Probabilistic Digital-Twin towards real-time human-robot collaboration

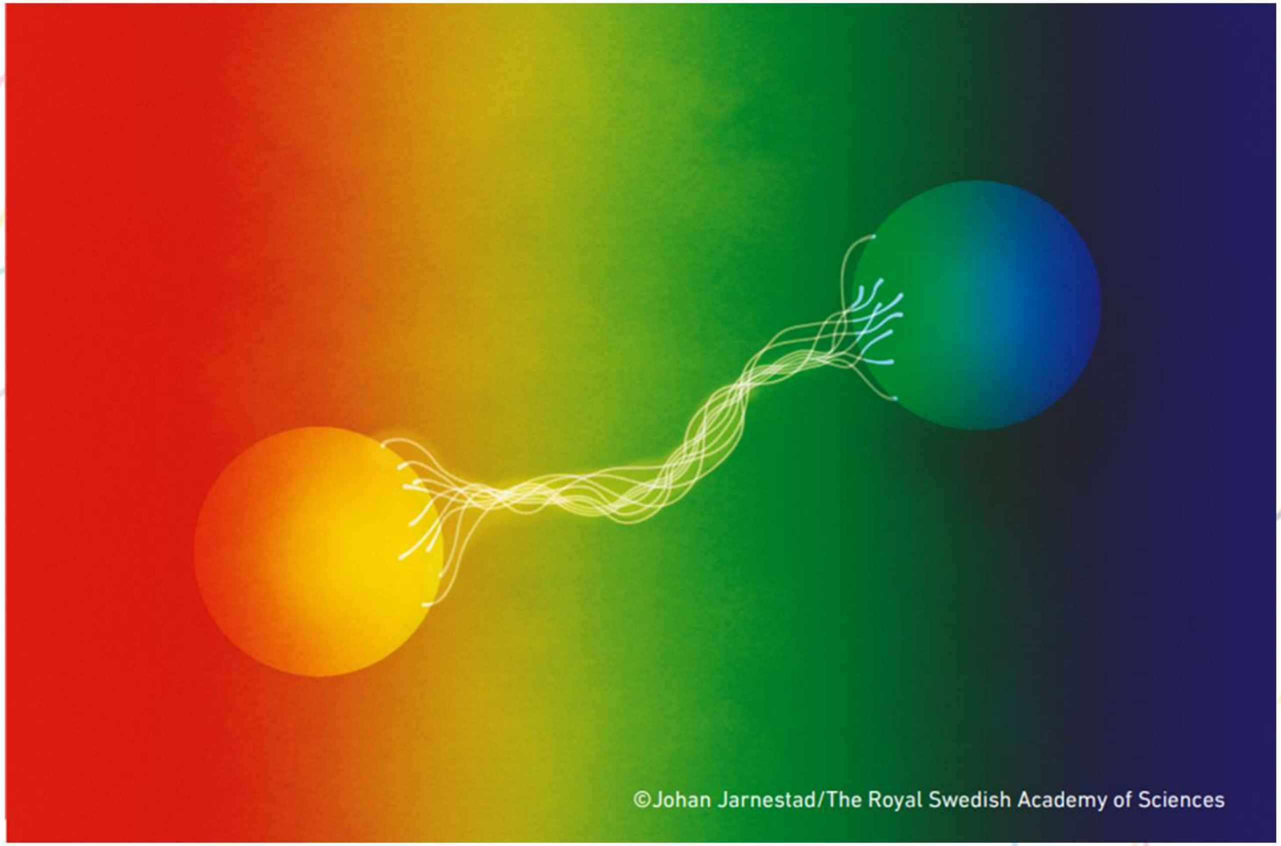
27th Oct. 2022

HIDEyuki Shimonishi,
Osaka University and NEC Corp.

大阪大学 OSAKA UNIVERSITY

OPEN 2021

大阪大学
OSAKA UNIVERSITY
OPEN 2021



https://www.nobelprize.org/uploads/2022/10/press-physics2022-figure1.pdf

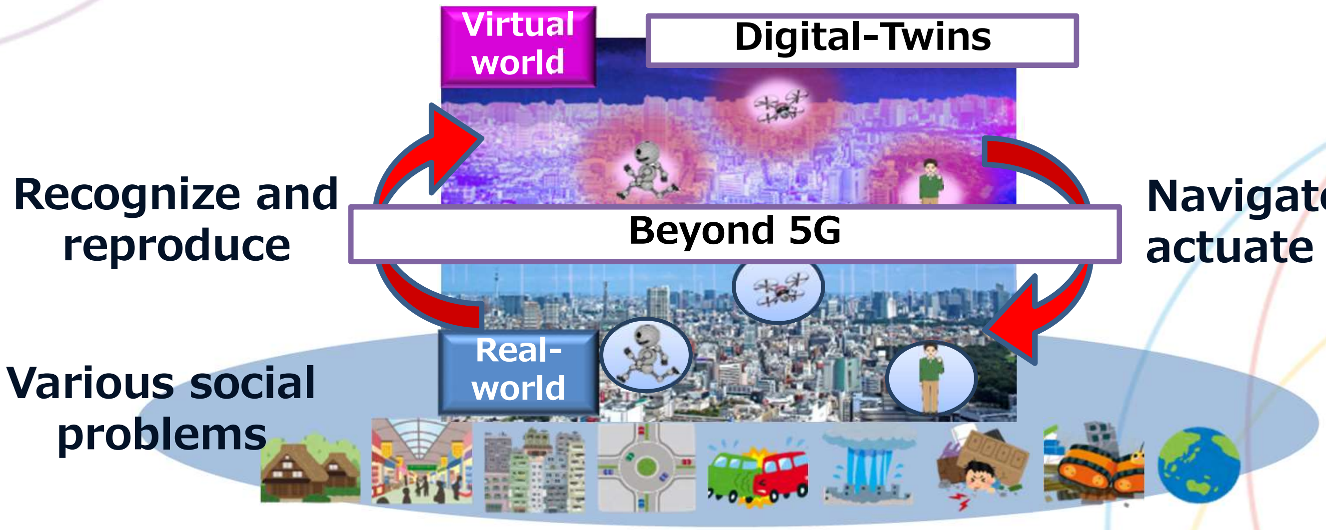
Expand IoT to Digital-Twin

- **Digital-Twin as real-time and high-precision representation of the entire space**
- **Explosive Evolution of Telecommunications with Digital-Twin**
  - Organisms acquired eyes during the Cambrian explosion
  - ICT systems acquire “eyes” towards Beyond 5G/6G era

Source: https://jpn.nec.com/nsp/5g/beyond5g/index.html

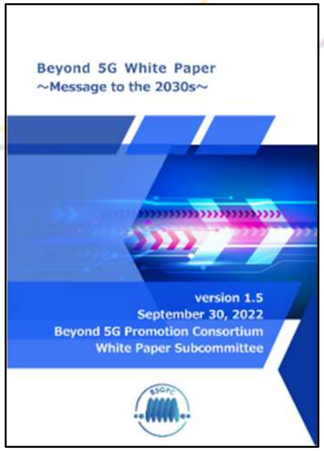
Digital-Twin (of our understanding)

- **Fusing real and virtual world to provide new value propositions to our society**
  - Digitalize the entire real-world in real-time, and reproduce them as a virtual world
  - Creating new services (such as future prediction and human-robot coexistence) by utilizing 4-dimensional (space + time) data structure in the digital-twin



Beyond 5G vision towards Digital-Twin

"Safety, security and peace of mind" would be a key driving force



Beyond 5G Promotion Consortium
Beyond 5G White Paper
https://b5g.jp/en/output.html

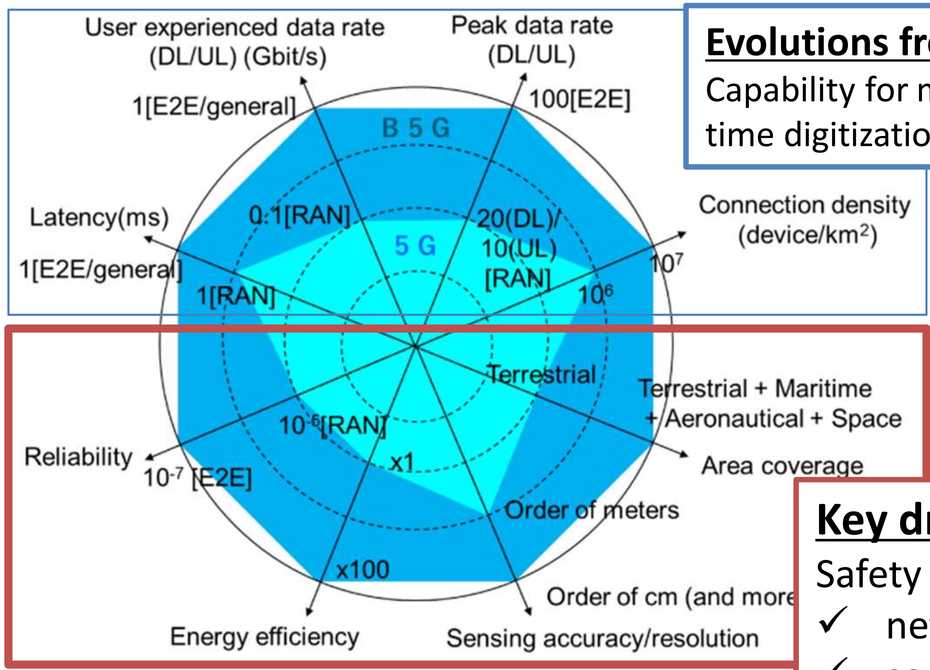


Figure 5.3-2 Target KPIs for Beyond 5G (Quantitative indicators)

Evolutions from 5G

Capability for massive and real-time digitization of the real-world

Key driving force towards B5G

Safety and security of both:

- ✓ network/ICT infrastructure, and
- ✓ social system and our daily life

“Deterministic” Digital-Twin

Constructing very precise copy of real-world to be safe and secure, with Beyond 5G and advance AI

- Robust ?
  - Uncertainty in network reliability and service quality
  - Uncertainty in recognition and control of real-world

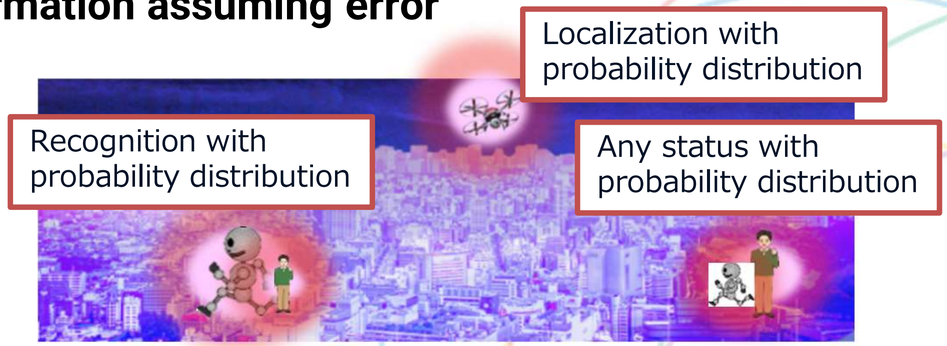
- Eco-friendly?
  - Huge traffic amount: very high-definition sensing data
  - Huge computation: very accurate recognition

“Probabilistic” Digital-Twin

- **A digital twin that:**
  - **probabilistically infers real-world from uncertain observations,**
  - **non-deterministically predicts the future,**
  - **and navigate human, actuate robots, control things, flexibly**

Real-world will tolerate sudden events and physical uncertainties by decision making based on ”**probabilistic information assuming error**”

Deterministic representation



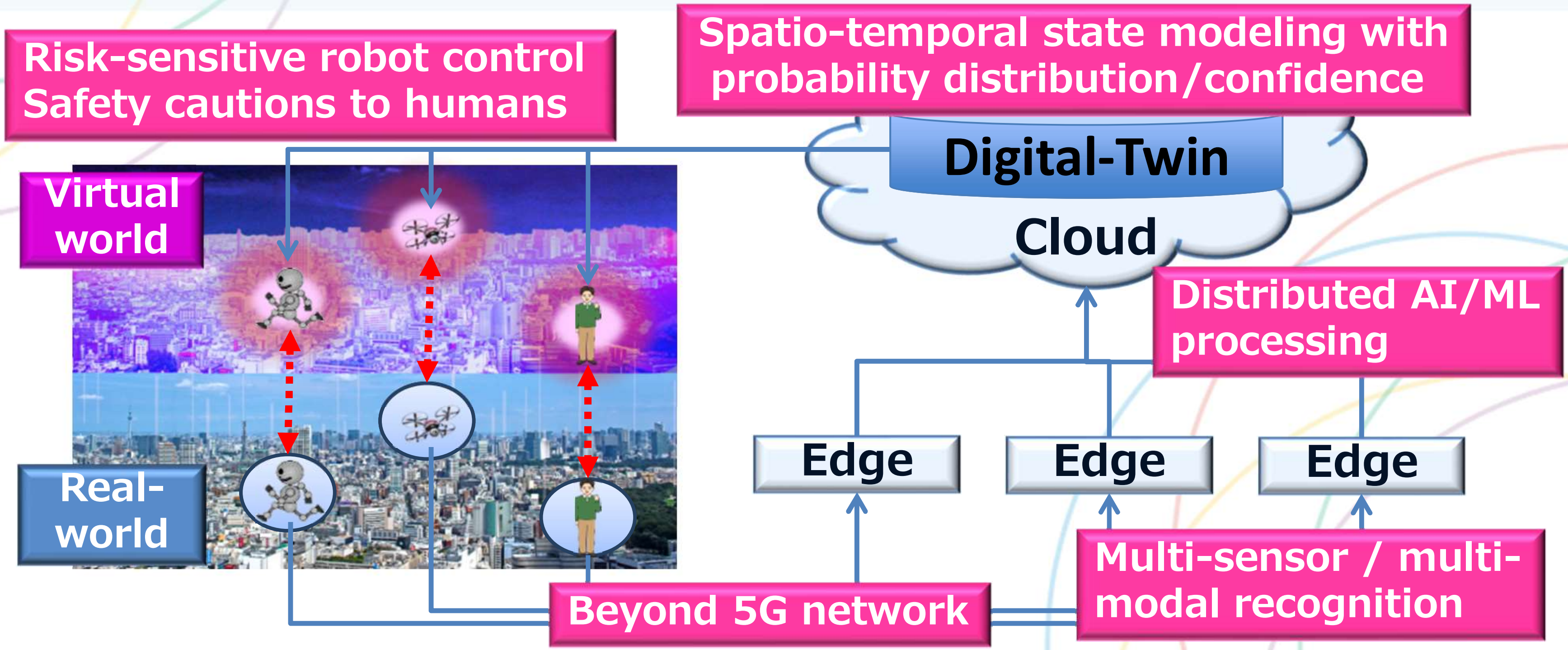
Probabilistic representation

Technologies for Digital-Twin

大阪大学 OSAKA UNIVERSITY

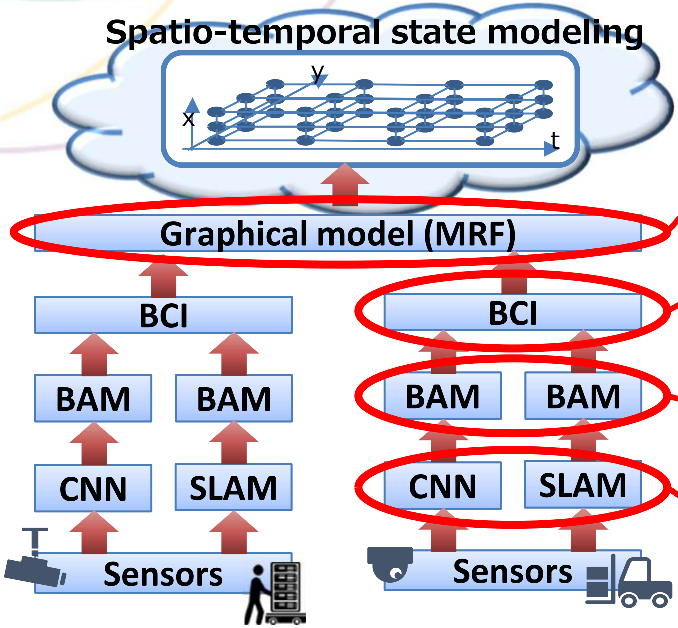
OPEN 2021

Digital-Twin of human, robots, cars, cities, and things in physical space used for robots/humans



(1) Multimodal recognition (1/2)

- Probabilistic recognition of various objects in physical space from noisy and unstable monitoring
- Mathematical model of the brain's stochastic perceptual function



Cognitive decision making using graphical models
(most plausible interpretation of the entire real world)

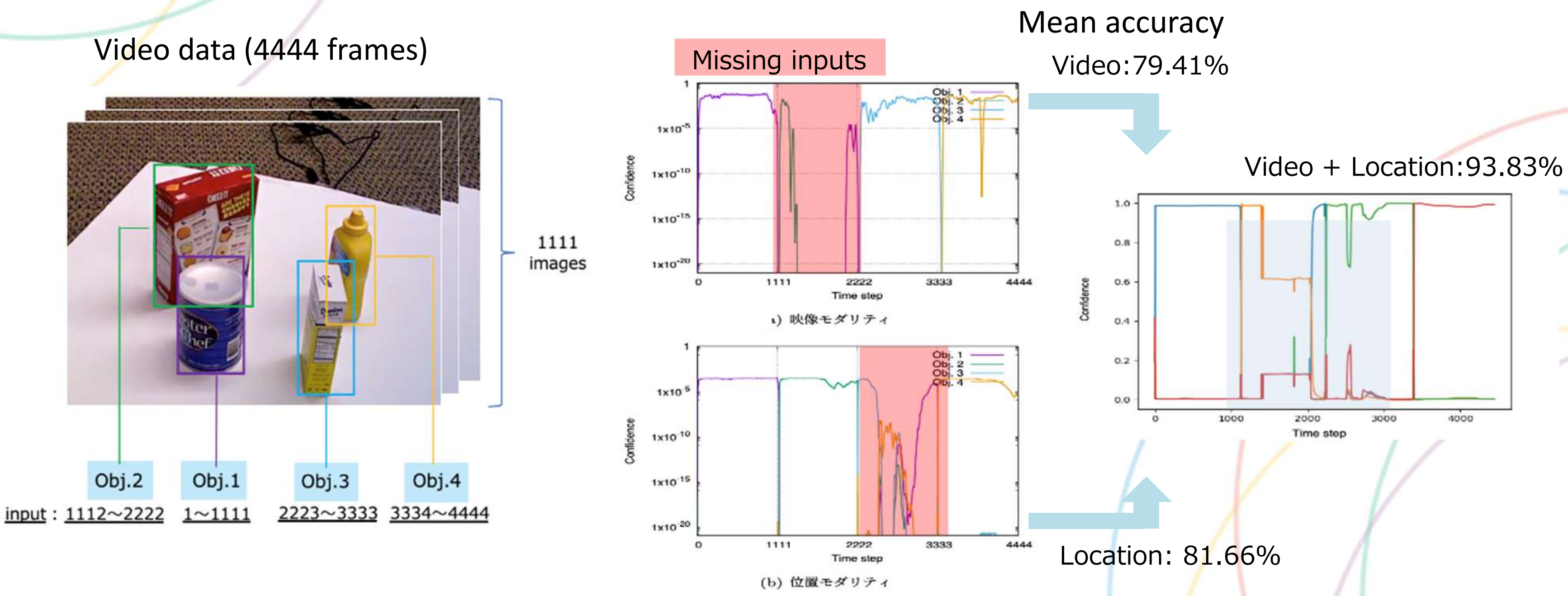
Multimodal integration using Bayesian Causal Inference
(BCI = Causal inference by perceptual integration processing in the brain)

Unimodal object recognition using Bayesian Attractor Model
(BAM = cognitive models of the human brain to identify system states)

Feature extraction using CNN and other ML models

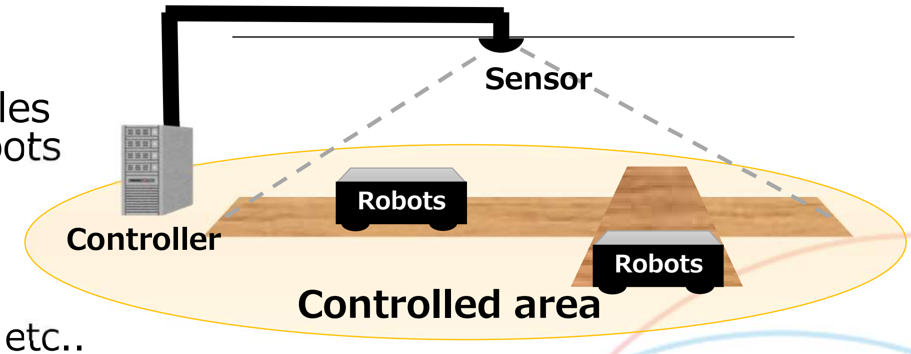
(1) Multimodal recognition (2/2)

- **Multimodal complementation combines missing inputs to generate plausible cognitive decision making**



(2) Spatio-temporal state modeling (1/3)

Application to the prediction of obstacles

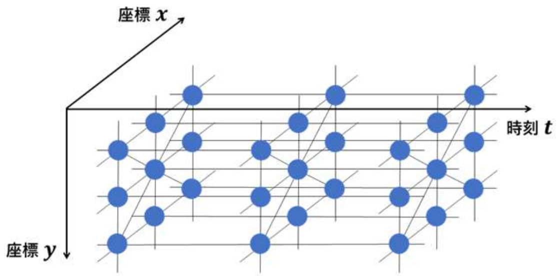


Objective:
- To predict the existence of moving obstacles in each area for the control of moving robots

Assumption:
- Observation
  - Multiple sensors such as camera, LiDAR etc..
- Obstacles
  - Obstacles may move (e.g., human, other robots etc..)

Approach:
- Construct spatio-temporal model of the area based of CRF
- Update the model based on the observations.

(3) Spatio-temporal state modeling (2/3)



- Application to the prediction of obstacles

$$P(S|D) = \frac{1}{Z(D)} \exp\left(-E(S;D)\right)$$

$$E(S;D) = \sum_{(t,x,y)\in N} f(s_{t,x,y}; d_{t,x,y}) + \sum_{((t_1,x_1,y_1),(t_2,x_2,y_2))\in E} f_{p((t_1,x_1,y_1),(t_2,x_2,y_2))}(s_{t_1,x_1,y_1} s_{t_2,x_2,y_2}; d_{t_1,x_1,y_1}, d_{t_2,x_2,y_2})$$

- $s_{t,x,y}$: State of the area x, y at the time slot t
  - No obstacles or the trajectory ID of the obstacle in the area x, y
- $d_{t,x,y}$: Observations based on depth camera of the area x, y at the time slot t
- $f(s_{t,x,y}; d_{t,x,y})$, $f_{p((t_1,x_1,y_1),(t_1,x_1,y_1))}\left(s_{t_1,x_1,y_1} s_{t_2,x_2,y_2}; d_{t_1,x_1,y_1} d_{t_2,x_2,y_2}\right)$: Defined based on the behavior of the past obstacles.

(2) Spatio-temporal state modeling (3/3)

Example of the results of the prediction

The area with obstacles can be predicted
based on the spatio-temporal CRF.

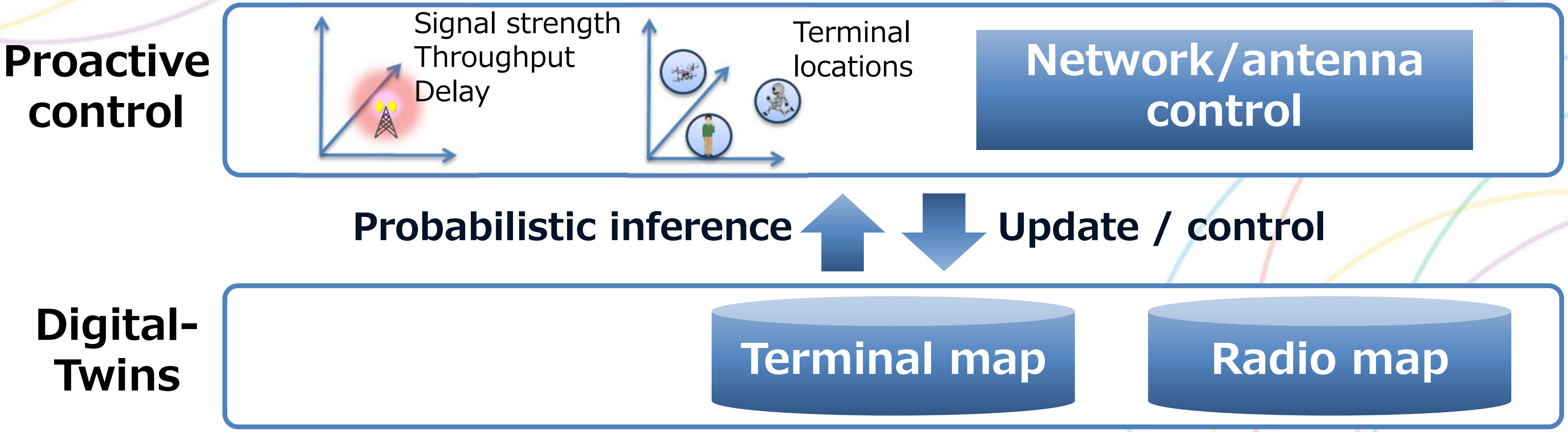
Actual Obstacles

Prediction
(3 seconds before the target)

Prediction
(5 seconds before the target)

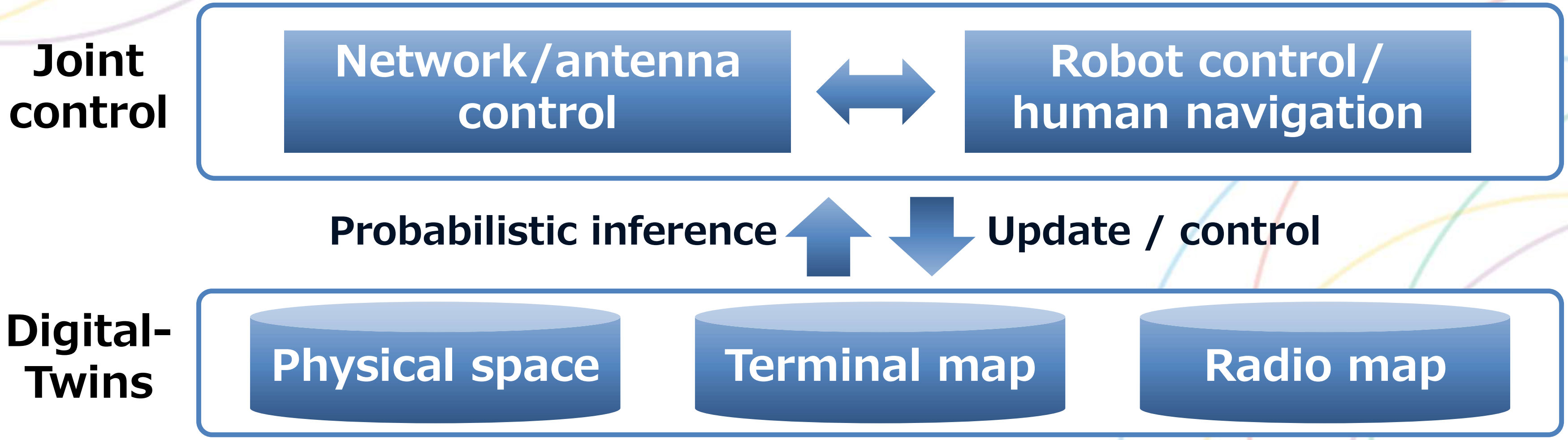
(3) Radio communication map (1/2)

- Practical use of higher frequency radio in mobile environment
- Controlling system and antennas using “Radio communication map”
  - Map mesh > Wavelength; deterministic representation would be hard
  - Probabilistic representation of signal strength/throughput/delay map



(3) Radio communication map (2/2)

- Coordination of Digital-Twins of radio map and physical space
  ➔Maps them into common spatial axes
- Coordinated control of radio and robot
  - Robots take a path of good radio, and move cautiously otherwise
  - Network controls antenna beam targeting moving robots



(4) Distributed AI processing (1/4)

- Reduce power consumption of the entire system is required for digital twin construction
- Optimization of distributed processing is necessary to perform video analysis with the required recognition accuracy and delay time.

Video analysis at cloud

- Video data is collected from cameras to the cloud
- Consuming large amounts of network bandwidth

Video analysis at edge

- Reduces network bandwidth through local processing
- Limited computing resources

(4) Distributed AI processing (2/4)

- Minimize power consumption: sum of terminal/edge/cloud/wireless NW/backbone NW power consumption

$$P = f_D^{comp}(\alpha_D)P_D W_D + DP_W(1 - W_D) + f_E^{comp}(\alpha_E)P_E W_E + DP_B(1 - W_D - W_E) + f_C^{comp}(\alpha_C)P_C(1 - W_D - W_E)$$

- E2E latency constraint: sum of analysis time at terminal/edge/cloud and transfer time at wireless NW/backbone NW

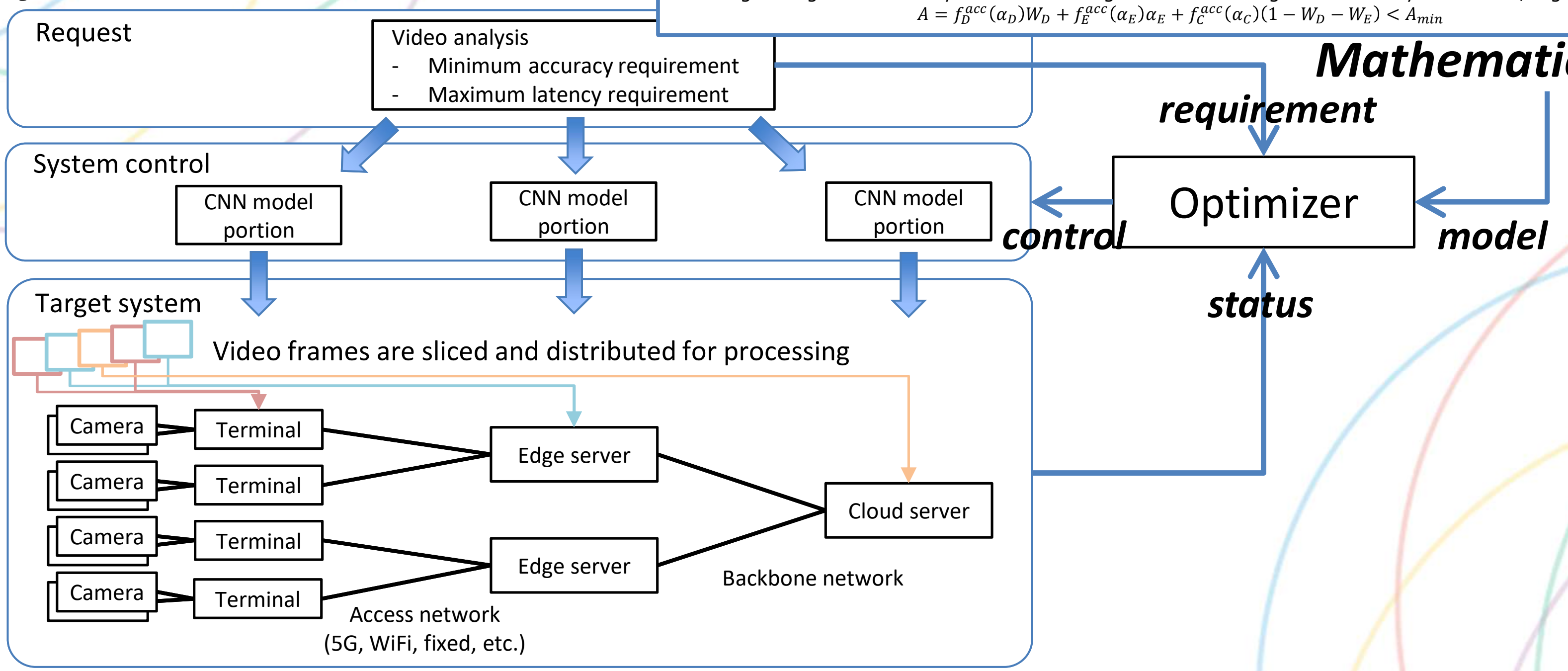
$$L = \frac{f_D^{comp}(\alpha_D)}{C_D}W_D + \frac{D}{B_W}(1 - W_D) + \frac{f_E^{comp}(\alpha_E)}{C_E}W_E + \frac{D}{B_B}(1 - W_D - W_E) + \frac{f_C^{comp}(\alpha_C)}{C_C}(1 - W_D - W_E) < L_{max}$$

- Image recognition accuracy constraints: weighted sum of recognition accuracy at terminal/edge/cloud

$$A = f_D^{acc}(\alpha_D)W_D + f_E^{acc}(\alpha_E)\alpha_E + f_C^{acc}(\alpha_C)(1 - W_D - W_E) < A_{min}$$

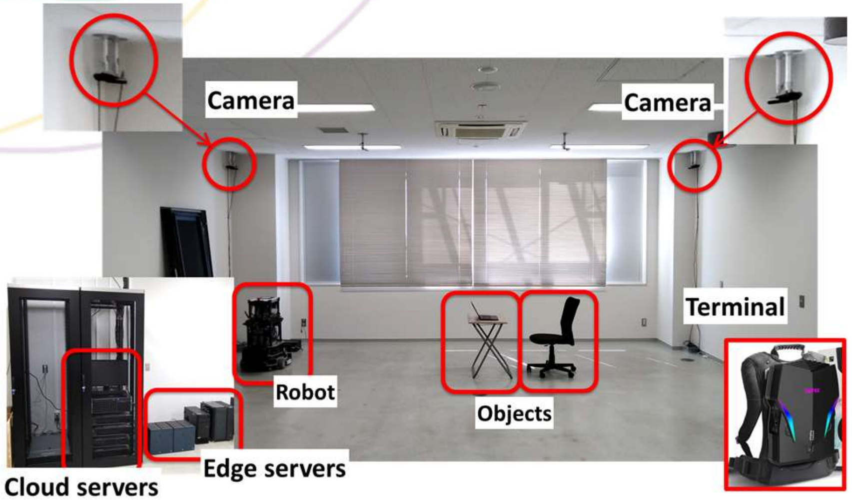
System model

Mathematical model



(4) Distributed AI processing (3/4)

- Experiment to get machine modeling for YoLo-v3
- Estimate GPU load, processing delay, and power consumption based on machine spec, CNN model, and video frame rate.



Machine spec.

	Terminal	Edge/cloud server	
CPU	Core i7-8700T	Core i9-10940X	Xeon GLOD 6226R x2
CPU TDP	35W	165W	205W x3
GPU	Nvidia GeForce GTX1070	Nvidia RTX A5000	Nvidia Tesla T4 x2
GPU FP32	6.463Tflops	27.77Tflops	8.141Tflops x2
GPU TDP	150W	230W	70W x2

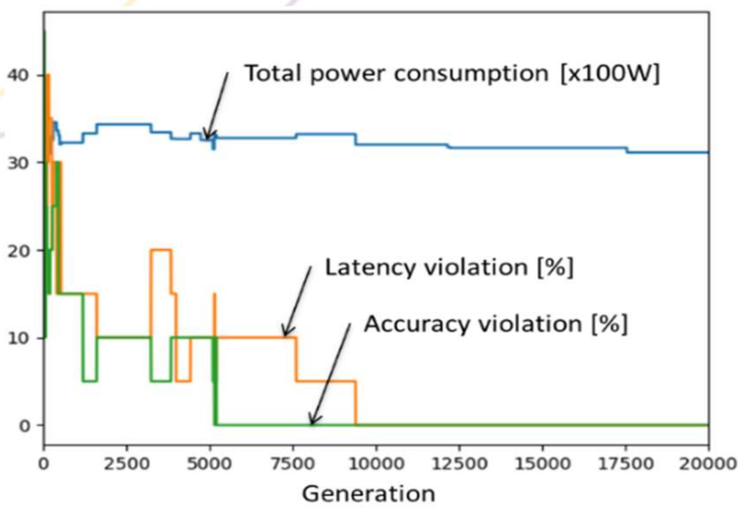
CNN models

Model	mAP	Floating operations
Yolov3-tiny	33.1%	5.6B / frame
Yolov3	55.3%	65.9B / frame
Yolov3-spp	60.6%	141.5B / frame

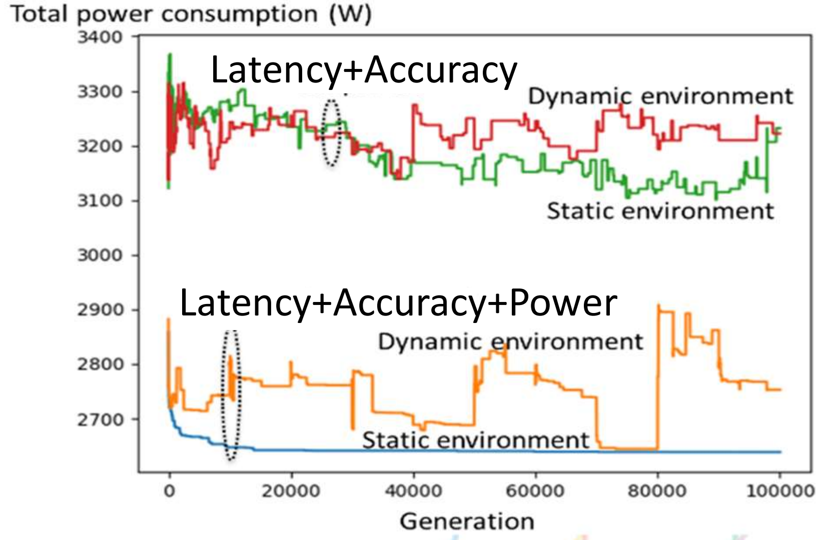
(4) Distributed AI processing (4/4)

- System optimization with Genetic Algorithm solver

Constraints violation

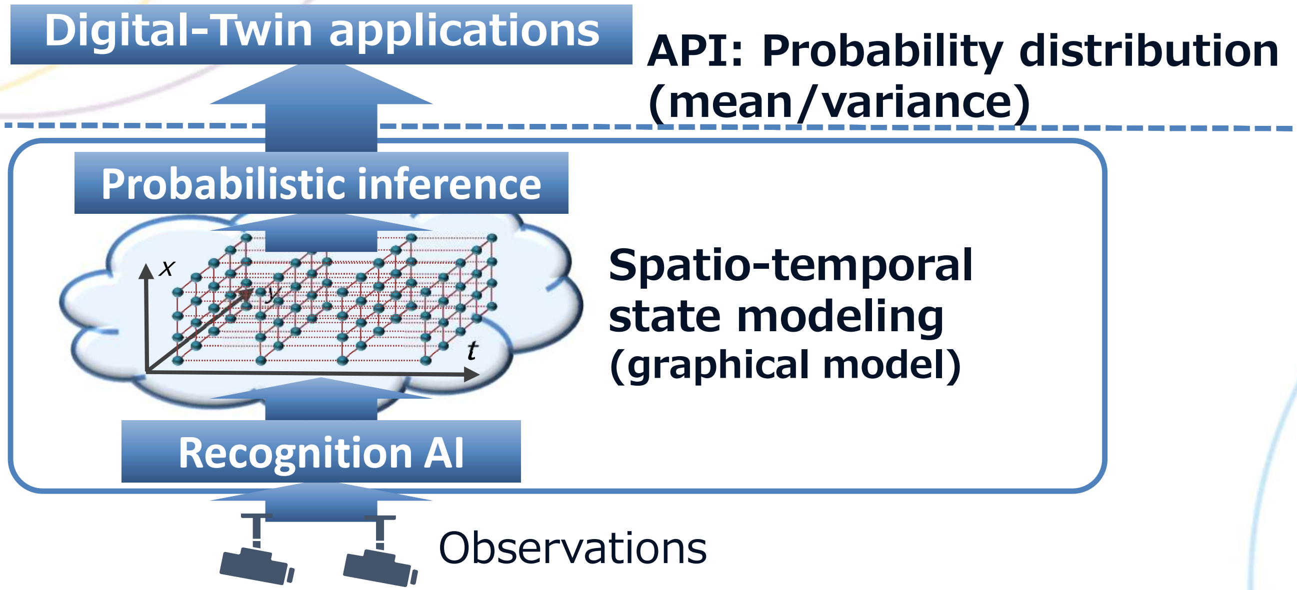


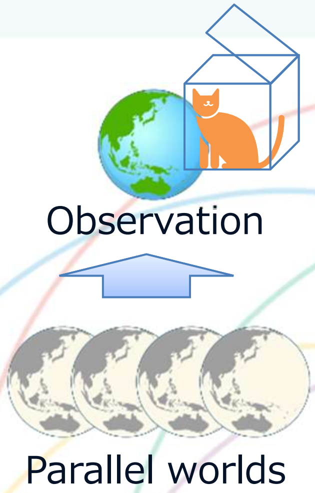
Power saving



Possible discussions:
Probabilistic data structure and API

- Probabilistic internal data structure that allow for **<u>any parallel understanding of real world</u>**
- APIs that provide probabilistic information of **<u>maximum likelihood understanding of the moment</u>**





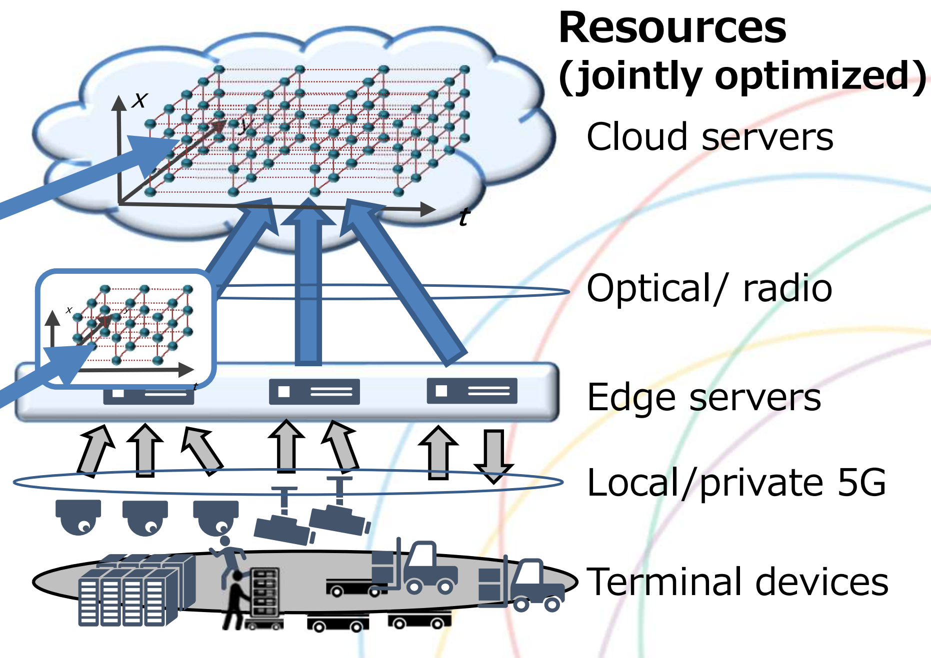
Possible discussions:
Digital-twin computing platform

- Hierarchical edge-cloud data structure meeting space and time resolution requirements
- Integrated optimization of network and computing resources

<u>Let us build campus-wide testbed as easy as possible.</u>

Digital-twin (cloud)
Global and long-term data

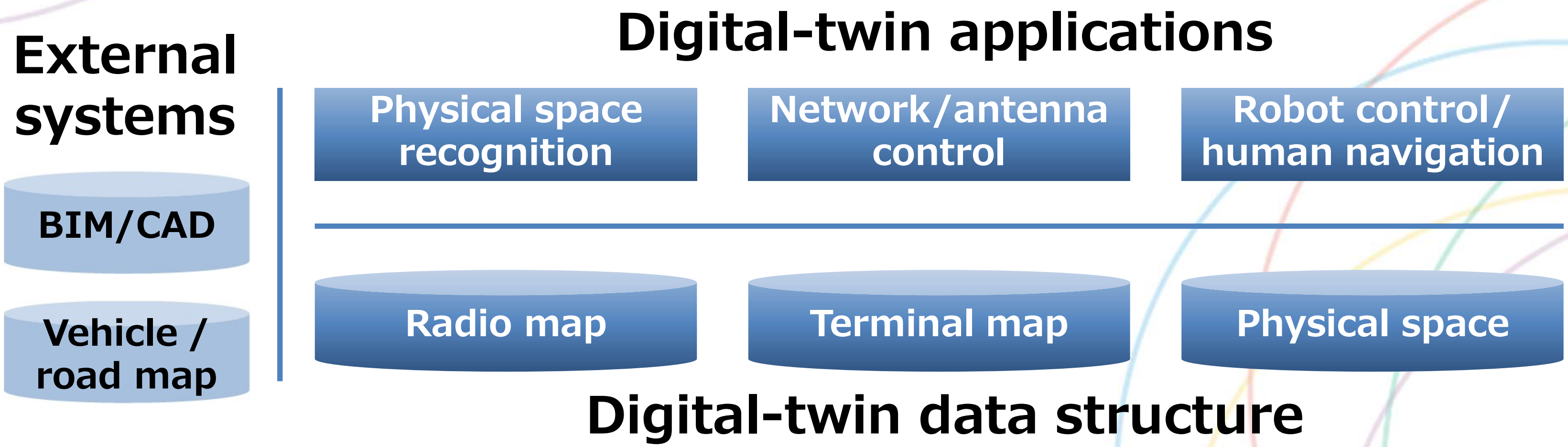
Digital-twin (local)
Local data / real time control



Possible discussions:
Digital-twin computing framework

- Real time data exchange for Cyber Physical Systems (10-200msec)
- Flexible data exchange with internal/external systems

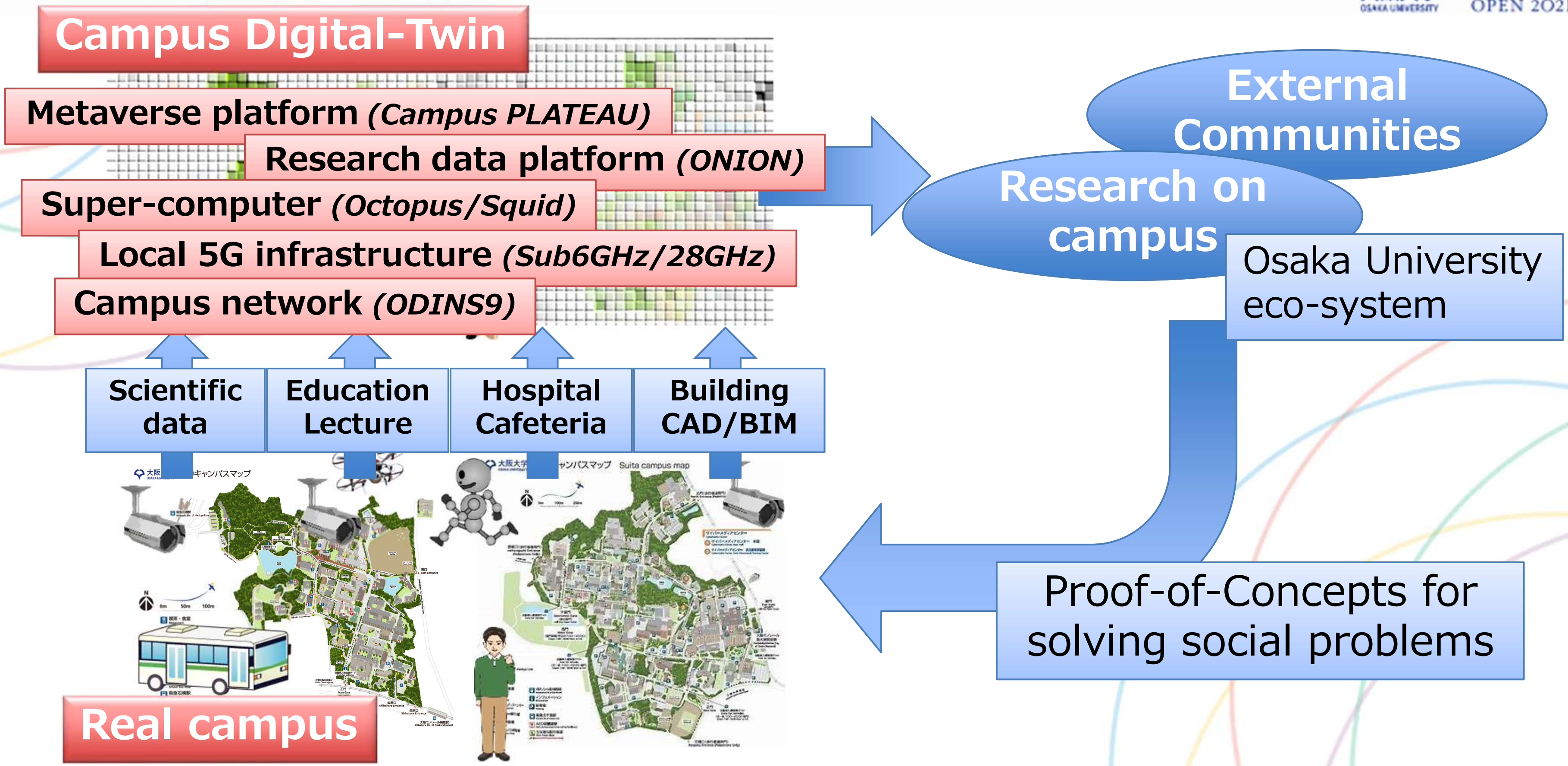
Let our students write those experimental/PoC applications as easy as possible.



Campus Living-Lab vision

大阪大学 OSAKA UNIVERSITY

OPEN 2021



Summary

- Let us digitize probabilistically
  - Physical space; robots and humans
  - Network and radio, and others

- (1) Multimodal recognition
- (2) Spatio-temporal state modeling
- (3) Radio communication map
- (4) Distributed AI processing