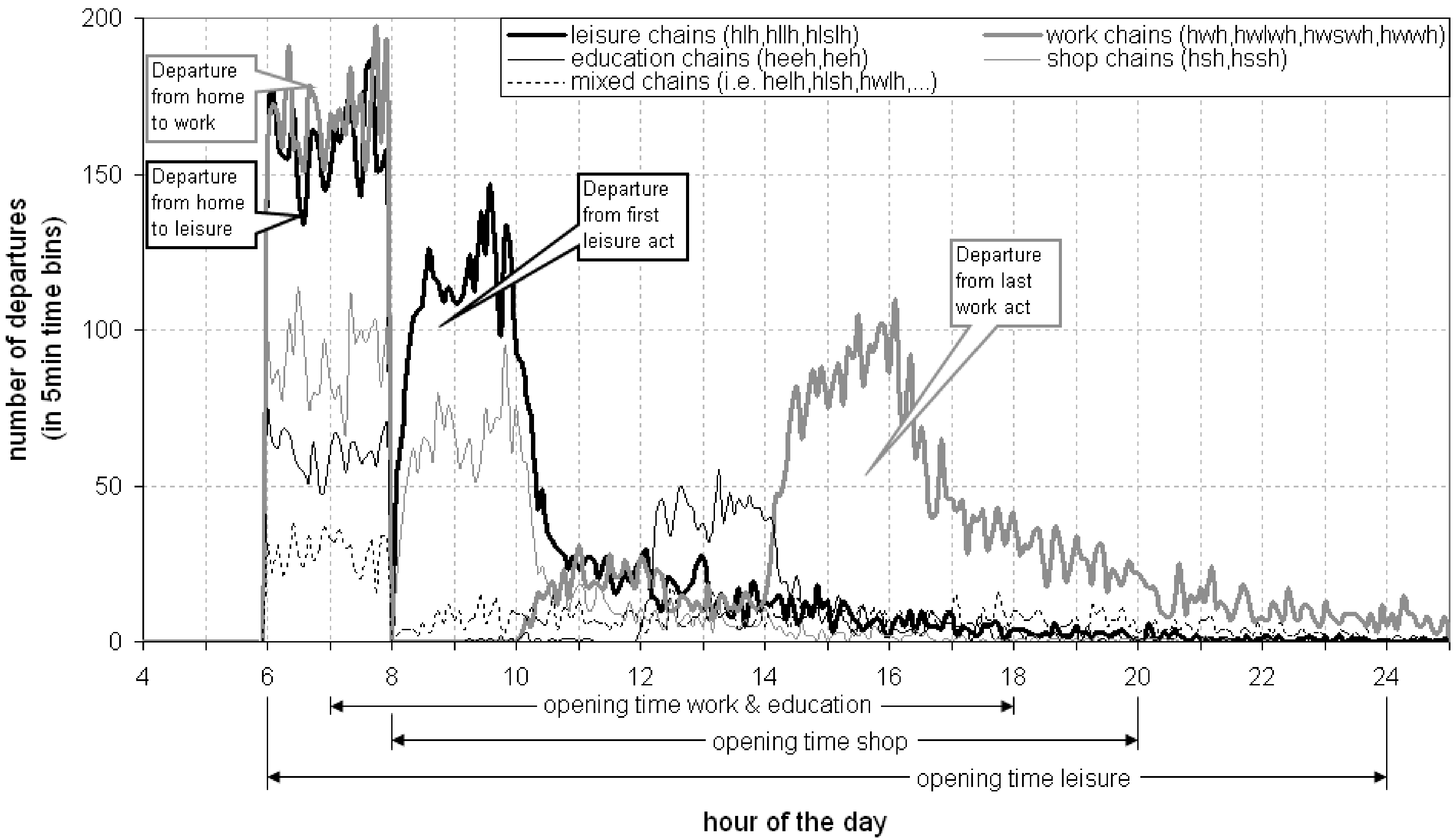
# What the tools allow: Kt. Zürich (2)



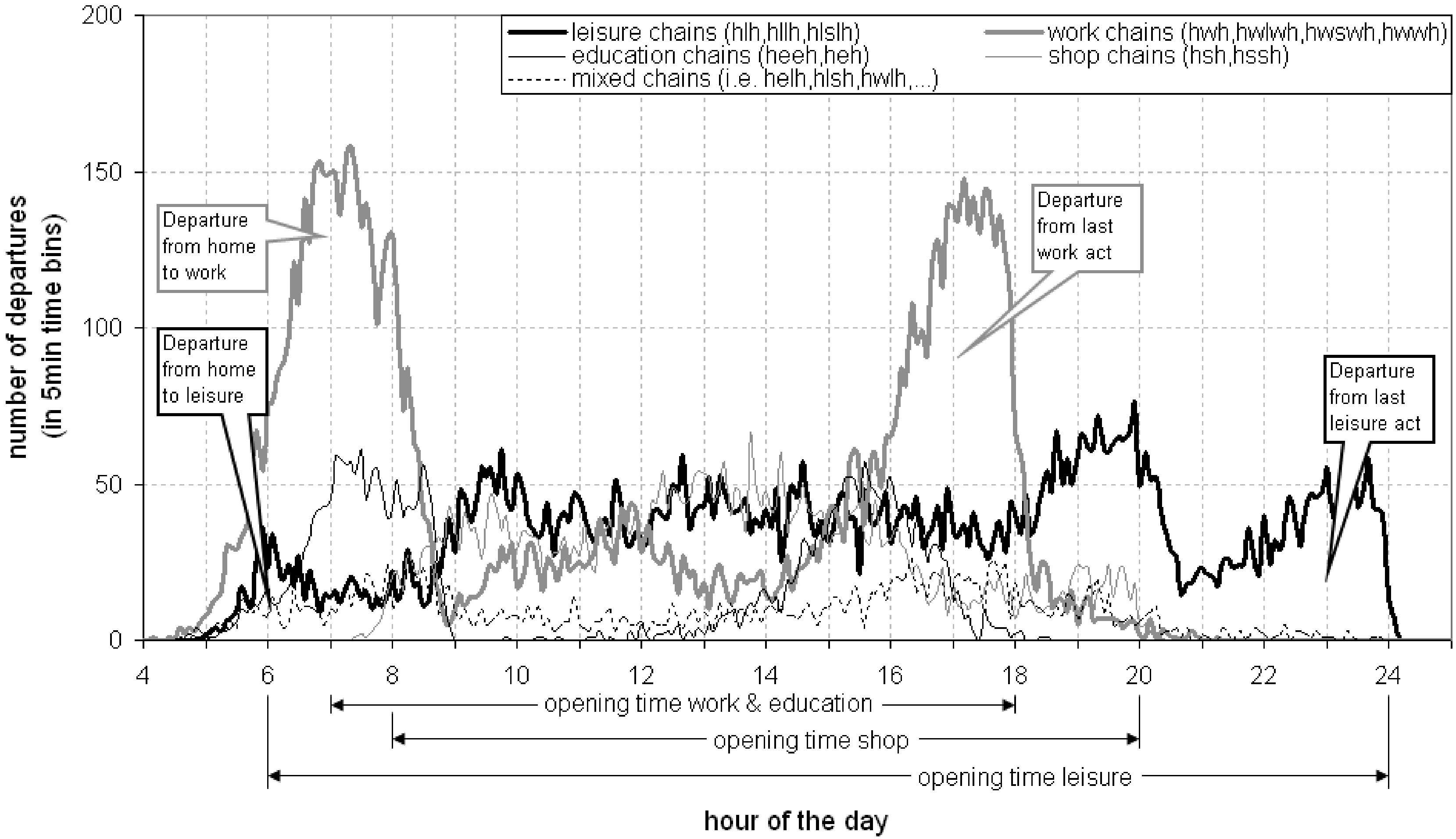
# MATSim-EA (C++) with „planomat“ (1% sample)



# MATSim-EA (C++) with „planomat": Iteration 0



# MATSim-EA (C++) with „planomat“: Iteration 400



# Current tasks: Functionality

- Improving the realism of the scenario (e.g. Opening hours, parameter distributions)
- Validation of the current Switzerland scenario
- Switzerland scenario in 12h to steady state
- Functional expansion the Planomat (Mode choice, Destination choice)
- Parameter estimation for Planomat
- Interface to UrbanSim

# Future tasks: Functionality

- Integration of social network data structures
- Explicit social network choices

- Addition of supply agents (car sharing, demand responsive transport, retail location, parking pricing, road pricing) (Traffic control)

# Current tasks: Usability (Shareability)

- Analysis tools
- Anonymous test data sets
- Improved documentation
- Web and sourceforge maintenance
- User support (Visits to ETH/Berlin)

# Acknowledgements


| | |
|---|---|
| Facilities: | Konrad Meister |
| Mobility tools: | Francesco Ciari, Sigrun Beige |
| Tour mode choice: | Francesco Ciari |
| Router: | Nicolas Lefebvre |
| Traffic flow model: | David Charypar |
| Planomat: | Konrad Meister, David Charypar |
| | |
| Supply agents: | Michael Löchl, Francesco Ciari |
| Social networks: | Jeremy Hackney |
| | |
| System architecture: | Michael Balmer |
| System integration: | Michael Balmer, Marcel Rieser |
| System management: | Michael Balmer, Marcel Rieser |

# Acknowledgements

Past contributors:

- Michael Bernard, ETH
- Sigrun Beige, ETH
- Ulrike Beuk, TU-Berlin
- Dr. Nurhan Cetin
- Phillip Fröhlich, Modus, Zürich
- Dr. Christian Gloor
- Dr. Bryan Raney, HP America
- Marc Schmitt, ETH
- Andreas Vogel, TU Berlin

# More information

www.matsim.org

www.vsp.tu-berlin.de

www.ivt.ethz.ch/vpl/publications/reports