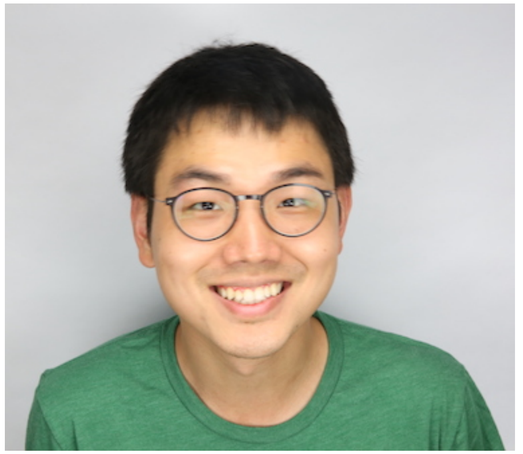
Sensing and Privacy

- The Yin-Yang of Ubiquitous Computing

UC San Diego
HALICIOĞLU DATA SCIENCE INSTITUTE

Haojian Jin
haojian@ucsd.edu

Bio



Haojian Jin (http://haojianj.in/)

Asst. Prof @ UCSD-HDSI

HCI, Privacy, Mobile Computing

Ph.D. from CMU Human-Computer Interaction Institute

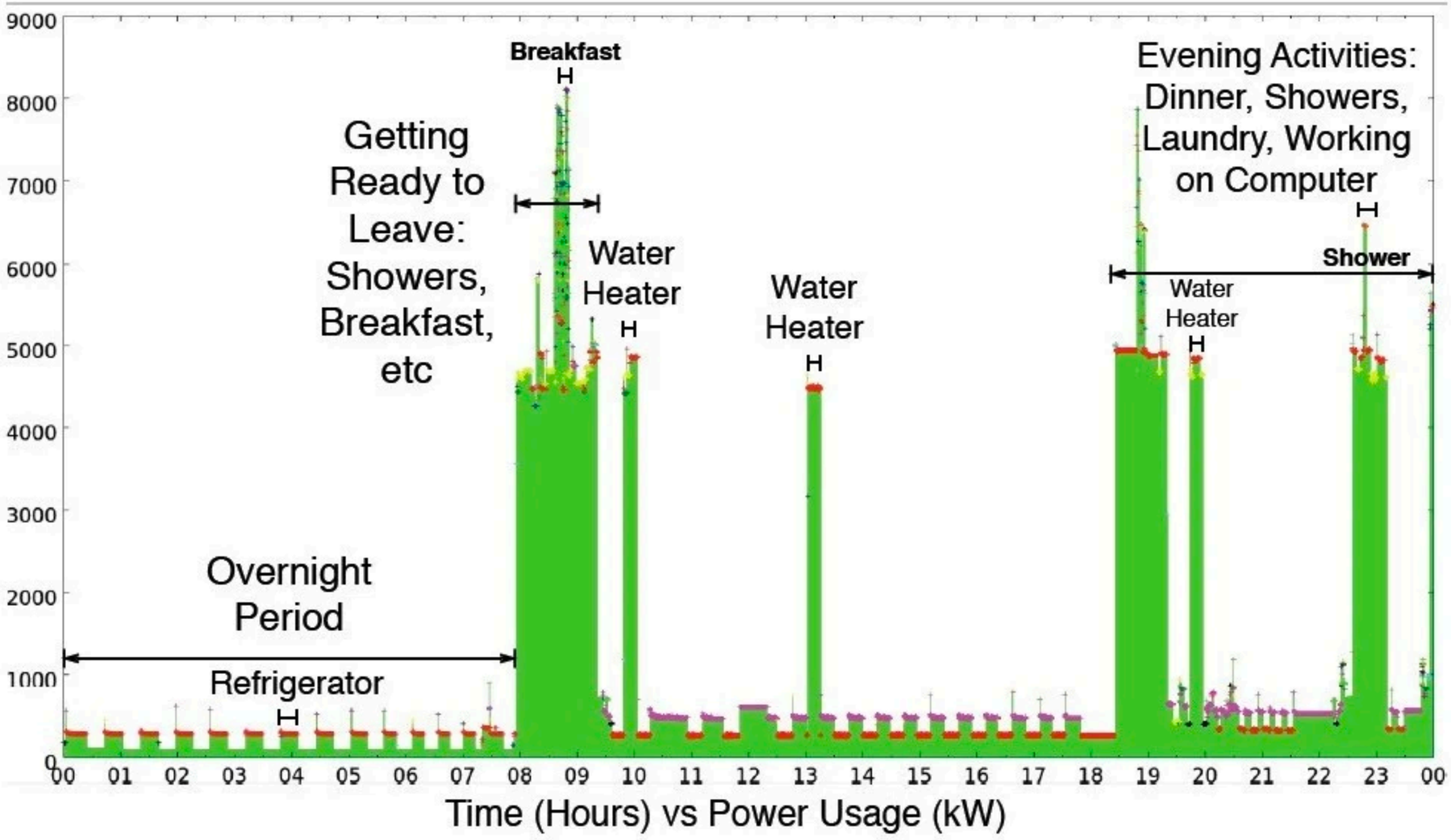
Before Ph.D.: worked at Yahoo Research, ran a startup

Looking for students and collaborators

What would our world be in 2035?

- Cheaper sensors
- More sensor deployments
- Better data mining algorithms
- Better and cheaper network/ storage/computing
- ……

Smart devices will know everything about us!



Private Memoirs of a Smart Meter, BuildSys 2010

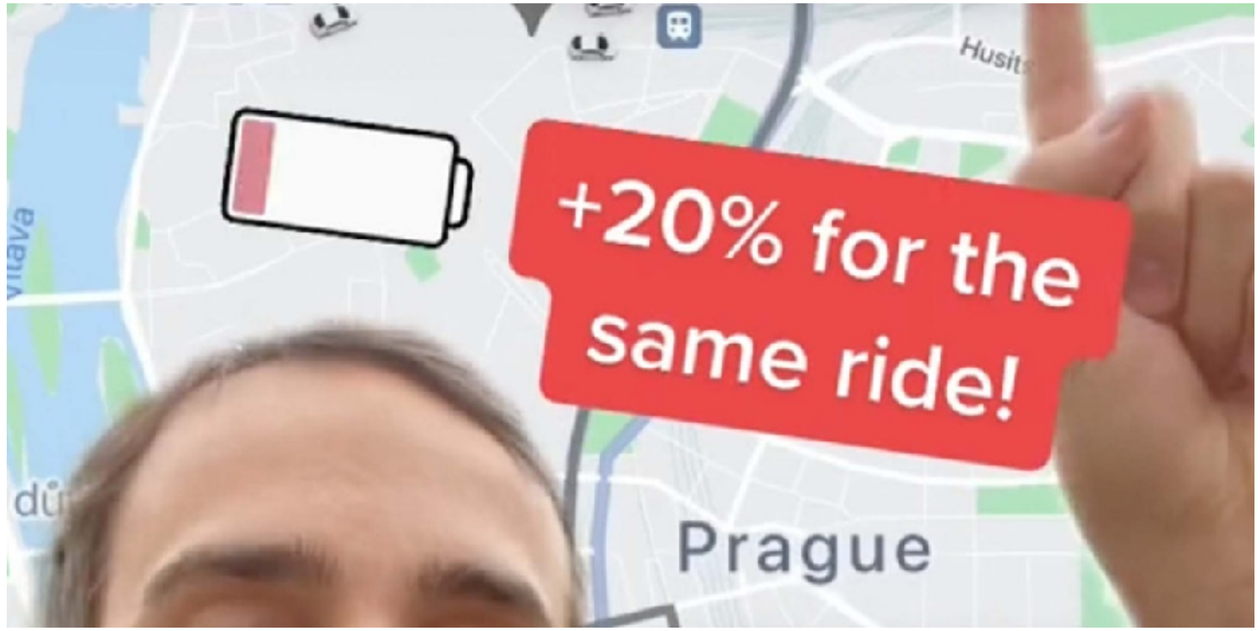
How can we protect users' privacy in a smart world?

Kiip:

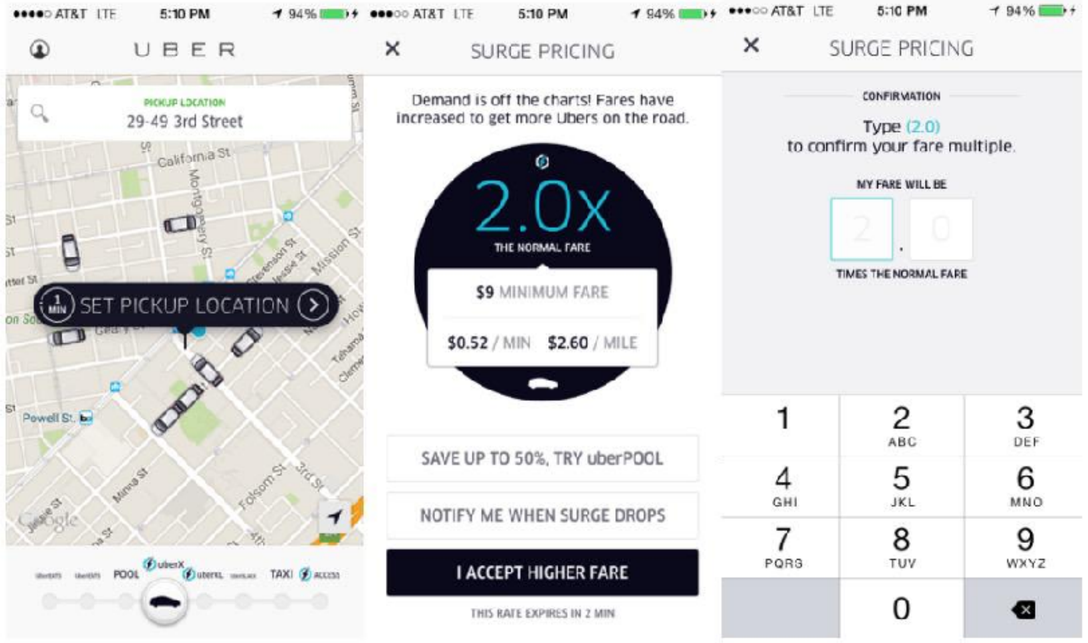
Real-life rewards advertising

How can we protect users' privacy in a smart world?

Low battery



High surge price?



Jin et al., Lean Privacy Review, 2021

A ubiquitously connected world that ~~advertisers want~~

people want to live in

self-contradicting?

improves peoples' live

respect users' privacy

Data Smith Lab

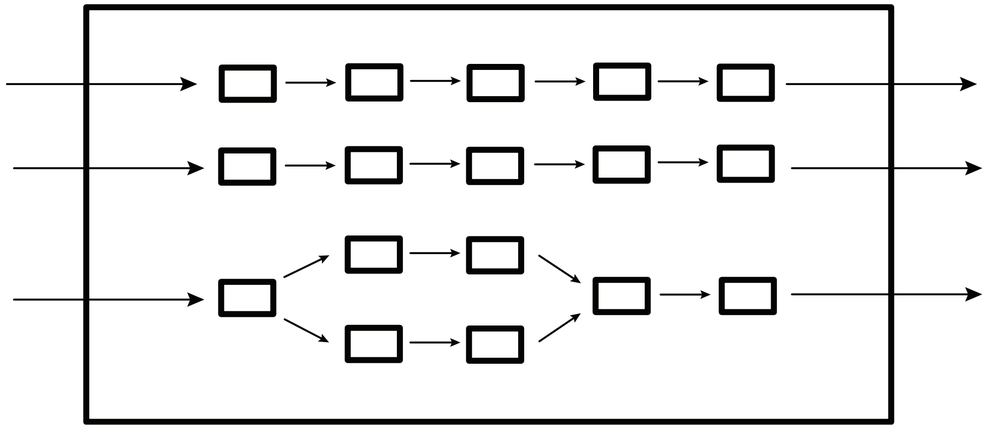
Safe, fair, cheap, accessible data economy

Casey Neistat's Guide to Filmmaking, https://www.youtube.com/watch?v=nLSUrTxquyE

This talk.

Software Defined Cooking

[MobiCom' 19, Featured in Communications of the ACM]



Modular Privacy Flows

[Ph.D. dissertation]

Mind Of A Chef. S1:E3 Memory Lane

Two important ingredients of a graduate student

Free pizza

Microwave oven

Today's Microwave: a blunt heating device

uneven & unpredictable heating

What is software-defined cooking?

Cooking is the **application of heat** to ingredients to transform them via chemical and physical reactions

Jeff Potter. Cooking for Geeks: Real Science, Great Hacks, and Good Food.

Cooking is the application of heat to ingredients to transform them via chemical and physical reactions

programmable heating

heat the food in a software-defined thermal trajectory (recipe).

Jeff Potter. Cooking for Geeks: Real Science, Great Hacks, and Good Food.

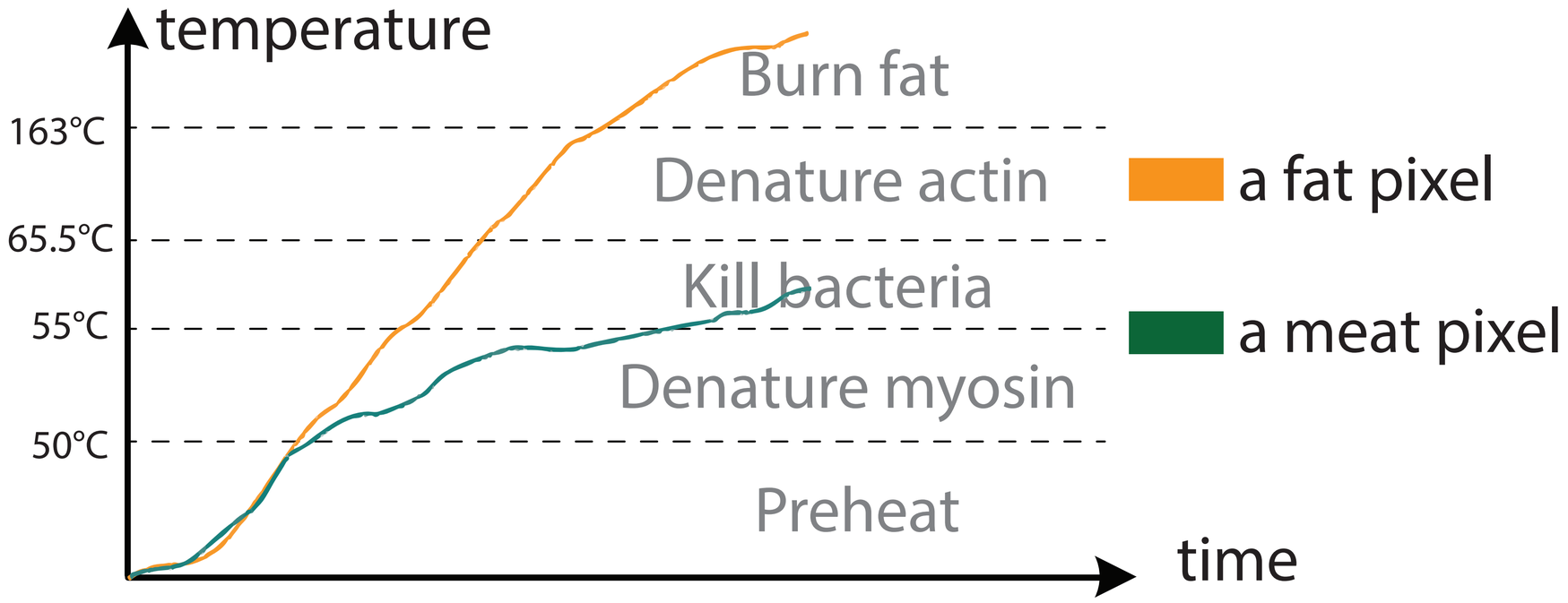
Perfectly-cooked bacon

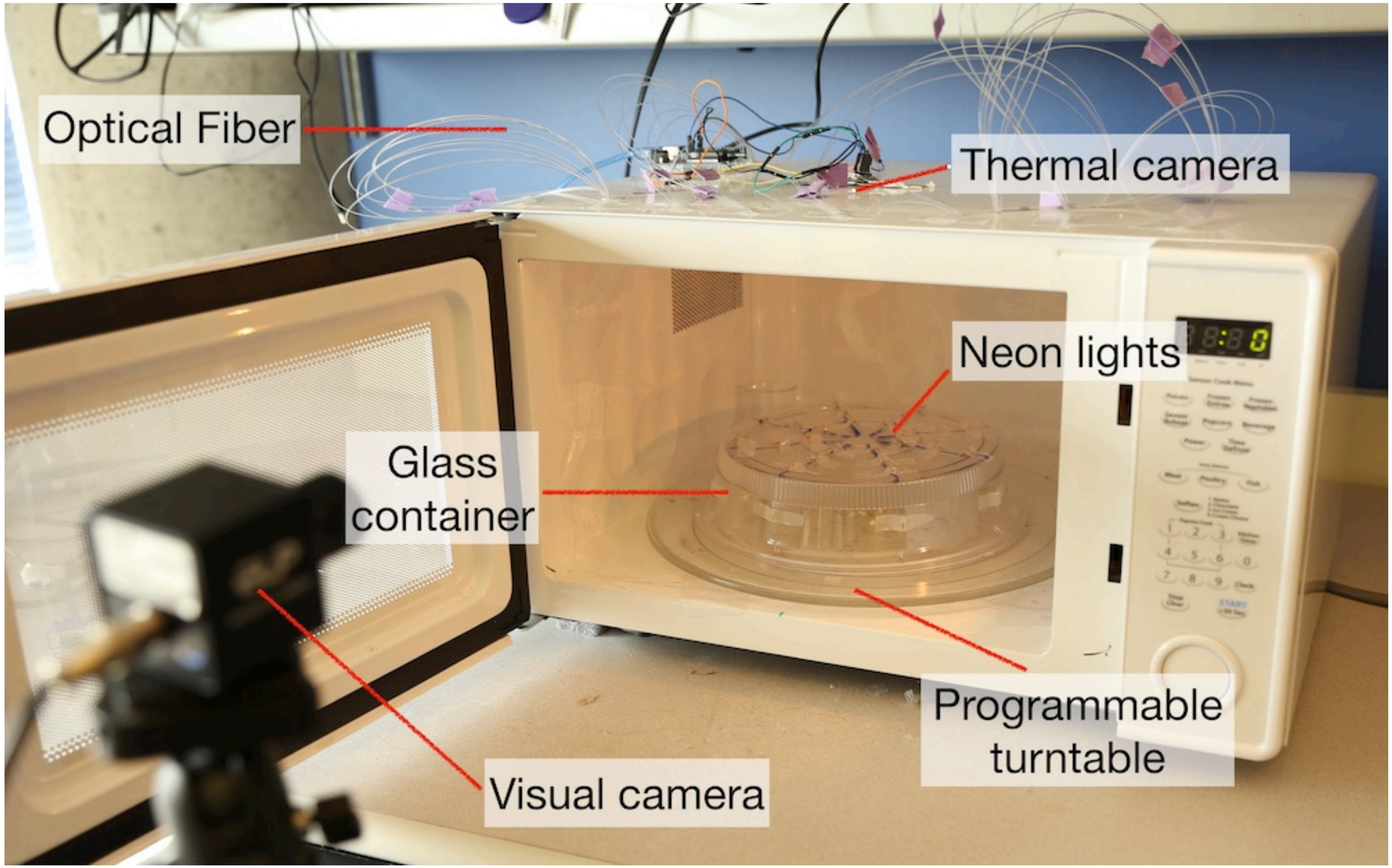
overcooking fat,

without burning the meat.

https://www.huffpost.com/entry/bacon-mistakes-how-to-cook_n_3111706

SDC recipe is a progression of desired temperature vs. time per-pixel of food.





SDC (software-defined cooking): a novel low-cost **closed-loop** system that can **sense** and **control** heating at a **fine-grained** resolution.

Spoiler alert

No Turntable

Default Turntable

SDC Uniform Heating

SDC Arbitrary Heating

Spoiler alert

No Turntable

Default Turntable

SDC Uniform Heating

SDC Arbitrary Heating

Spoiler alert

No Turntable

Default Turntable

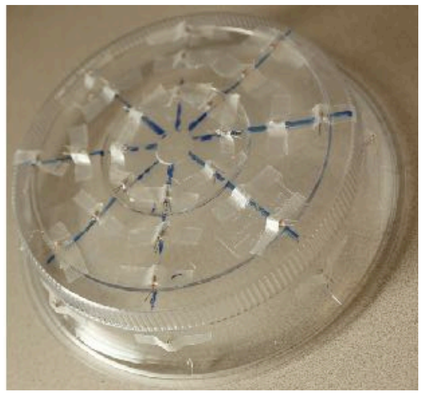
SDC Uniform Heating

SDC Arbitrary Heating

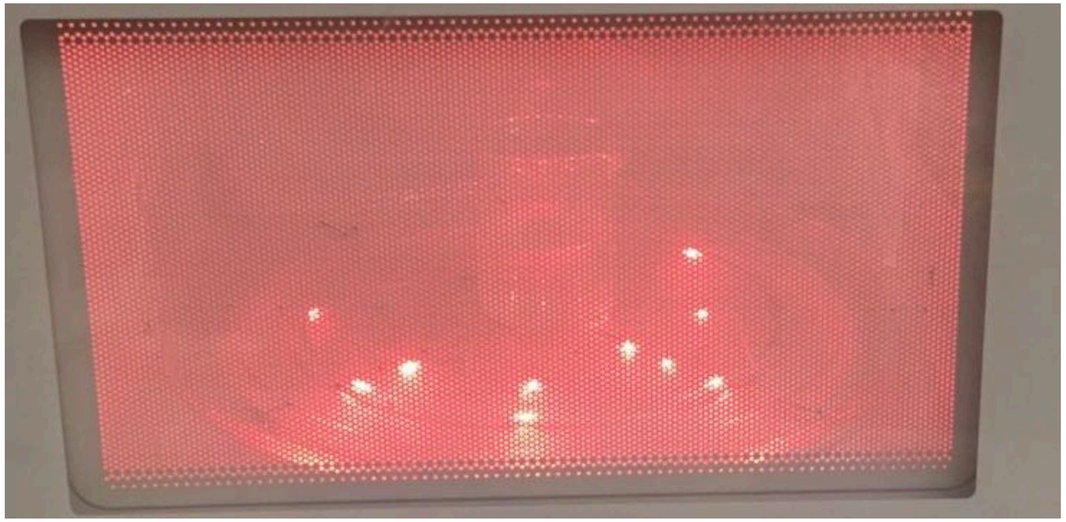
Today's Microwave: a blunt heating device

Sensing	Don't know how much heat each food pixel has absorbed.
Actuation	Have no way to actuate heating on a specific food pixel.

A closed-loop system to heat smartly



Sensing



Actuation

Heat Sensing

Microwave is dangerous

sharp-edged metals (e.g., forks, most sensors, motors)

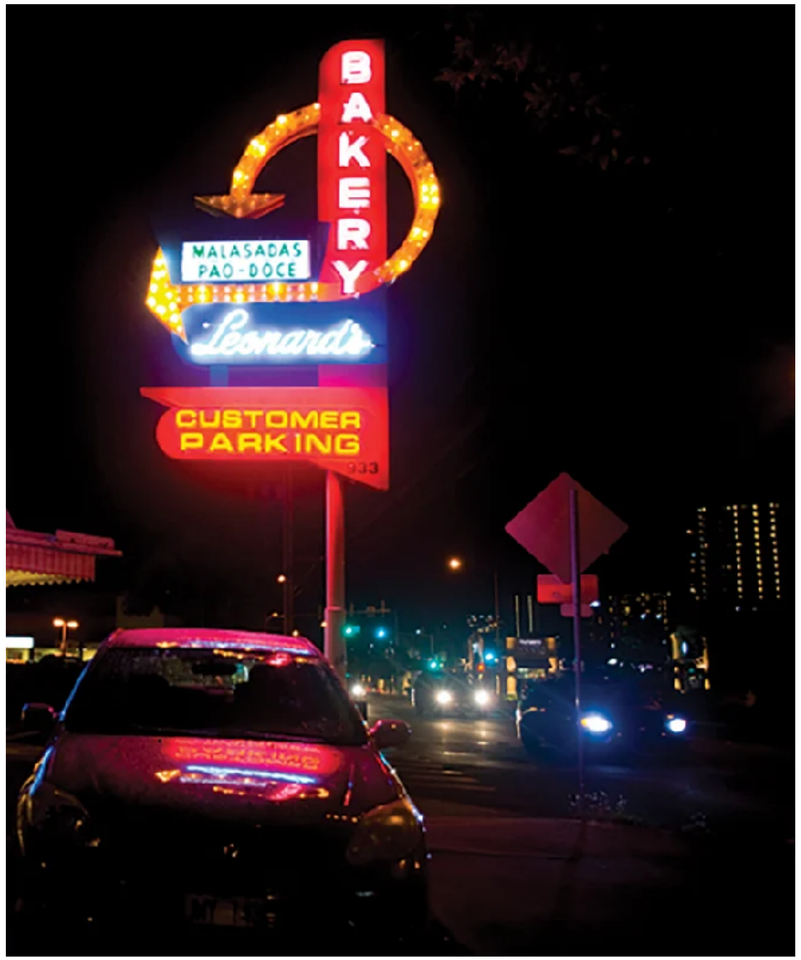
Microwave-safe plastic

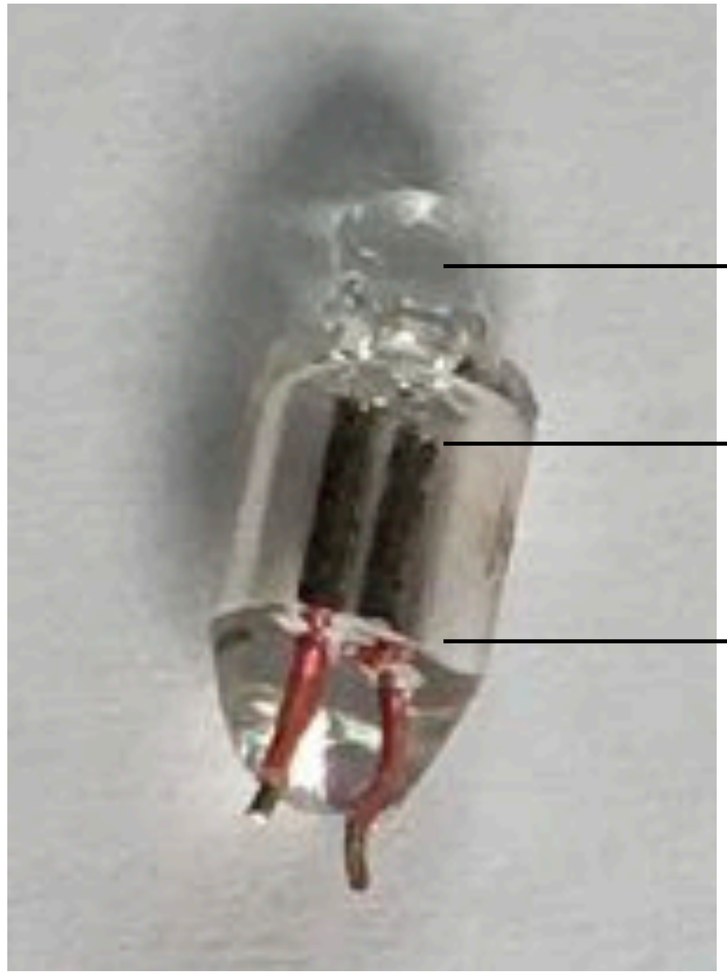
Eggs

battery

….

Neon lights

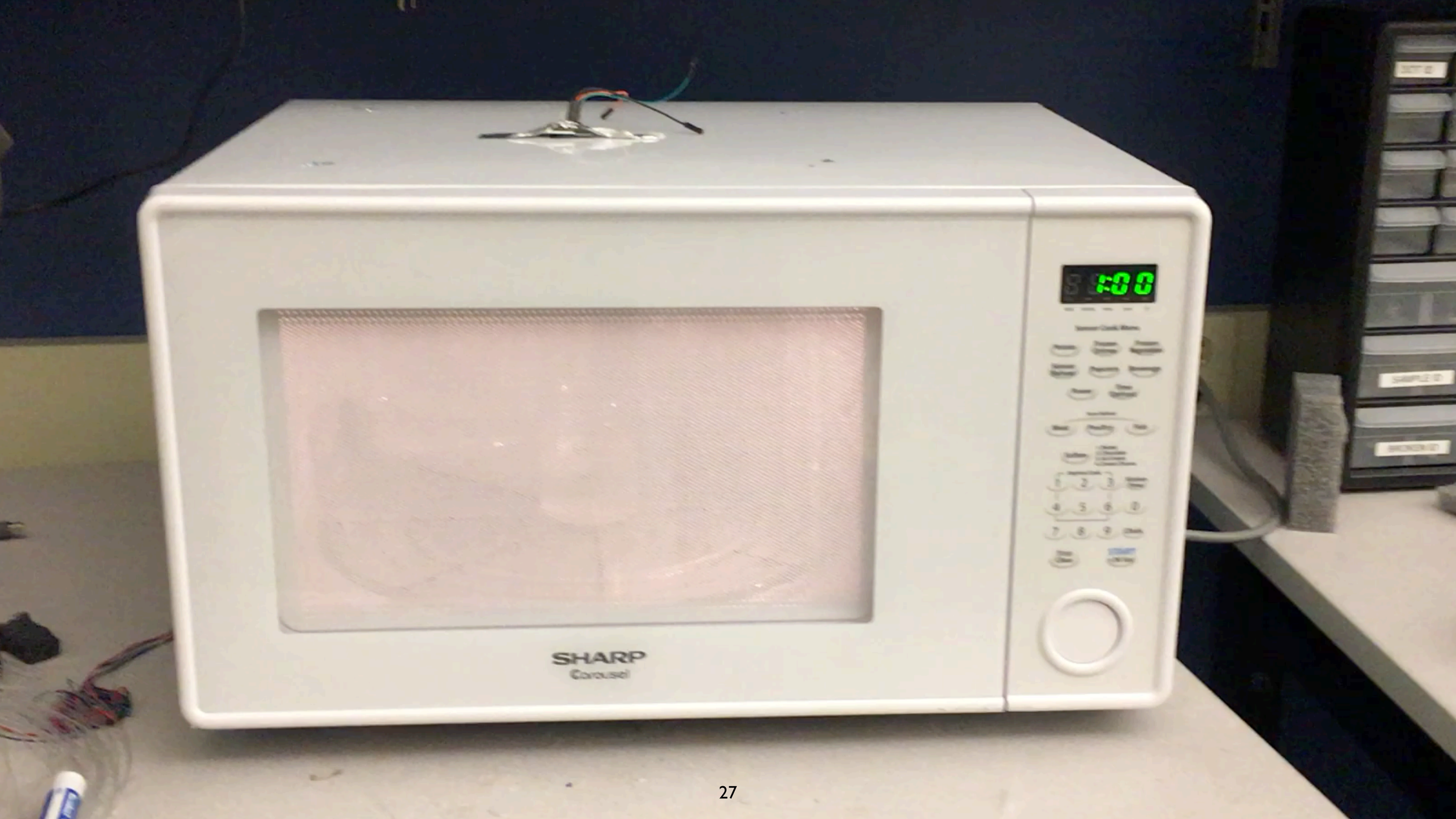




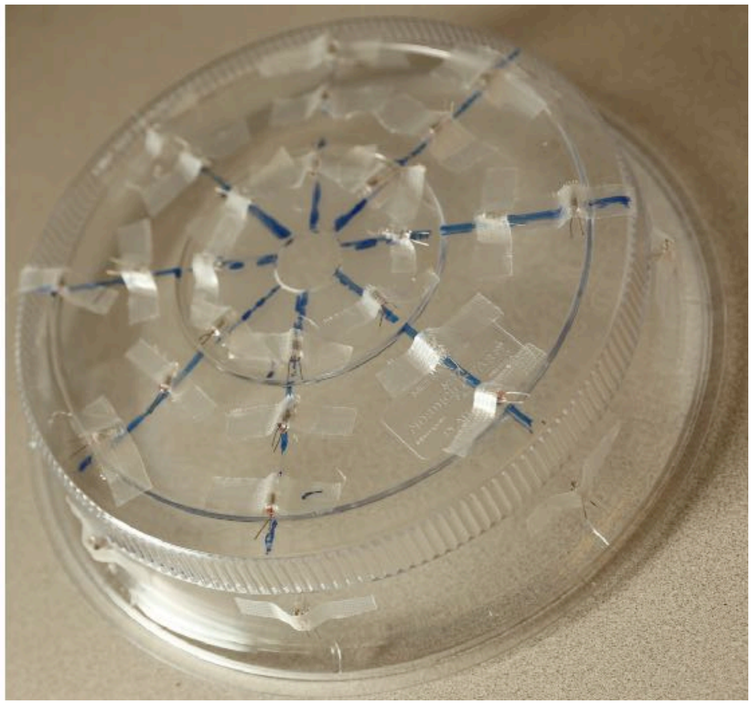
Glass

Electrodes

Low-pressure
Neon gas mixture


SHARP
Carousel

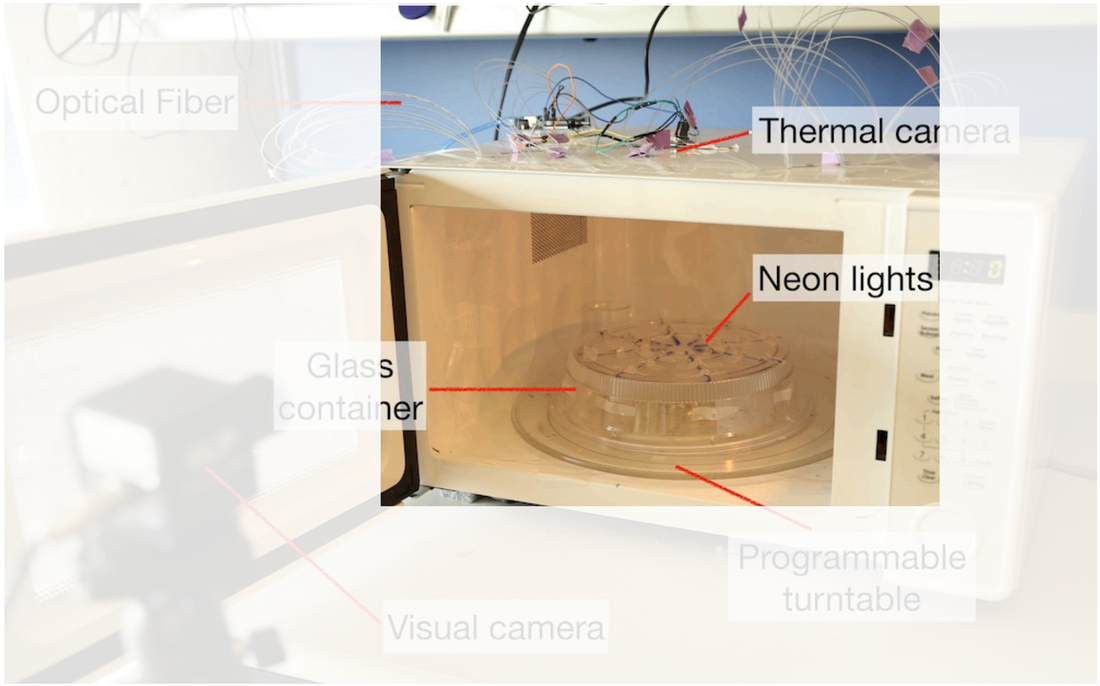
Placement of Neon Lights

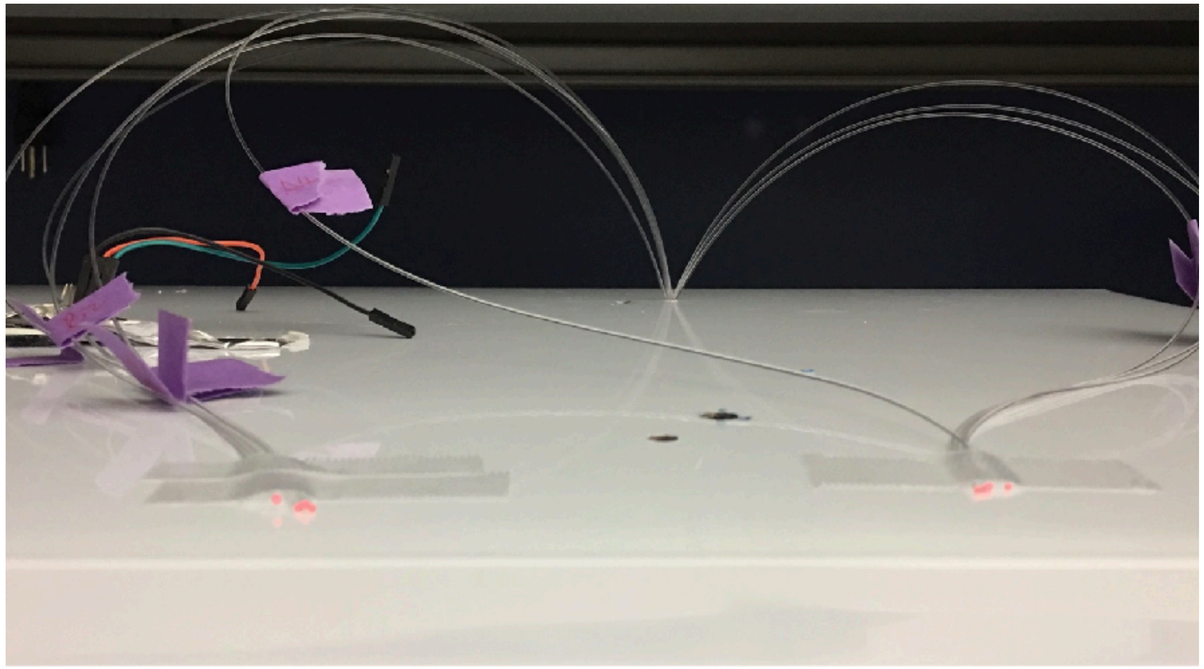


64 T2 neon lights on the turntable & cover

3 cm spacing < wavelength of 2.4 GHz (12.5 cm)

Optical fibers to conduct non-line-of-sight neon lights





Heat Actuation

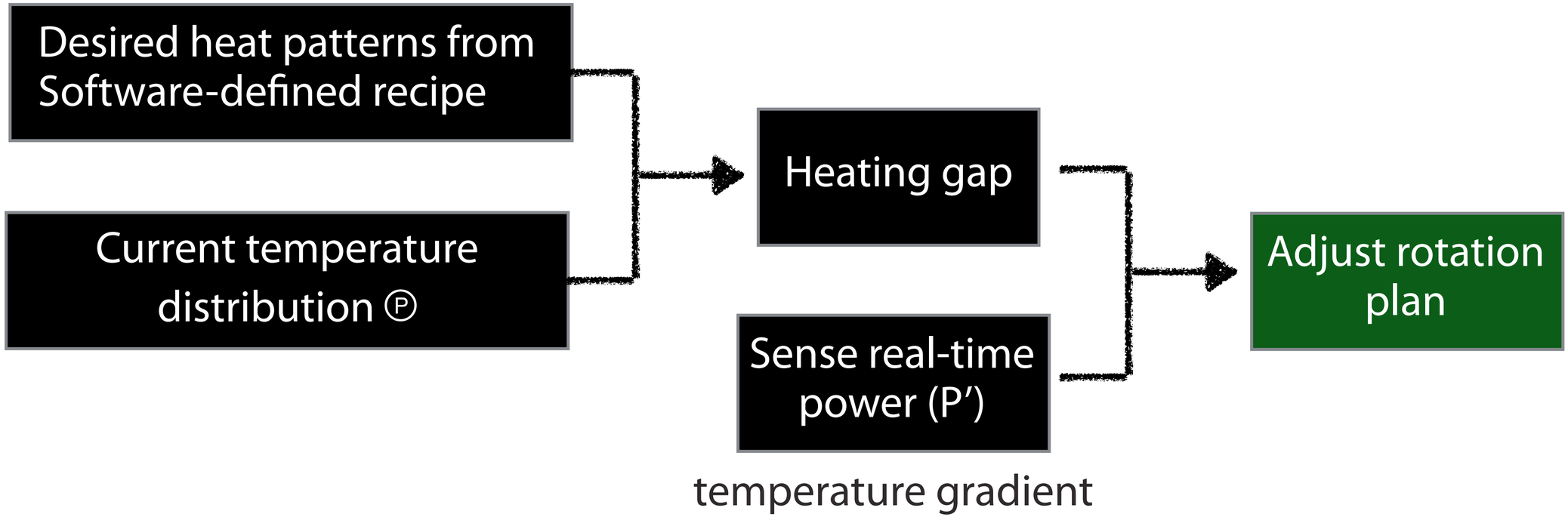
Blind rotation

turn table

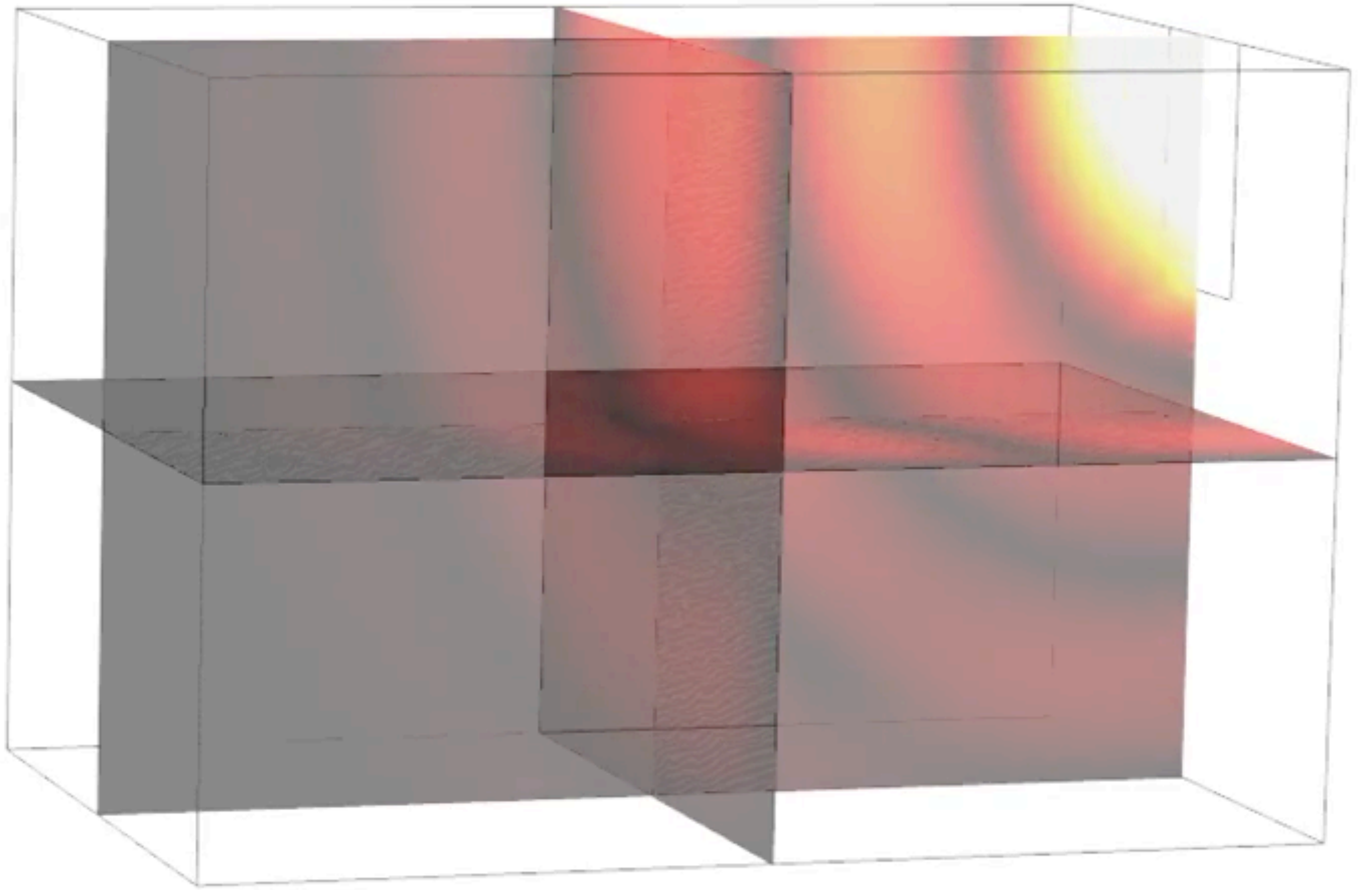
without turn table

with turn table

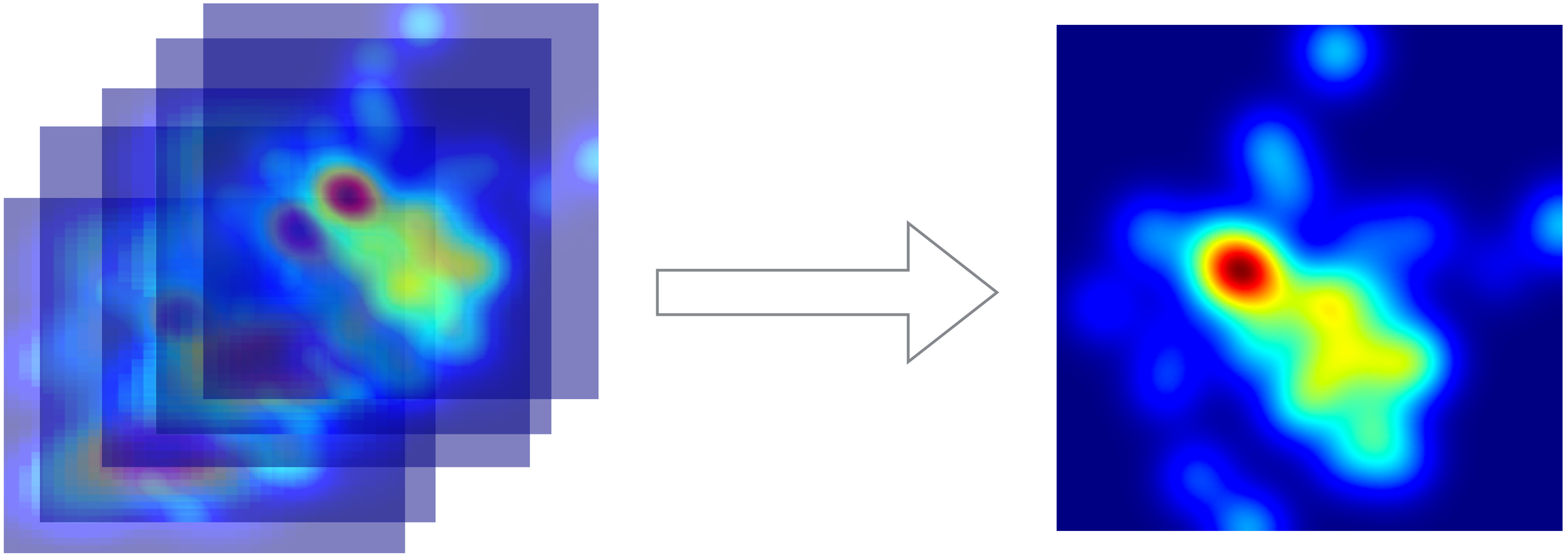
A closed-loop turntable



Microwave cannot heat individual pixels independently.



Knapsack problem



select a set of patterns whose union is equivalent to the target heat pattern.

Spoiler alert

No Turntable

Default Turntable

SDC Uniform Heating

Spoiler alert

No Turntable

Default Turntable

SDC Uniform Heating

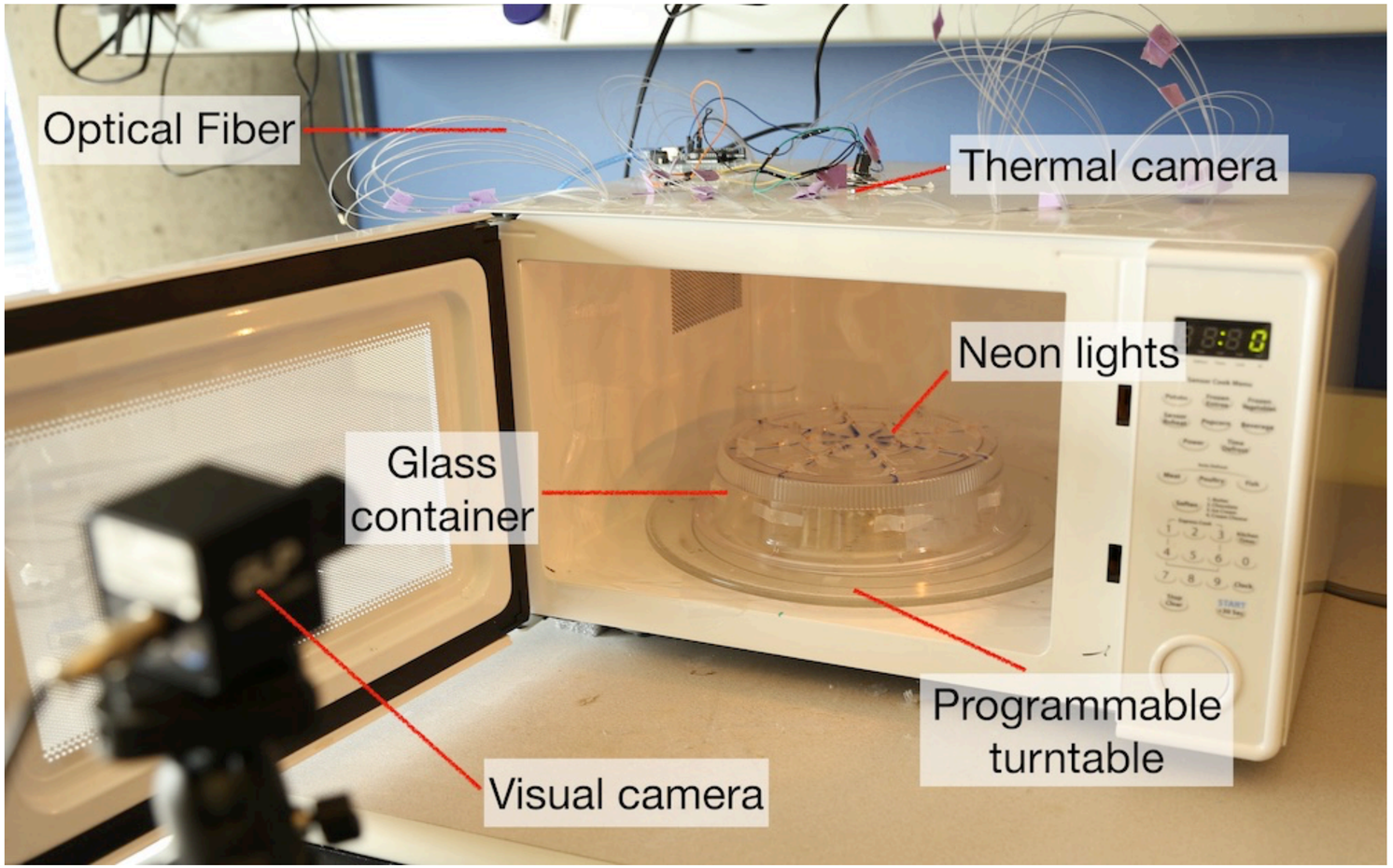
SDC Arbitrary Heating

patterned microwave susceptor

ensure **coverage** through SDC

Implementation & Evaluation

Prototype



Sharp SM1441CW ($110 from Ebay)

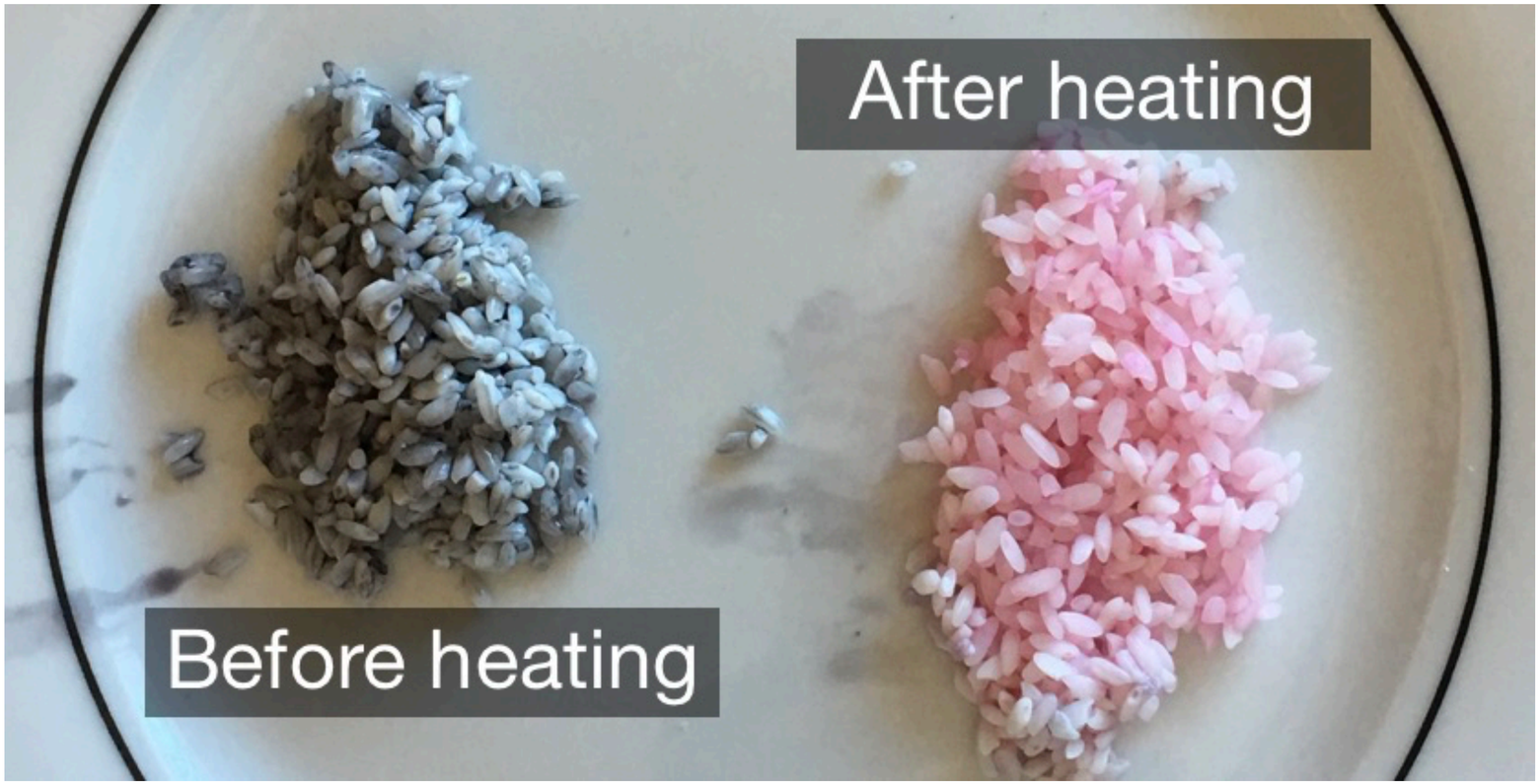
Neon lights

Camera x 2

Step motor

Arduino

Evaluation apparatus



thermal-chromatic pigment + rice

reusable

turn pink if p > 31 °C

Evaluation

1 **Uniform Heating:**

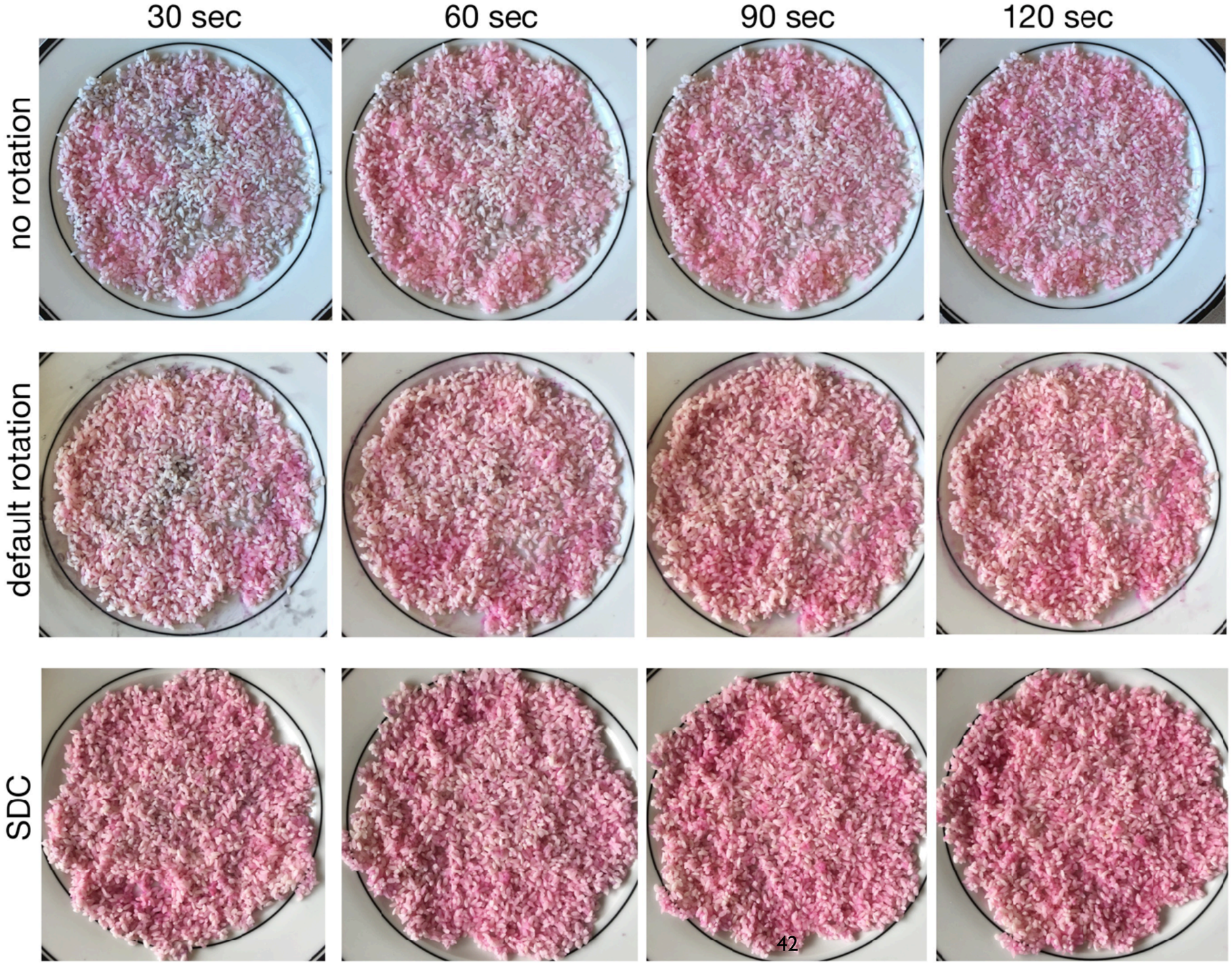
heat the rice in a plate uniformly to 60°C over 2 minutes.

2 **Arbitrary Heating:**

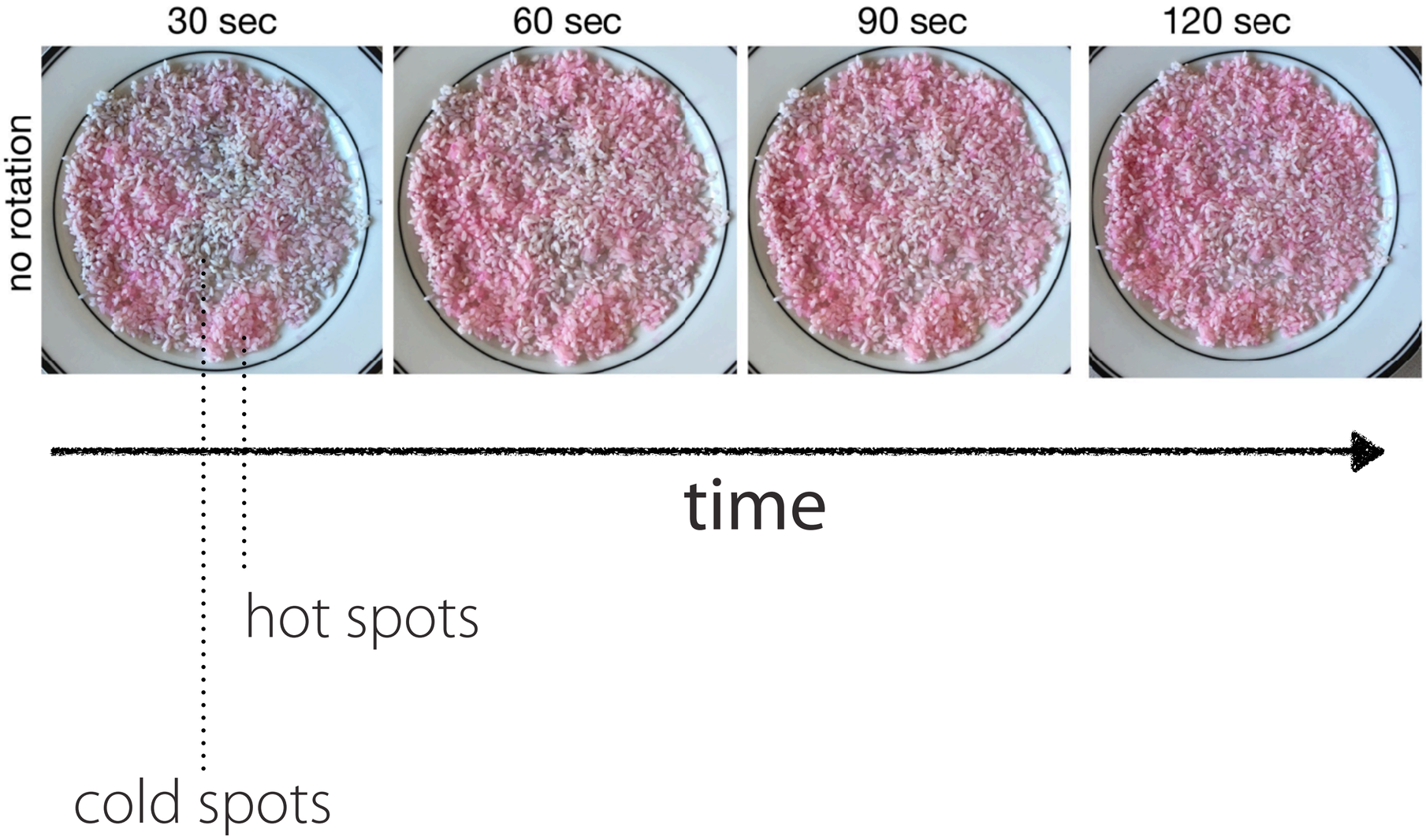
heat parts of the rice to 500°C while the rest are at 50°C.

(stress test)

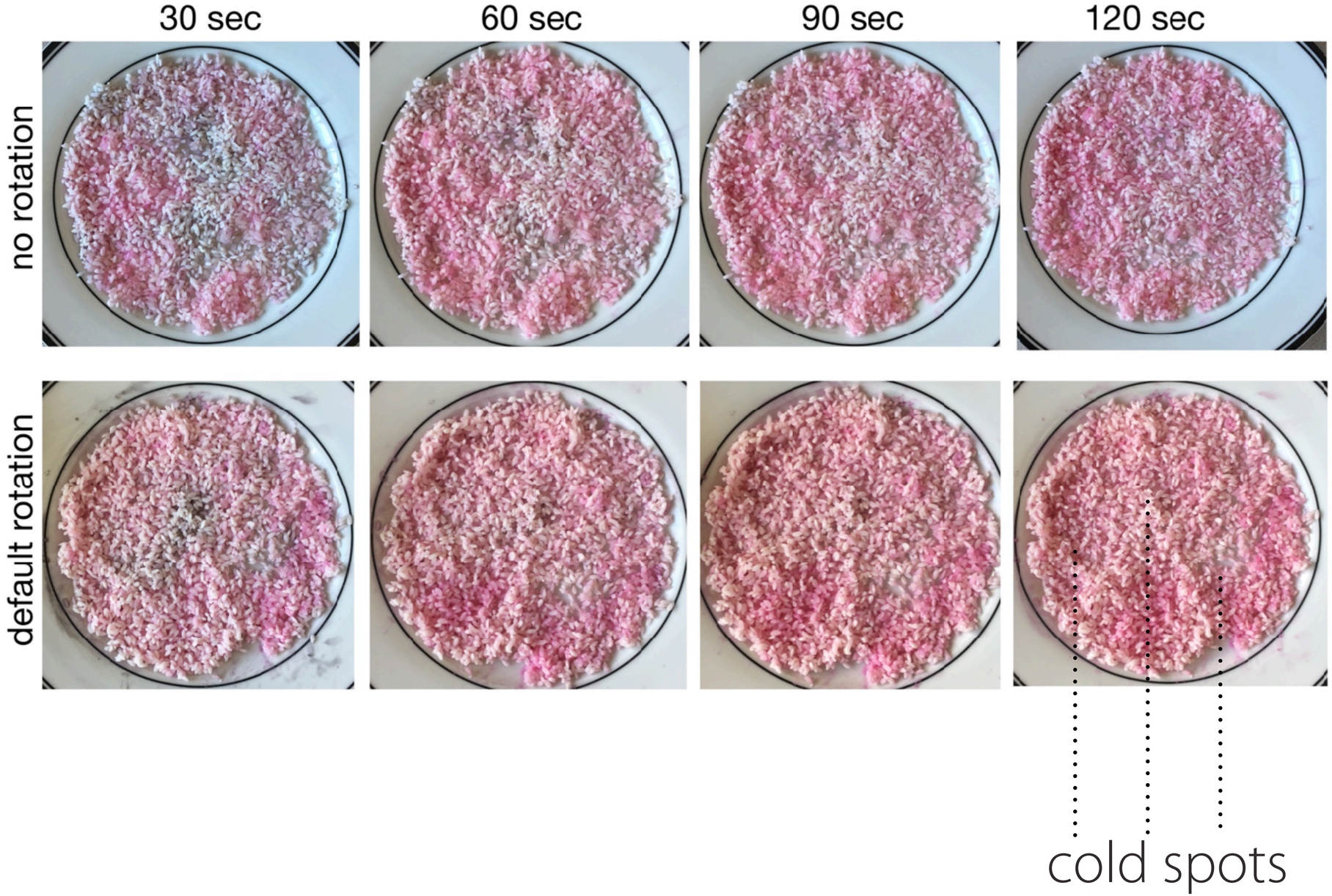
The room temperature is at 20°C.



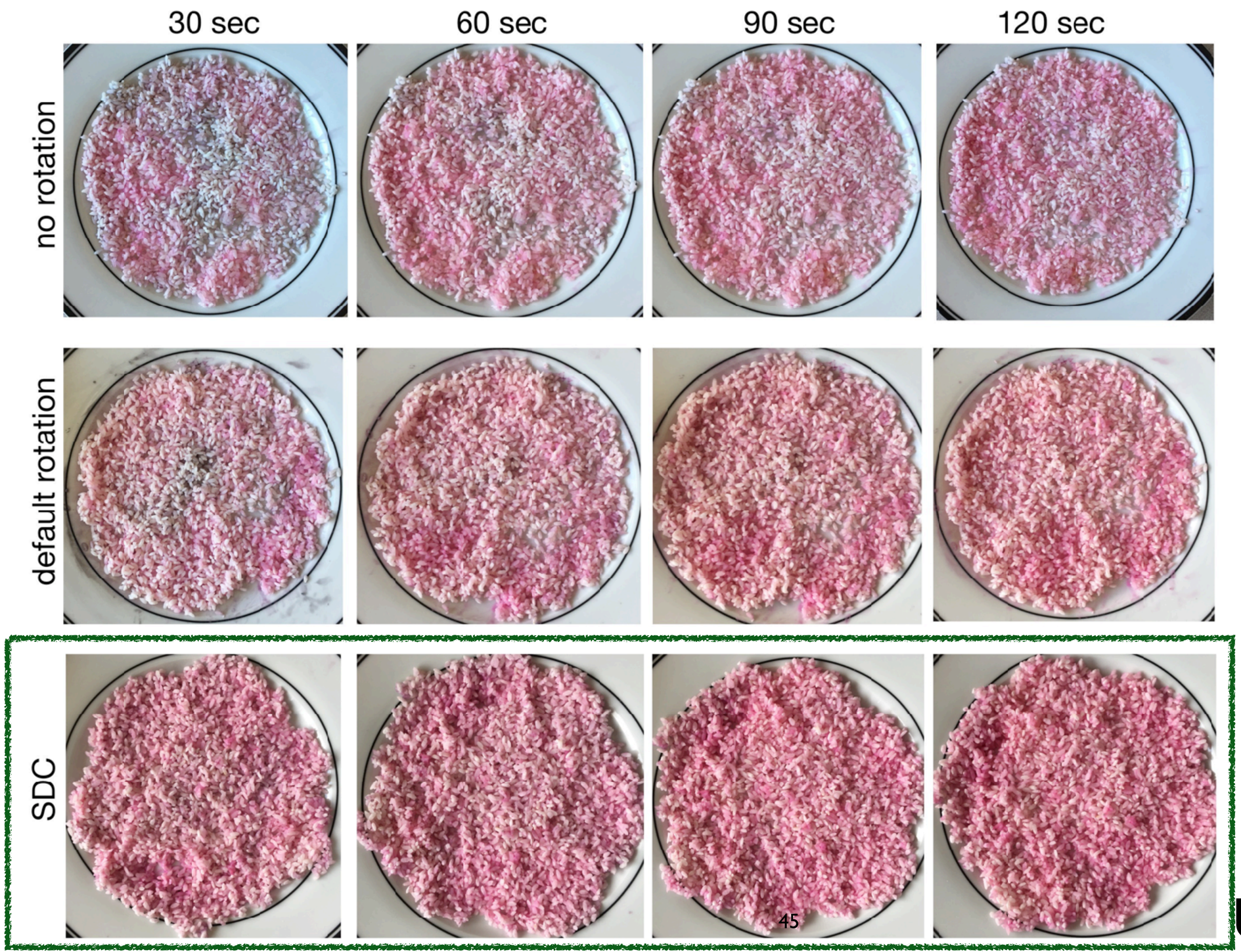
Uniform Heating



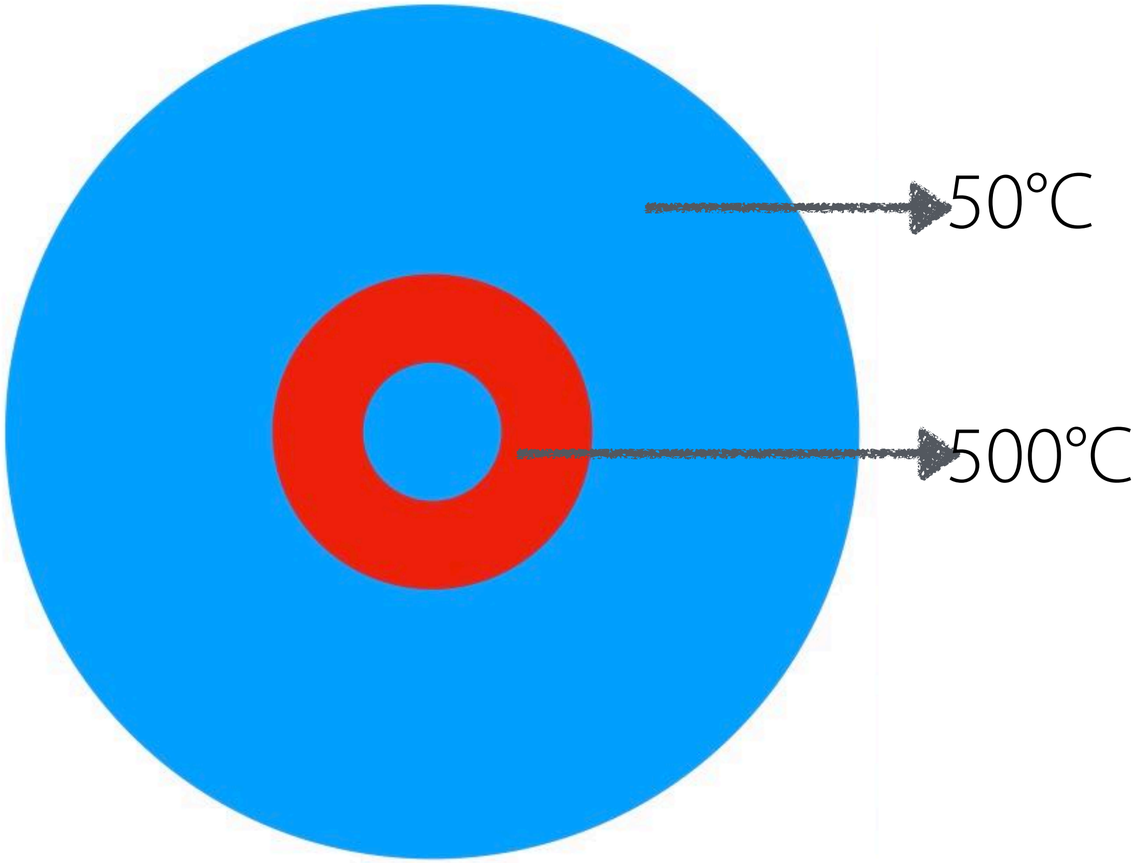
Uniform Heating



Uniform Heating



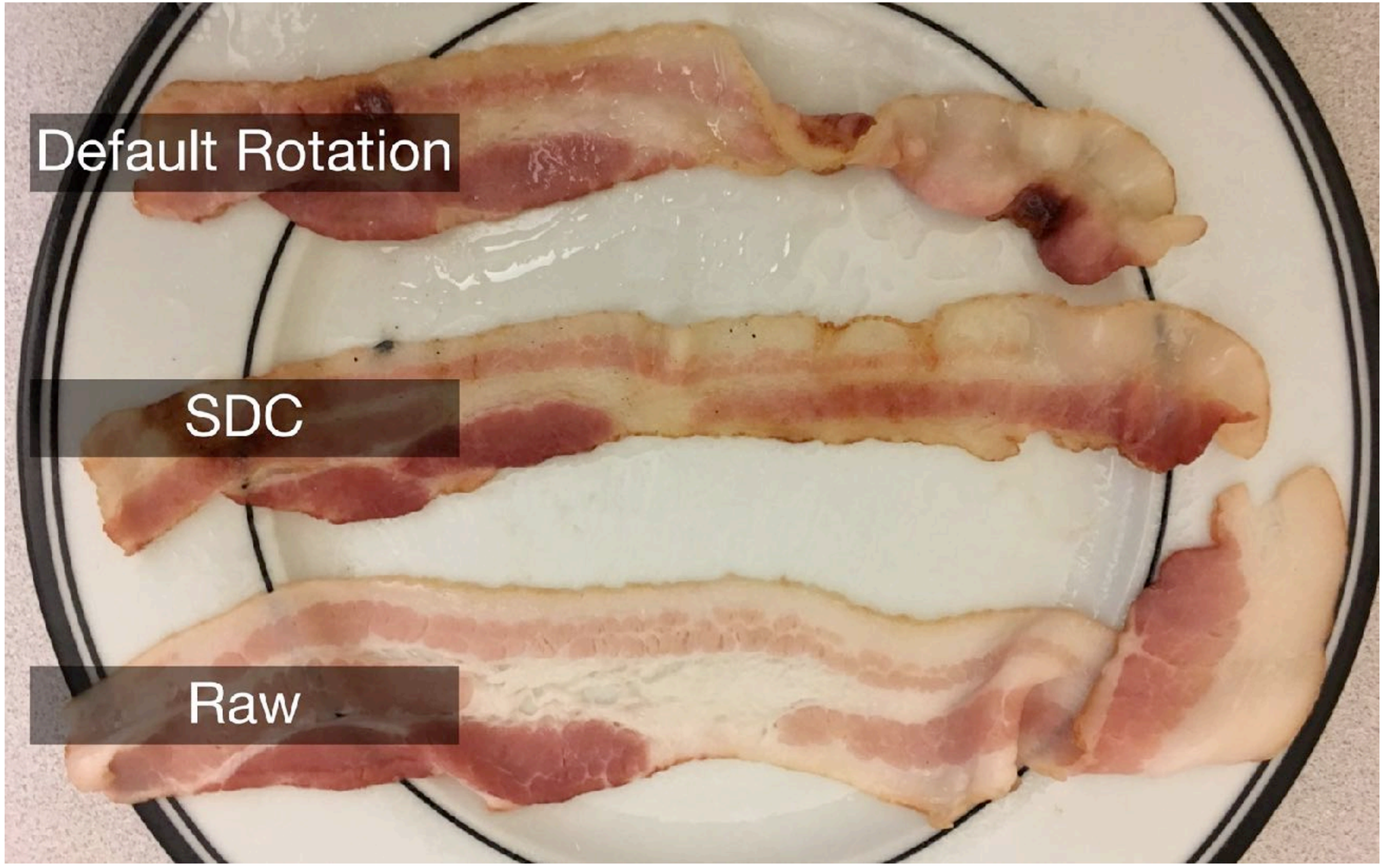
Uniform Heating



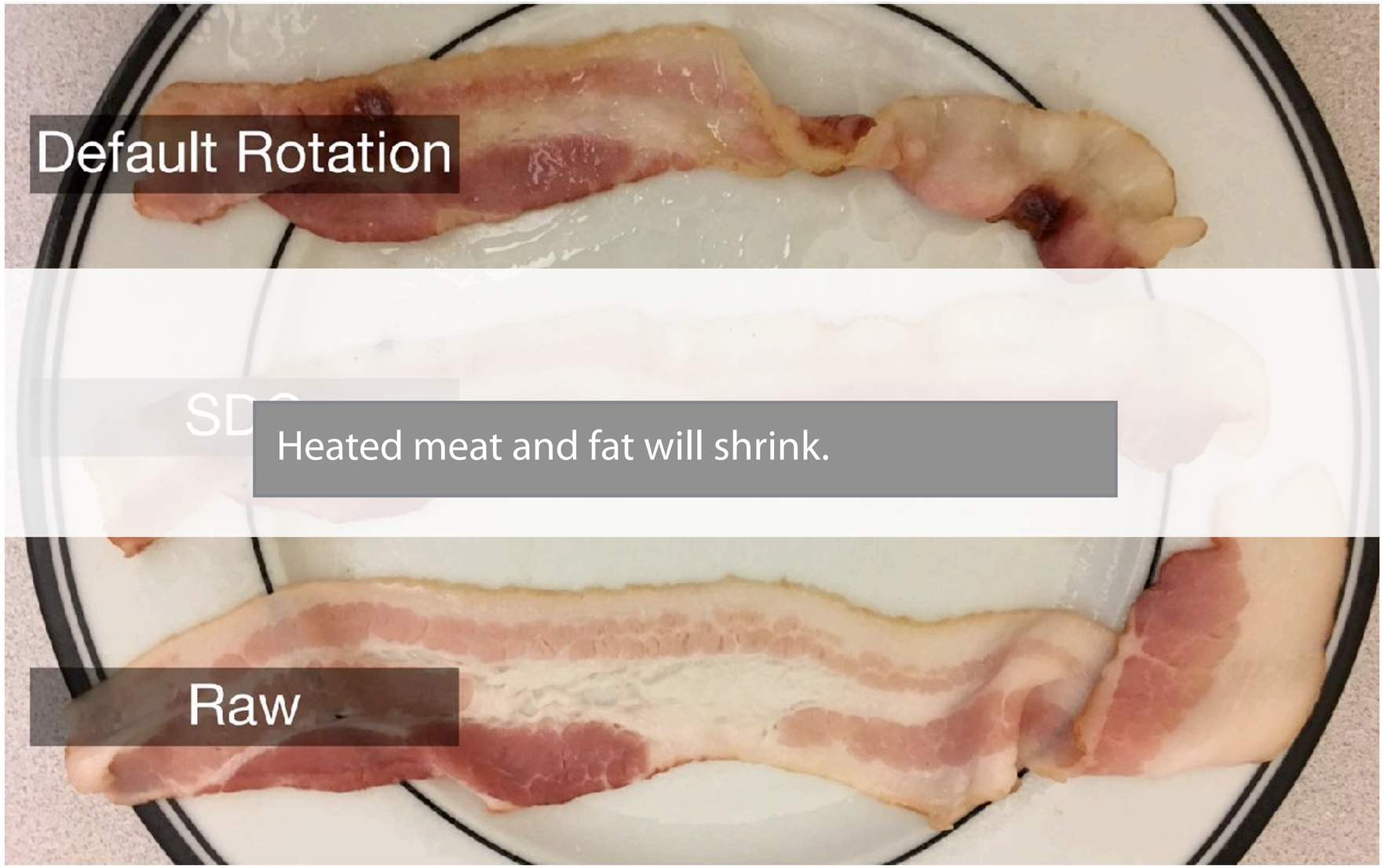
with a microwave susceptor ring

Maximum temperature delta: 183°C

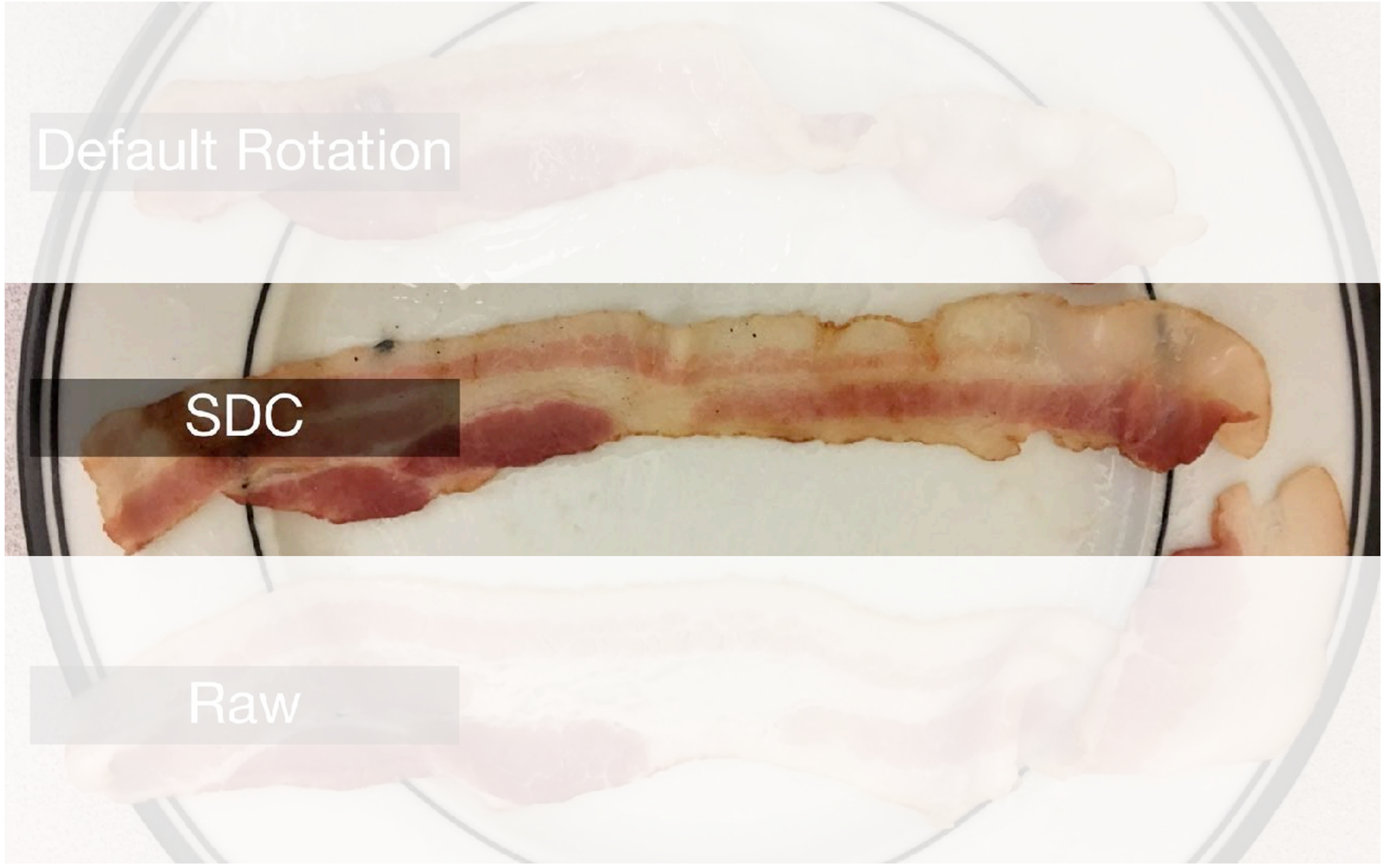
Arbitrary Heating



App: Cooking bacon
More apps => Paper

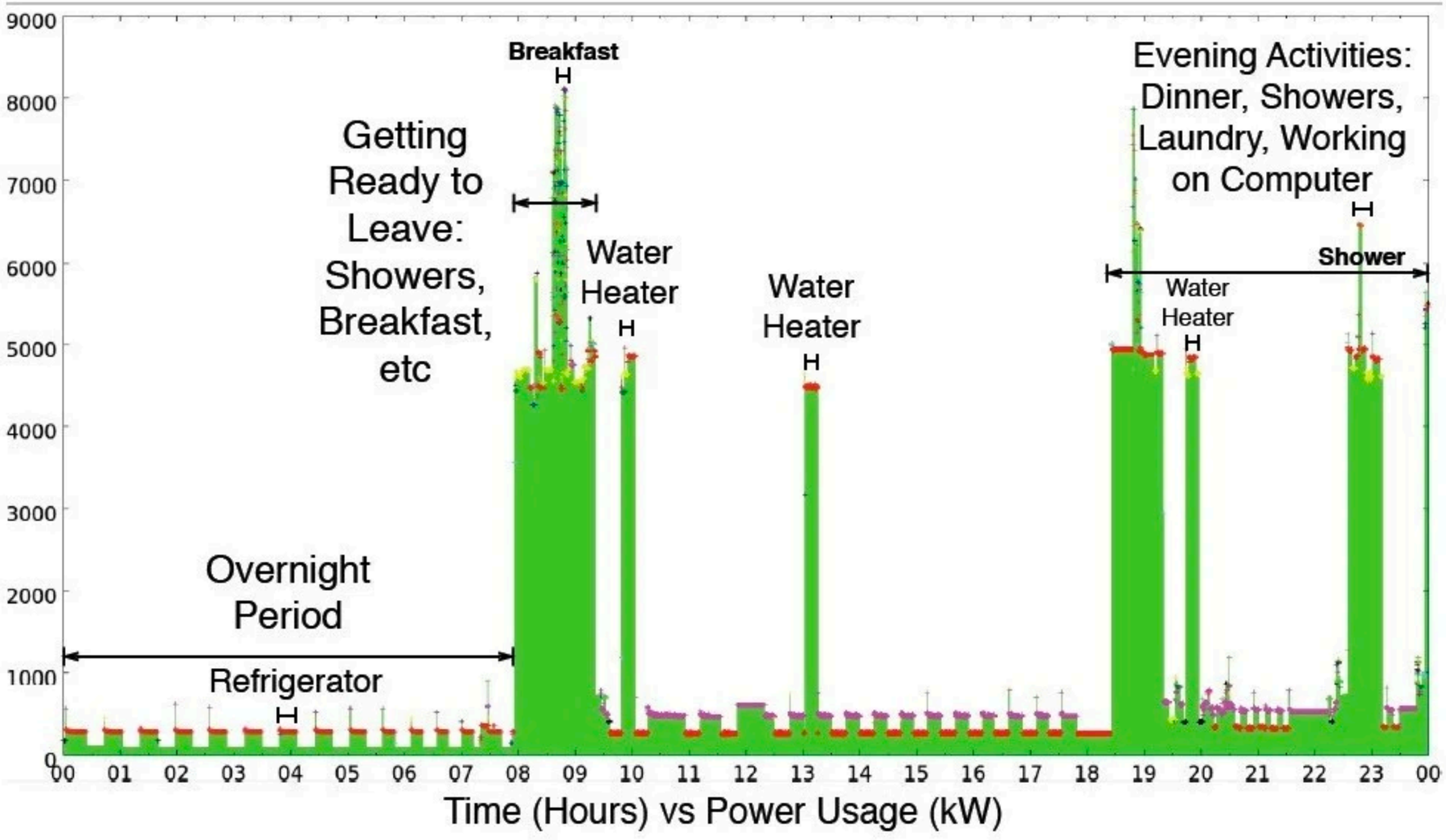


App: Cooking bacon
More apps => Paper



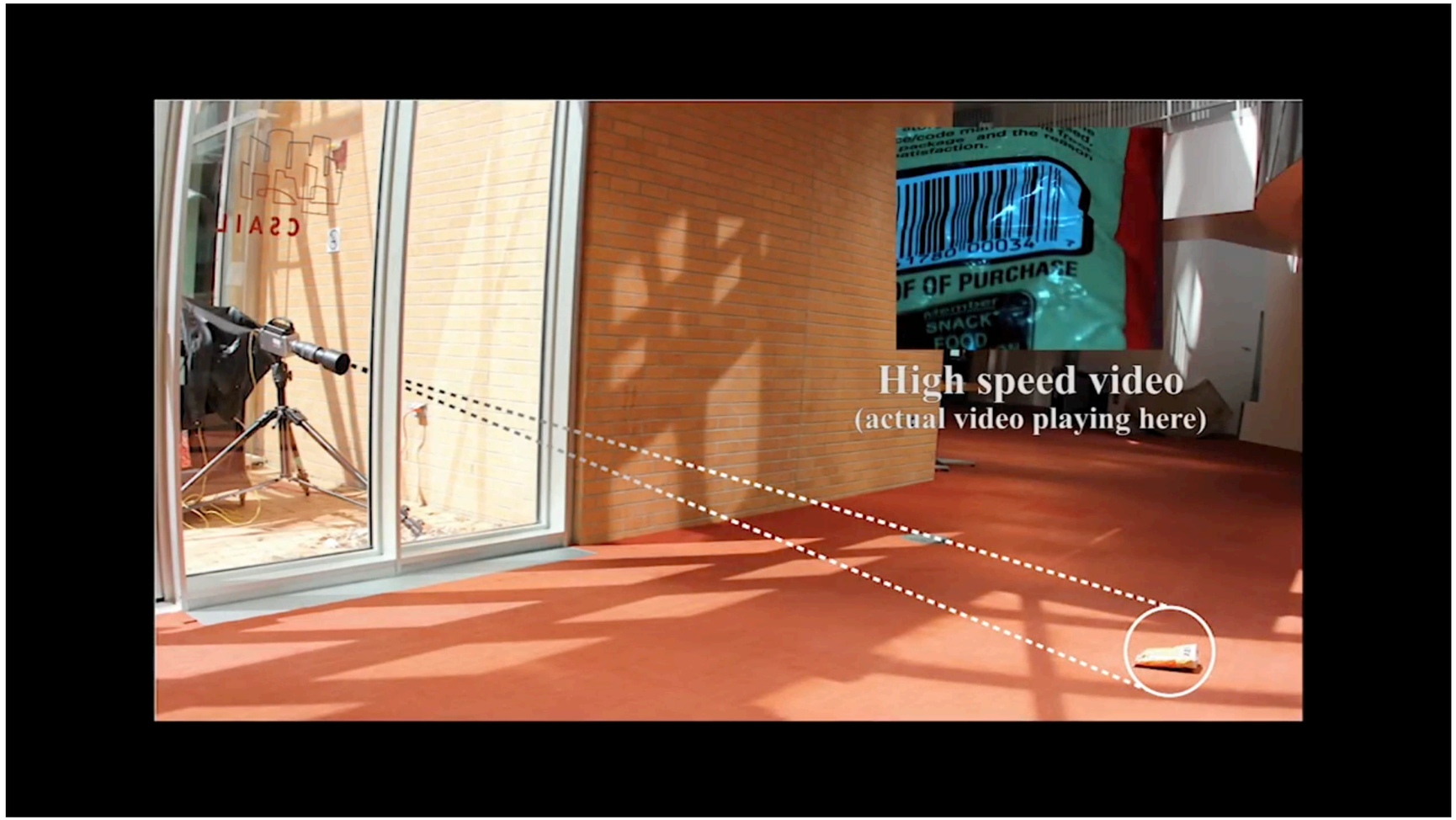
App: Cooking bacon
More apps => Paper

Smart devices will know everything about us!



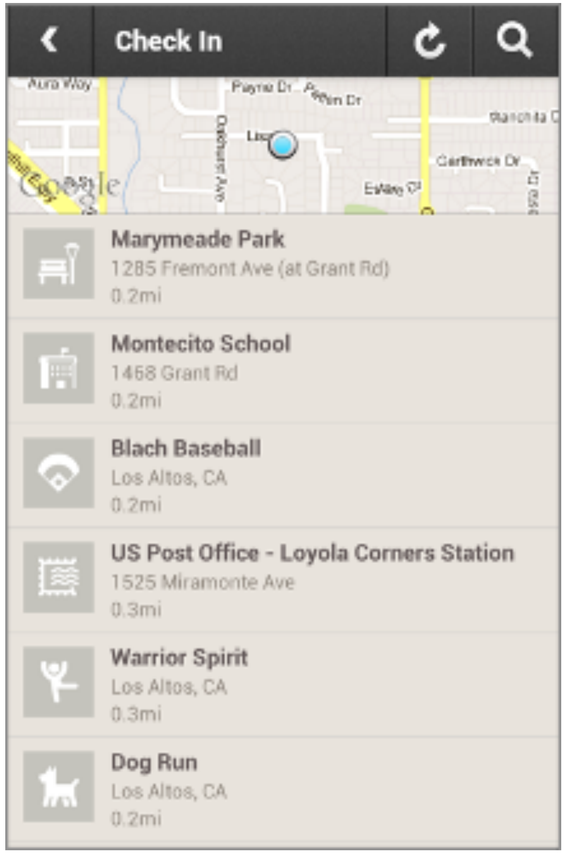
Private Memoirs of a Smart Meter, BuildSys 2010

Sky is the limit.



When sounds hits an object, it causes that object to vibrate.

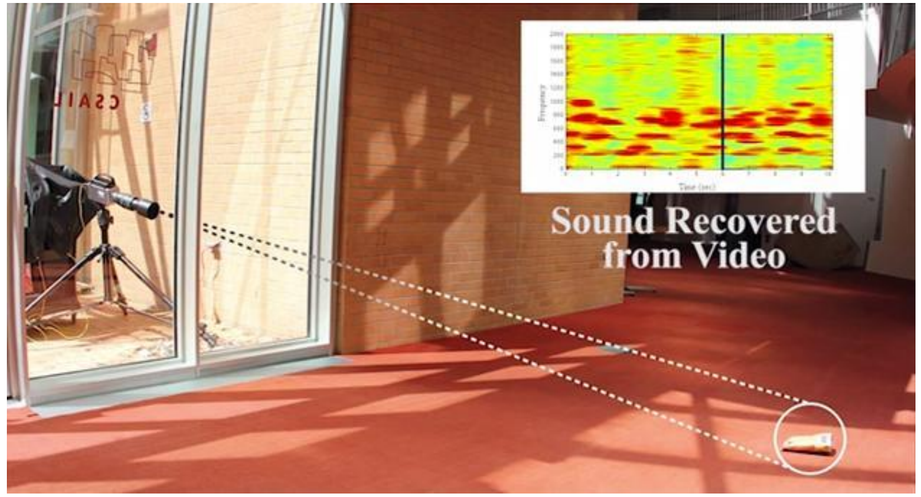
Users will lose control.

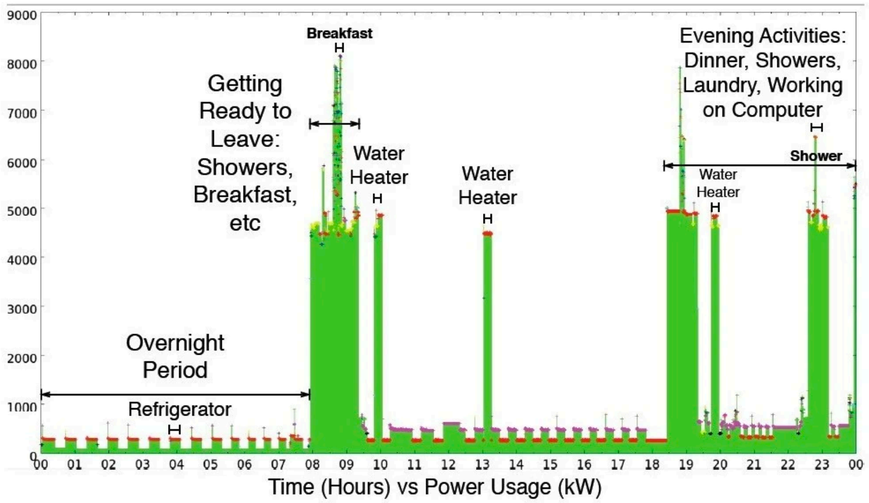


Foursquare

GirlsAroundMe

What's wrong? Or why is it wrong? How can we fix it?





Why is privacy hard?

GDPR CCPA

Law

Technology

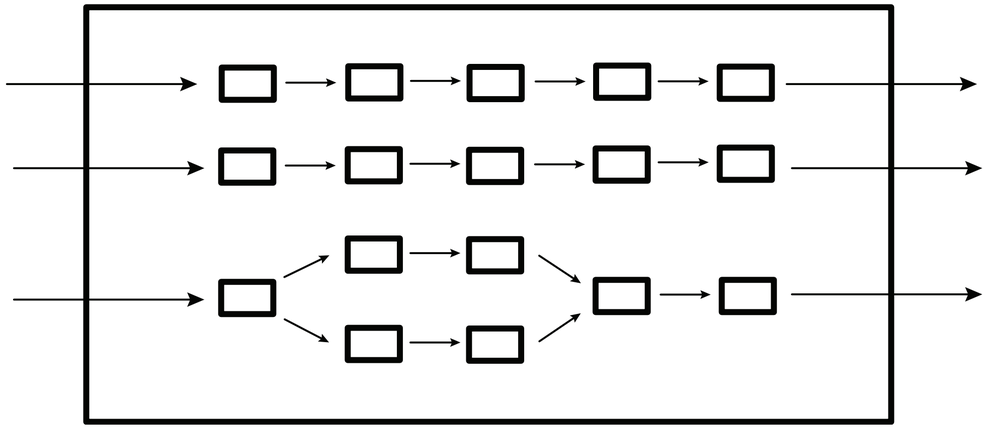
Users

Developers

This talk.

Software Defined Cooking

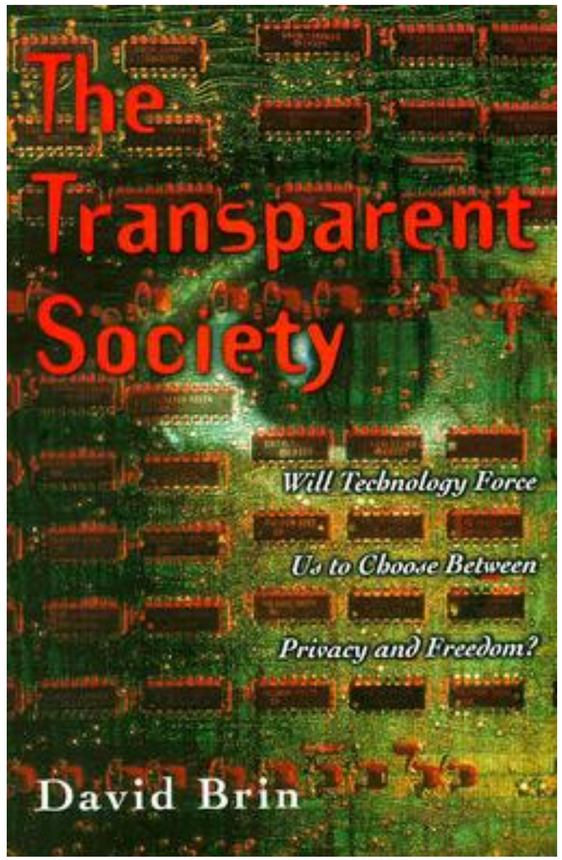
[MobiCom' 19, Featured in Communications of the ACM]



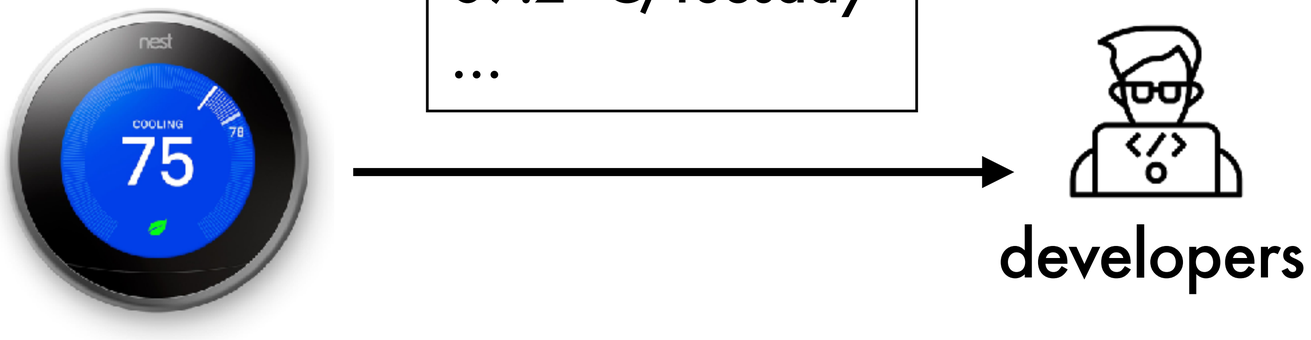
Modular Privacy Flows

[Ph.D. dissertation]

Sousveillance (inverse surveillance) - Transparency



How can Nest prove that they only collect aggregated data?



Open source?

Your TV watch history contains too much insights

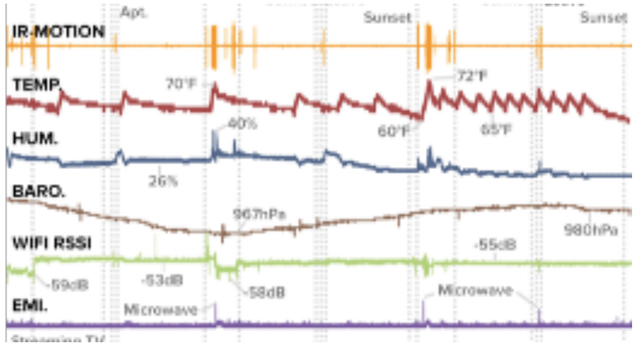
video #	duration	name	time	…
aaa	-	-	-	-
bbb	-	-	-	-
…	…	…	…	…

25 hours/week

How much time does the user spend on the TV?

- Is the user at home
- Activity routine
- User interests
- ….

Analyze data collection in 200+ smart home scenarios



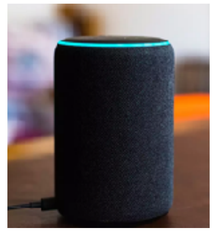
Sensing research (sensors)

Mobile privacy (developers) [1]

Design fiction interviews (users) [2]

[1] MobiPurpose, H. Jin et al, IMWUT'18/UbiComp'19
[2] Privacy Speeddating, H. Jin et al., SIGCHI'22

77% Apps do not need raw data.

	Sensor	*Raw*	*Needed data*
Hello visitor			
Noise level			55 db

Principle of data minimization

*"Personal data shall be limited to **what is necessary** in relation to the **purposes** for which they are processed."*

- GDPR, Article 5 (1) (c)

Best practice

Only collect the necessary data for a specific purpose.

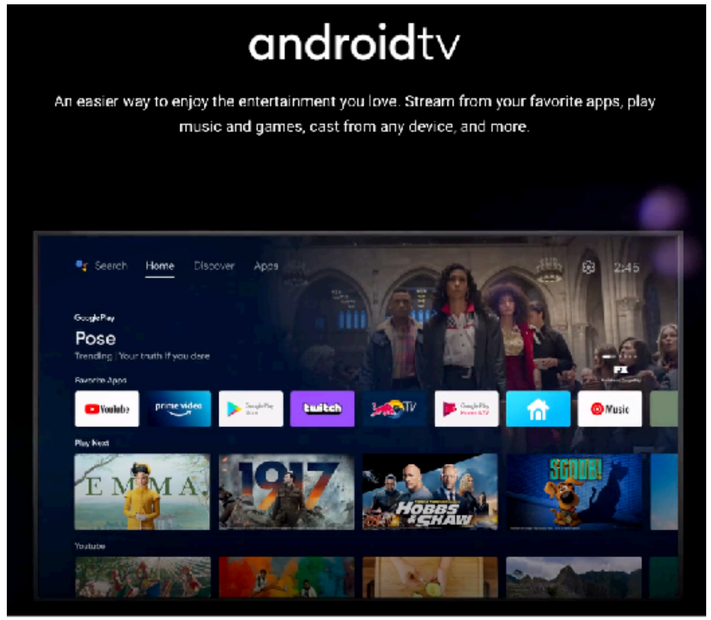
25 hours/week

Use weekly usage data to measure device engagement.

- ~~Is the user at home~~
- ~~Activity routine~~
- ~~User interests~~
- ....

How can developers prove themselves?

A strawman solution: fine-grained permission manifest



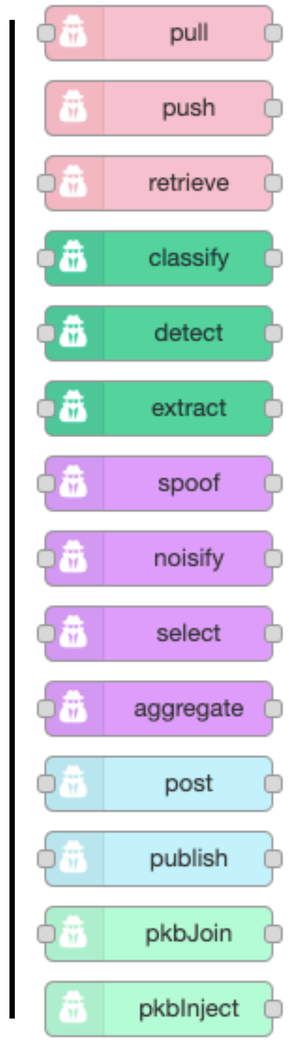
https://www.android.com/tv/

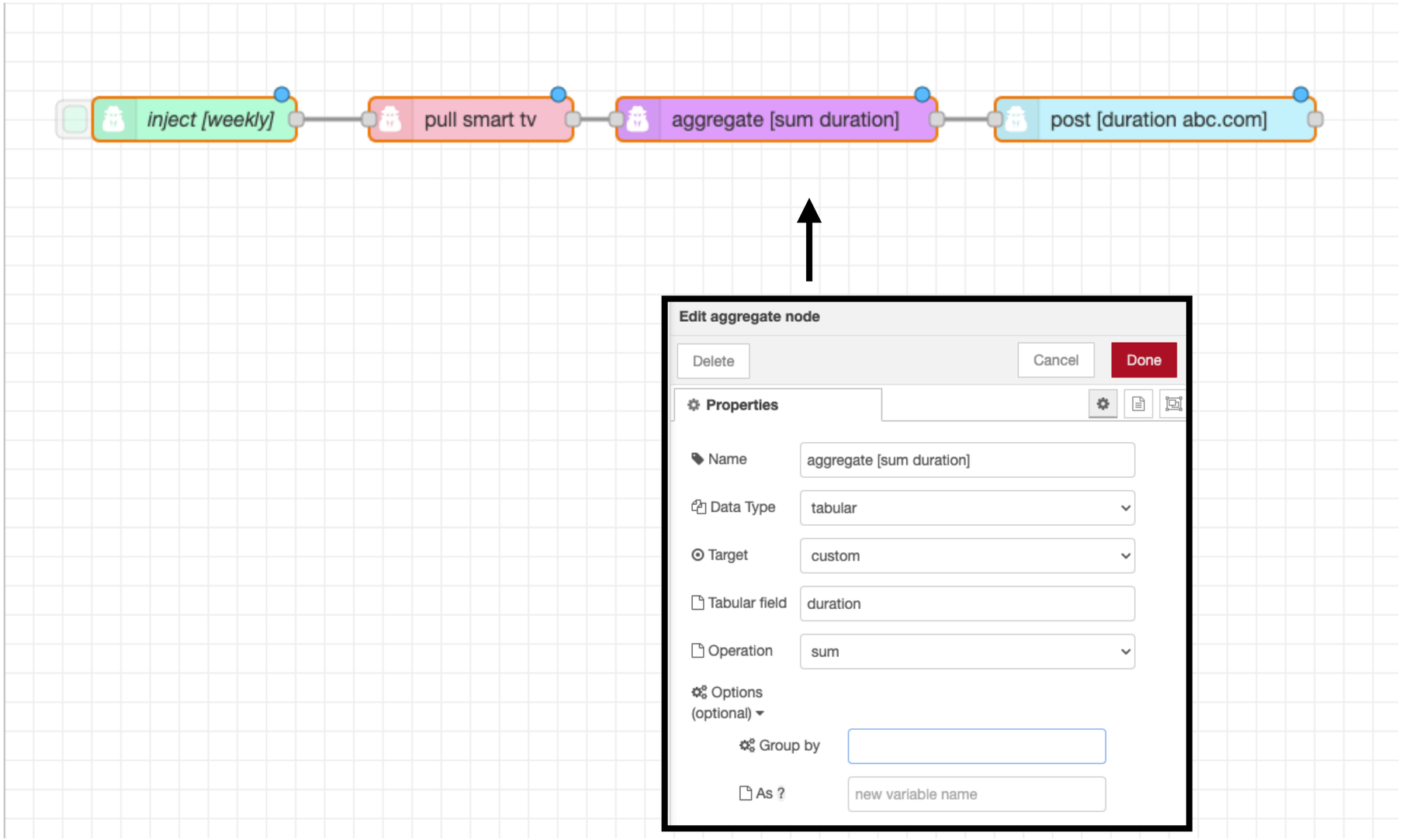
```
<manifest ...>
  <uses-permission android:name="android.permission.
      TV_AGGREGATED_DURATION_WEEKLEY" />

  <uses-permission android:name="android.permission.
      TV_AGGREGATED_DURATION_DAILY" />

  ......
</manifest>
```

Fine-grained permission manifest

Program pre-processing functions using chainable *operators*

A fixed set of operators





Peekaboo primitive (2)

A text-based whitelist *manifest* (i.e., program representation)

How much time does the user spend on the TV?

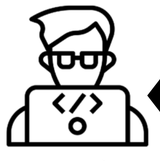
```
@purpose: To measure device engagement.
WeeklyUsageHours{
  // operator [properties]
  inject [weekly] ->
  pull [smart TV driver] ->
  aggregate [sum duration] ->
  post [duration]
}
```

Peekaboo primitive (3)

A trusted runtime with pre-loaded implementations

video #	duration	name	time	...
aaa	-	-	-	-
bbb	-	-	-	-
...	...	...	...	...

25 hours



How much time does the user spend on the TV?

A local hub

inject [weekly]

pull [smart tv]

video #	duration	name	time	...
aaa	-	-	-	-
bbb	-	-	-	-
...	...	...	...	...

aggregate [sum duration]

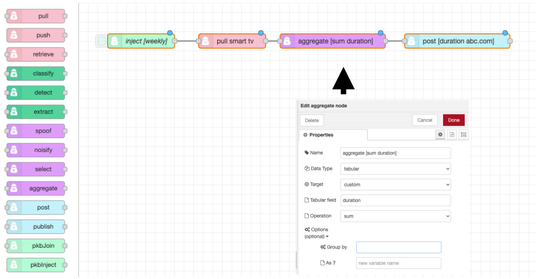
25 hours

post [duration]

```
@purpose: To measure device engagement.
WeeklyUsageHours{
  // operator [properties]
  inject [weekly] ->
  pull [smart TV driver] ->
  aggregate [sum duration] ->
  post [duration]
}
```

A text-based manifest

Smart home app store



Programming environment with operators

Runtime with preloaded implementations

App developers

```
@purpose: To measure device engagement.
WeeklyUsageHours{
  // operator [properties]
  inject [weekly] ->
  pull [smart TV driver] ->
  aggregate [sum duration] ->
  post [duration]
}
```

Manifest

Peekaboo adoption

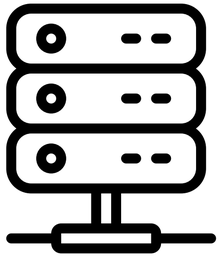
Smart home app store

Smart home app →

```
@purpose: To measure device engagement.
WeeklyUsageHours{
  // operator [properties]
  inject [weekly] ->
  pull [smart TV driver] ->
  aggregate [sum duration] ->
  post [duration]
}
```

Edge devices

A local hub



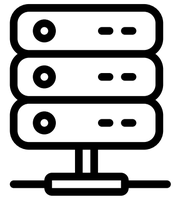
Cloud

"Privacy firewall"

Pre-process users' data to mitigate data overaccess.

video #	duration	name	time	…
aaa	-	-	-	-
bbb	-	-	-	-
…	…	…	…	…

25 hours/week

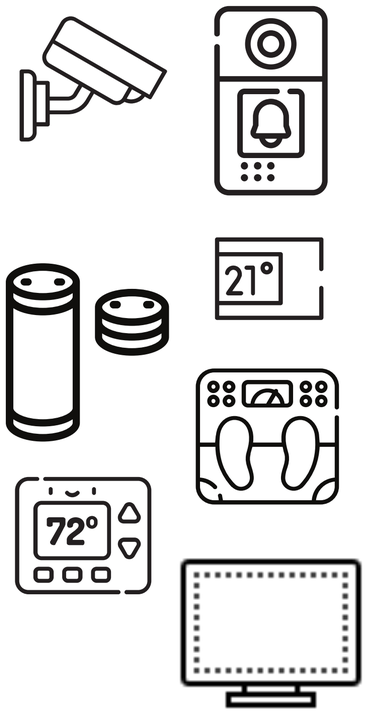


Edge devices

Hub

Cloud

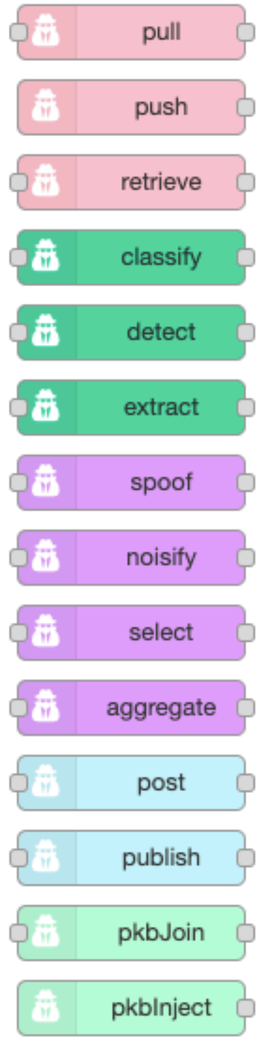
A fixed set of operators



Edge devices

video, image, audio, tabular, scalar

A *fixed* set of operators



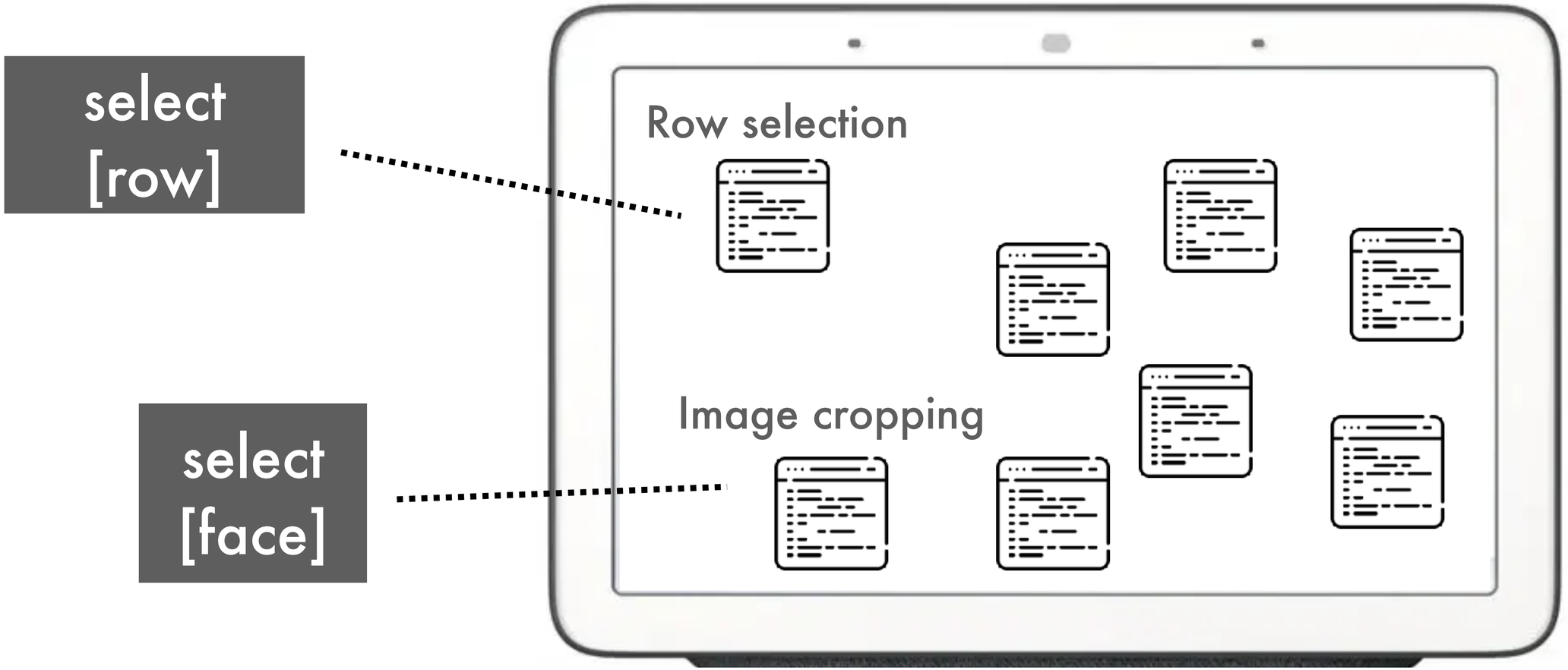
An operator = A verb keyword

select [row]

	product_id	product_name	inventory_received	starting_inventory	inventory_on_hand	minimum_required
1	2	Booth	29pcs	27pcs	56pcs	20pcs
2	3	Maclean	23pkts	25pkts	48pkts	25pkts
3	4	Closeup	24pkts	25pkts	49pkts	25pkts

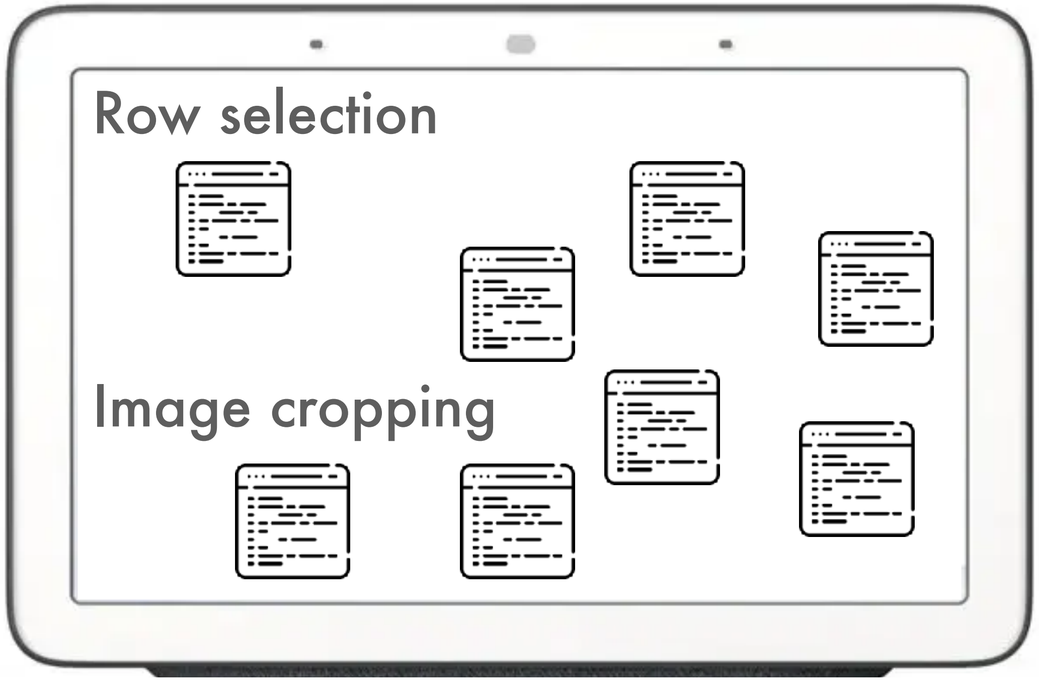
→ detect [face] → select [face] →

Operators are mapped to pre-loaded implementations



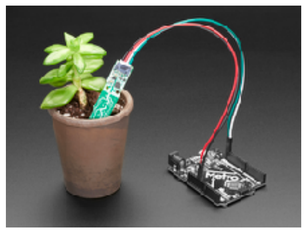
A small set of pre-processing algorithms improve privacy

video #	duration	name	time	...
aaa	-	-	-	-
bbb	-	-	-	-
...	...	...	...	...



25 hours/week

Implementation (hardware)



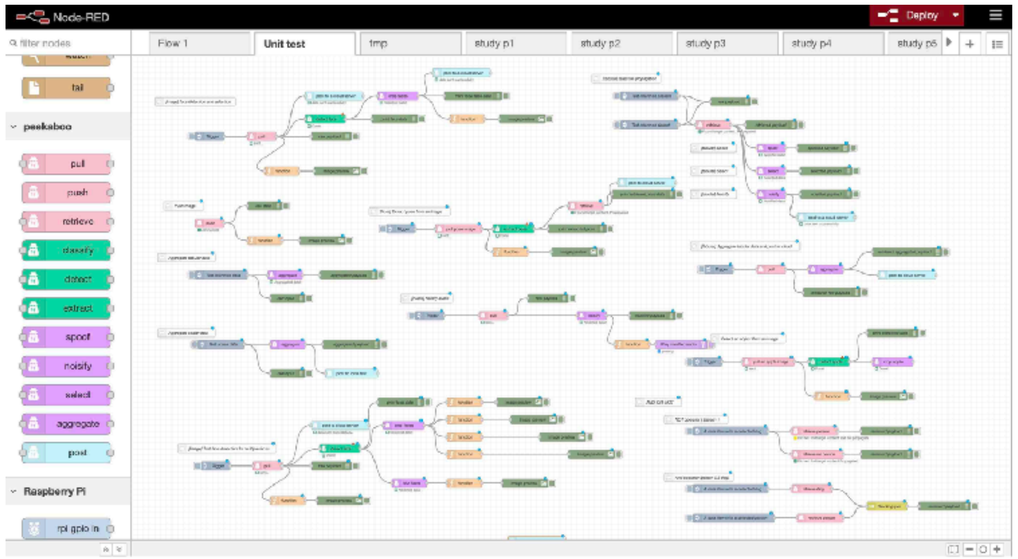
Edge devices

Raspberry PI + TPU

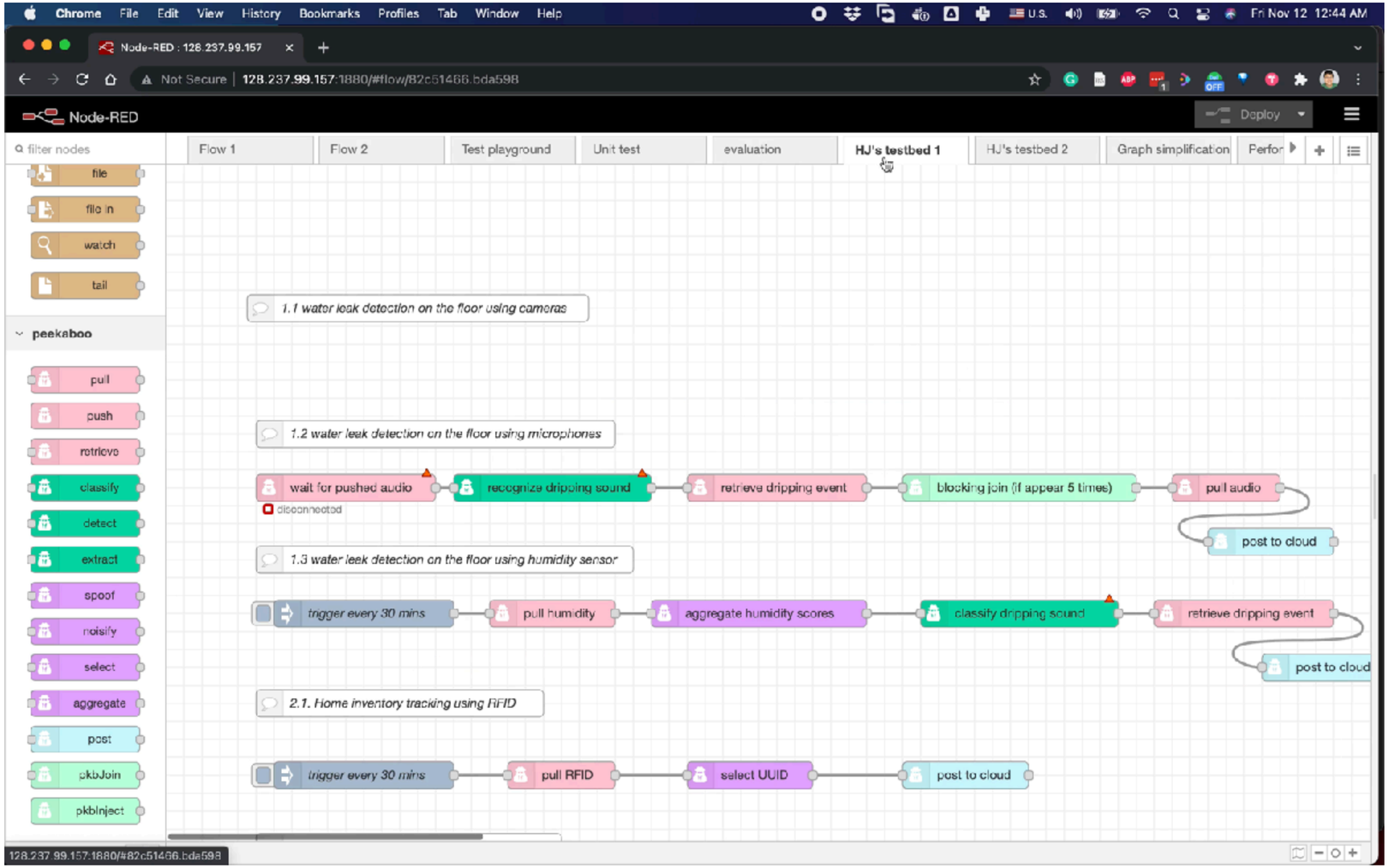
Cloud

Implementation (software)

1. Operators: Node.JS package
2. Programming IDE: NodeRed
3. Drivers: 5 data types
4. 23 Preloaded implementations



Expressiveness (200+ smart home cases)



Evaluation

System performance

≈$100

25 inference/s

100 filtering/s

1-80 ms per request

Utility privacy tradeoff example

incognito voice assistant

6 speakers
112 audio files [1]

noisify <5% random pitch shift

Microsoft Cognitive Services

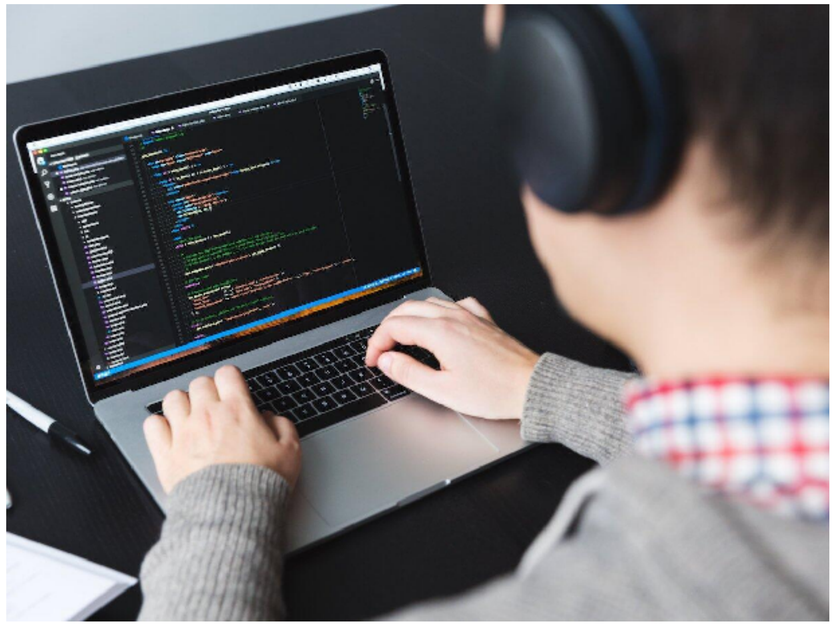
Speech word error rate:
9.27% → 11.88%

Speaker recognization:
100% → 27.7%

[1] CMU PDA Speech Database

Evaluation

Developer studies



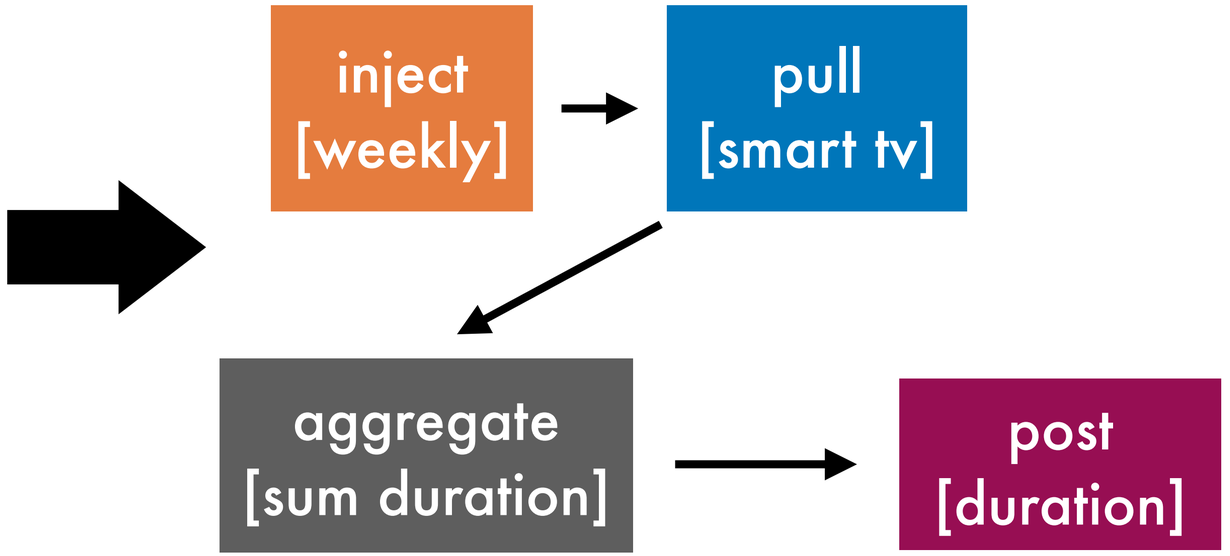
Task descriptions

IDE & Unit tests

6 - 15 mins to
author a manifest

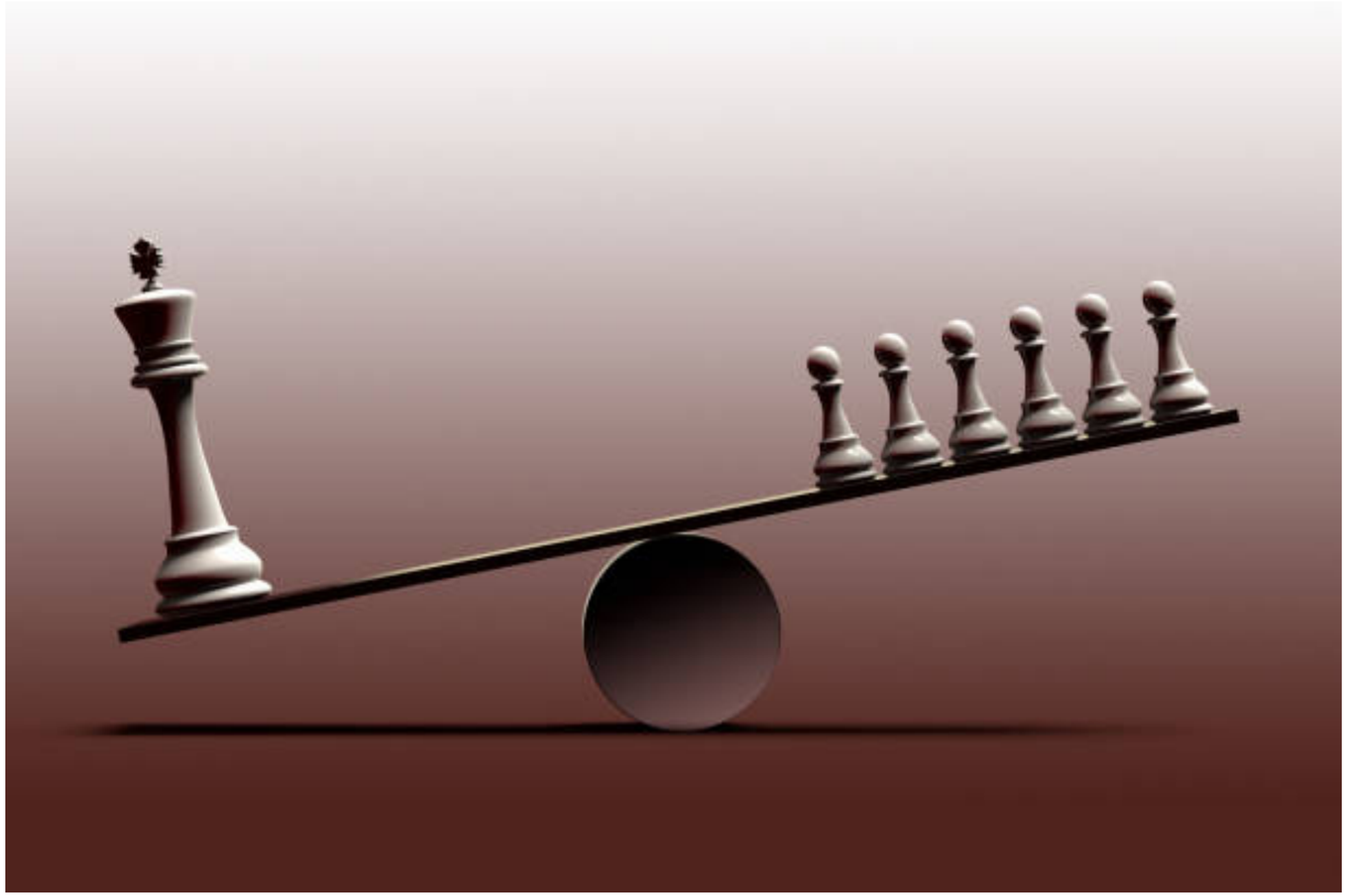
Manifests enforce fine-grained data collection

```
@purpose: To measure device engagement.
WeeklyUsageHours{
  // operator [properties]
  inject [weekly] ->
  pull [smart TV driver] ->
  aggregate [sum duration] ->
  post [duration]
}
```



public, non-proprietary

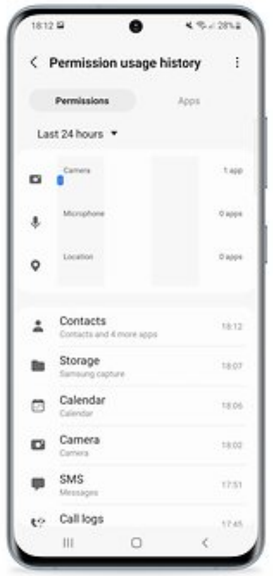
User-developer privacy negotiation

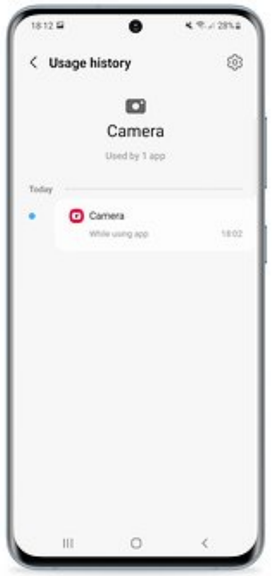


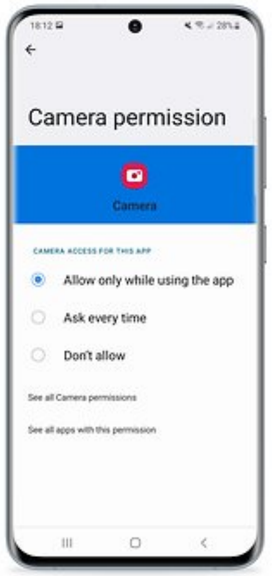
- *Empower users (e.g., Ad-blocker)*
- *Empower good developers*
- *Empower trusted third-parties*

Repetitive implementation and *distributed* interfaces

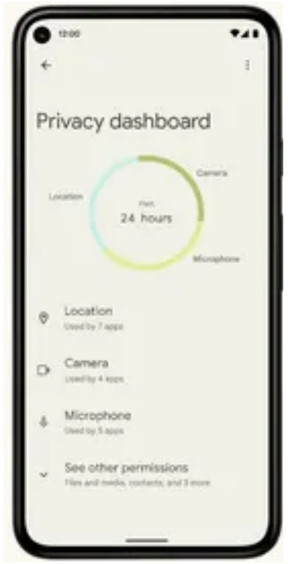
Samsung







Nest



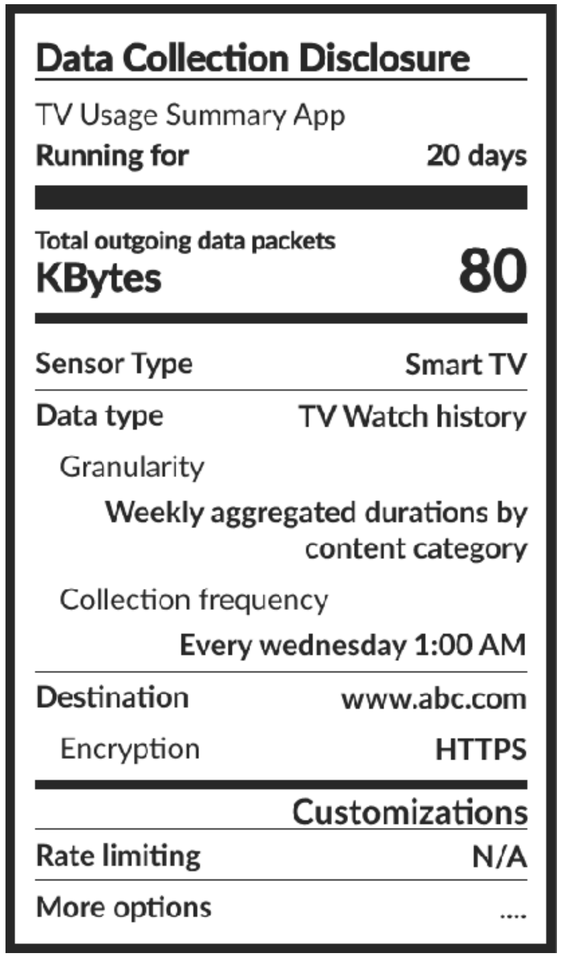
Small developers?

Users?

Advantages

Manifests → *enforceable/dynamic* privacy nutrition labels

```
@purpose: To measure device engagement.
WeeklyUsageHours{
  // operator [properties]
  inject [weekly] ->
  pull [smart TV driver] ->
  aggregate [sum duration] ->
  post [duration]
}
```

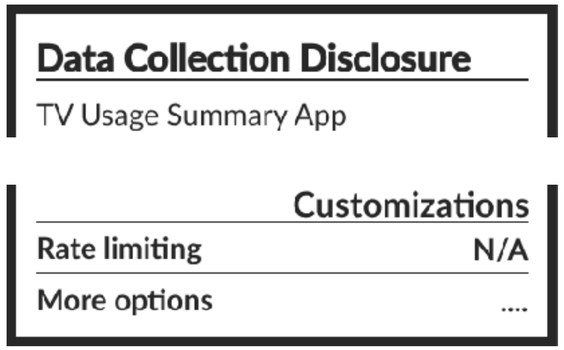


[1]

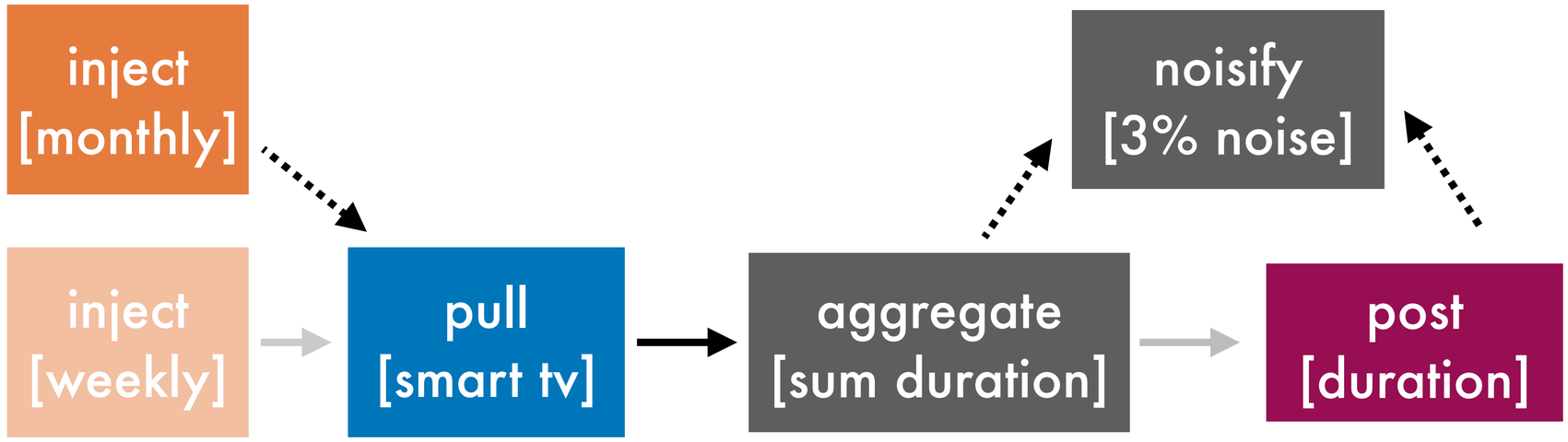
[1] Security and Privacy "Nutrition" Label, P. Emami-Naeini et al, IEEE S&P'20

Advantages

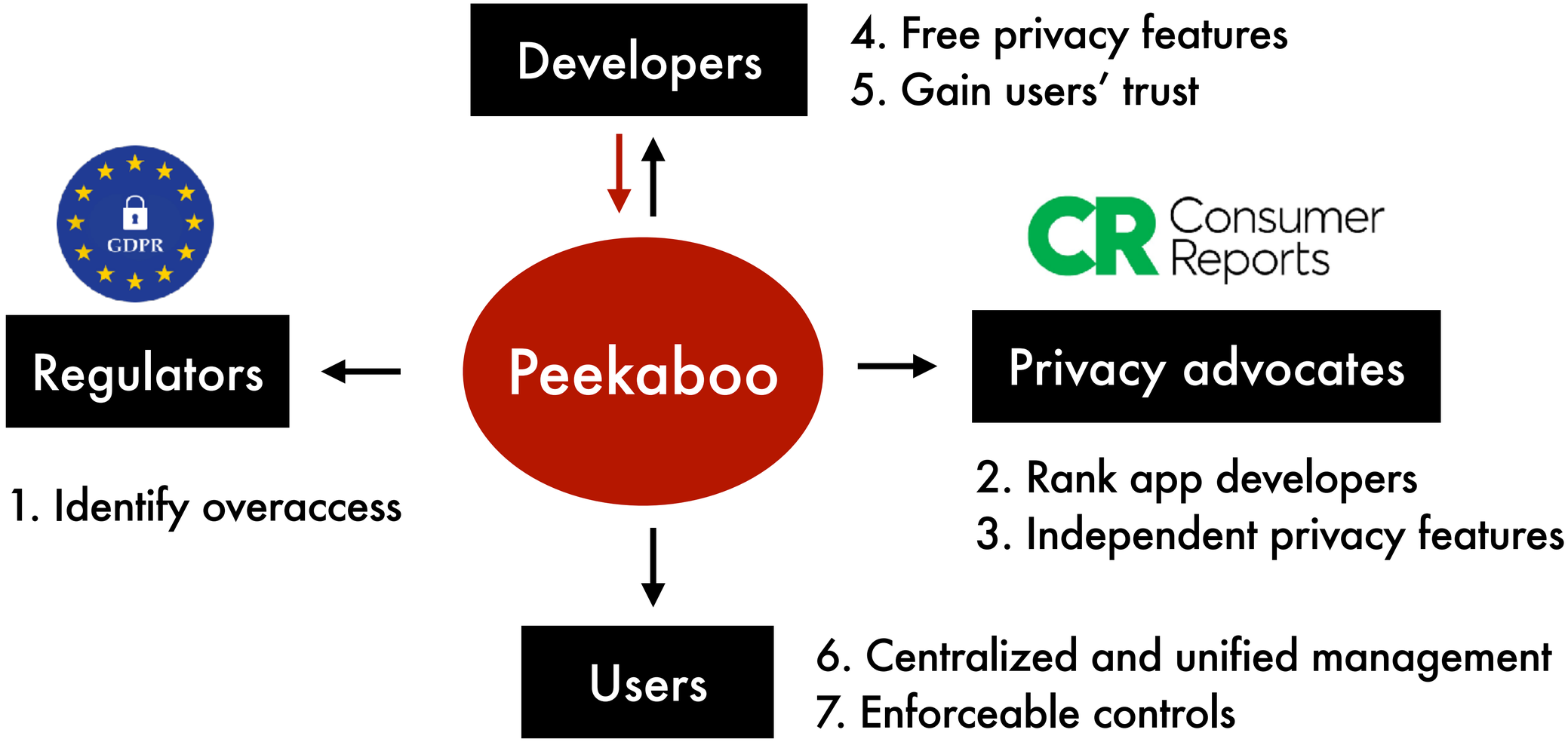
Built-in fine-grained control through manifest rewriting



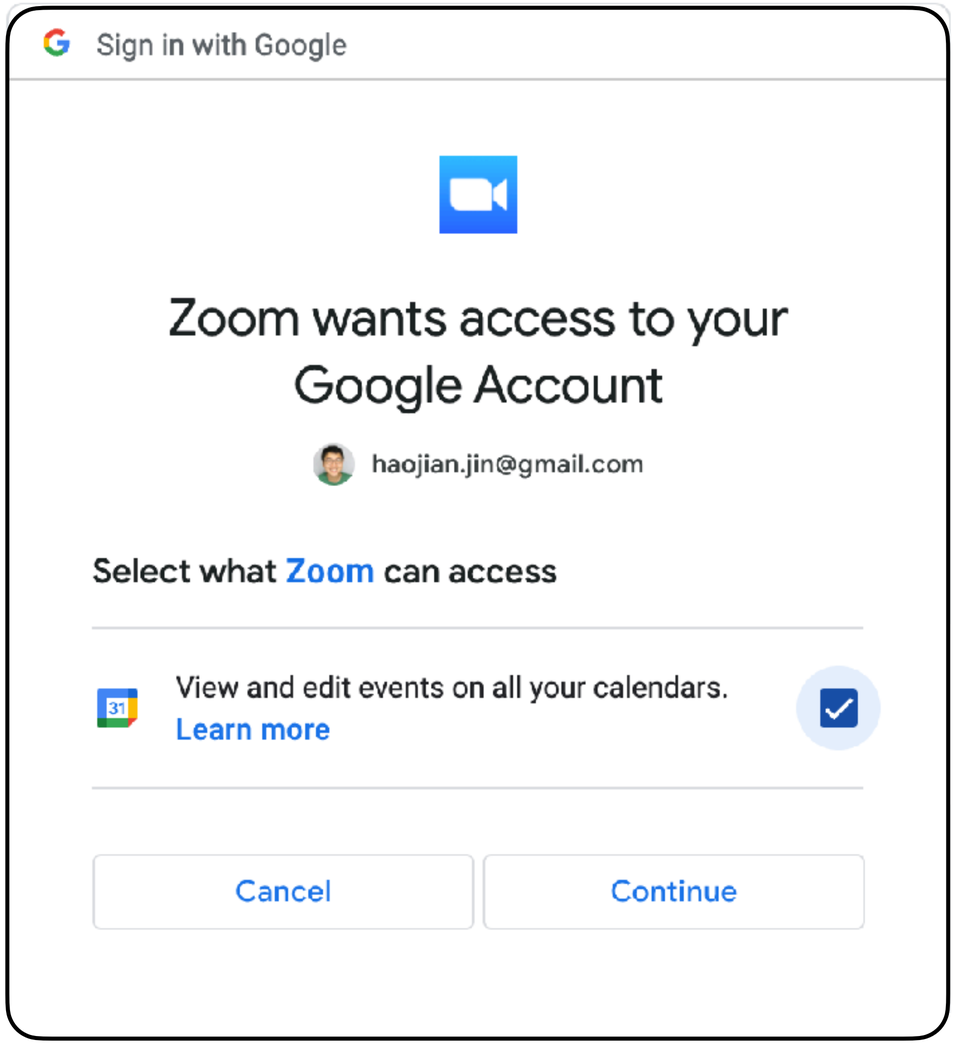
Change the rate to monthly



Let the good privacy drive out the bad privacy



Zoom accesses **all** your calendar events **continuously**!



Calendar events that contain https://zoom.us/xxxxx

Google APIs - All-or-nothing binary permissions

Scope	Meaning
`https://www.googleapis.com/auth/calendar`	read/write access to Calendars
`https://www.googleapis.com/auth/calendar.readonly`	read-only access to Calendars
`https://www.googleapis.com/auth/calendar.events`	read/write access to Events
`https://www.googleapis.com/auth/calendar.events.readonly`	read-only access to Events
`https://www.googleapis.com/auth/calendar.settings.readonly`	read-only access to Settings
`https://www.googleapis.com/auth/calendar.addons.execute`	run as a Calendar add-on

https://developers.google.com/calendar/api/guides/auth

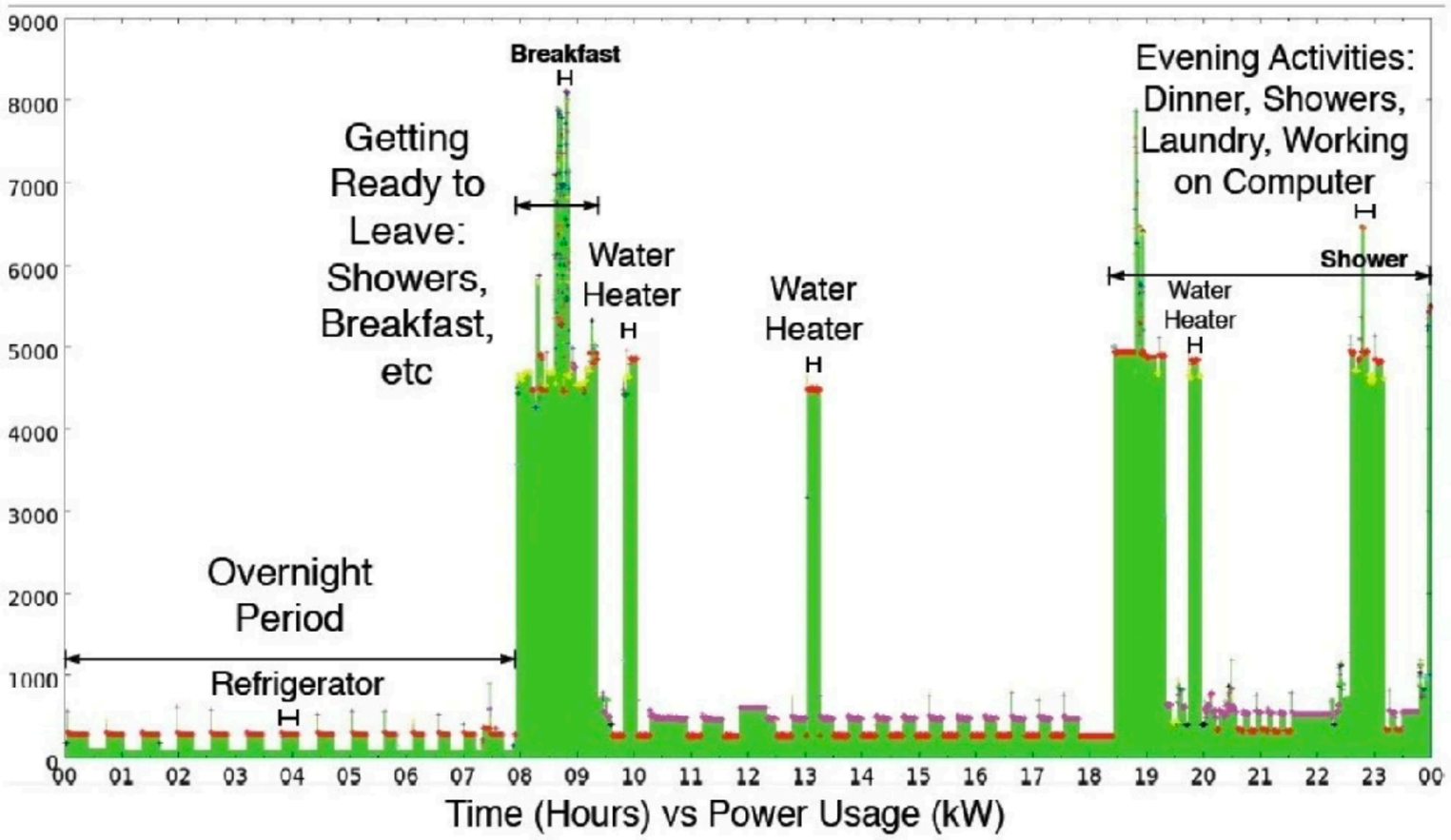
MPF in a nutshell

"All-or-nothing binary permissions" is insufficient. (1)

Sender	Recipient	Attribute	Transmission Principle
a sleep monitor	the local police	{subject}'s location	if {subject} has given consent
a security camera	government intelligence agencies	{subject}'s eating habits	if {subject} is notified
a door lock	{subject}'s doctor	the times {subject} is home	if the information is kept confidential
a thermostat	an Internet service provider	{subject}'s exercise routine	if the information is anonymous
a fitness tracker	its manufacturer	{subject}'s sleeping habits	if the information is used to perform maintenance on the device
a refrigerator	other devices in the home	audio of {subject}	if the information is used to provide a price discount
a power meter	{subject}'s immediate family	video of {subject}	if the information is used for advertising
a personal assistant (e.g. Amazon Echo)	{subject}'s social media accounts	{subject}'s heart rate	if the information is used to develop new features for the device
		the times it is used	if the information is not stored
			if the information is stored indefinitely
			if its privacy policy permits it
			in an emergency situation
			null (no transmission principle)

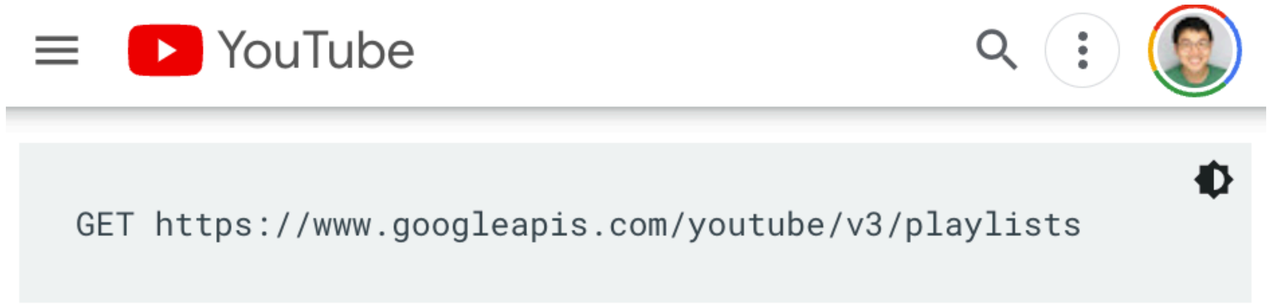
Discovering Smart Home Internet of Things Privacy Norms Using Contextual Integrity, IMWUT'18

"All-or-nothing binary permissions" is insufficient. (2)

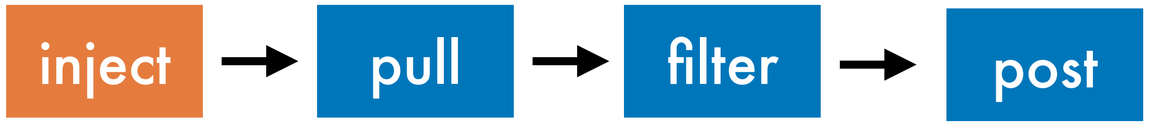


Program data transformation functions using chainable *operators*

URL-based APIs

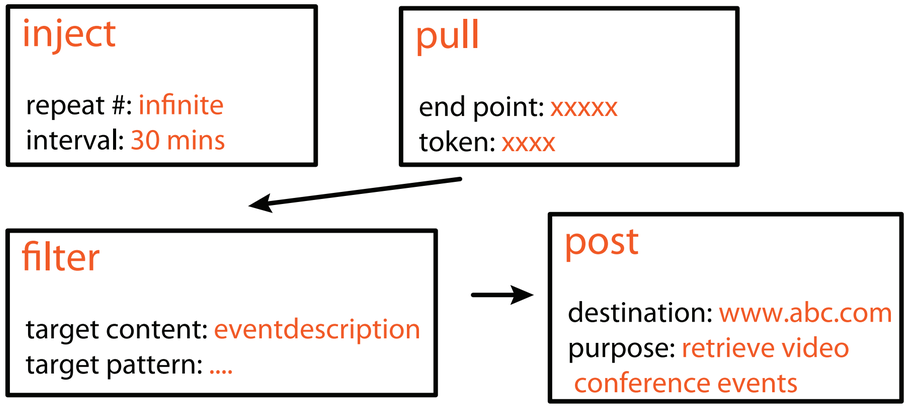


Operator-based APIs



MPF in a nutshell

A text-based whitelist *manifest* (i.e., program representation)




```
@purpose: The app can access calendar events
which contains a zoom link.
ZoomCalendarIntegration{
  // operator [properties]
  inject[...] -> pull Calendar[...] ->
  filter [Zoom join link pattern] ->
  post [Zoom events]
}
```

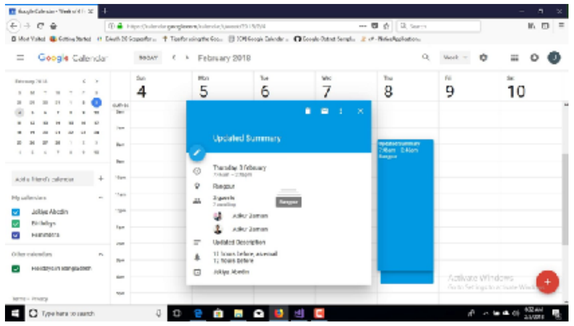
Broader application domains

Smart Home

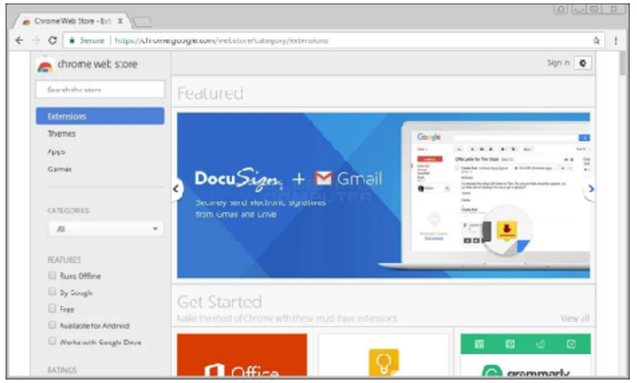
Smart City

Mobile apps

Social network



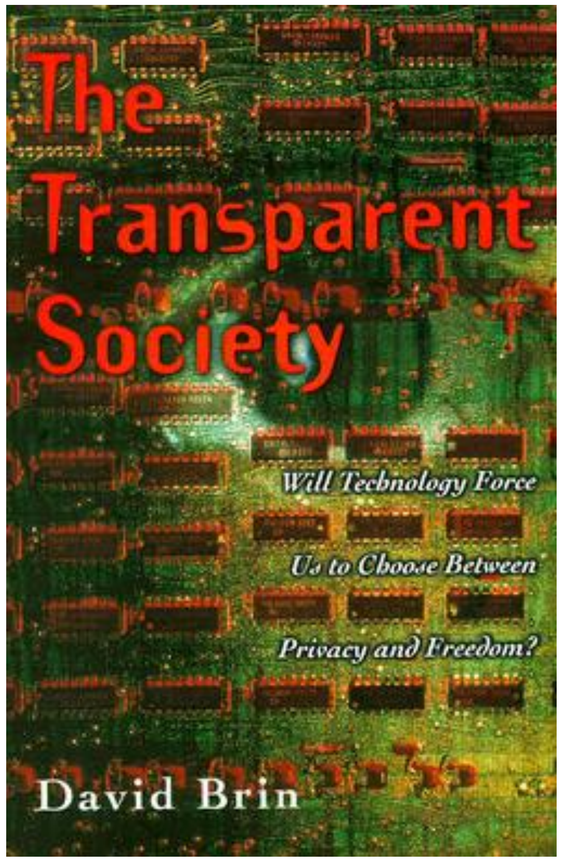
Personal data API



Browser extensions

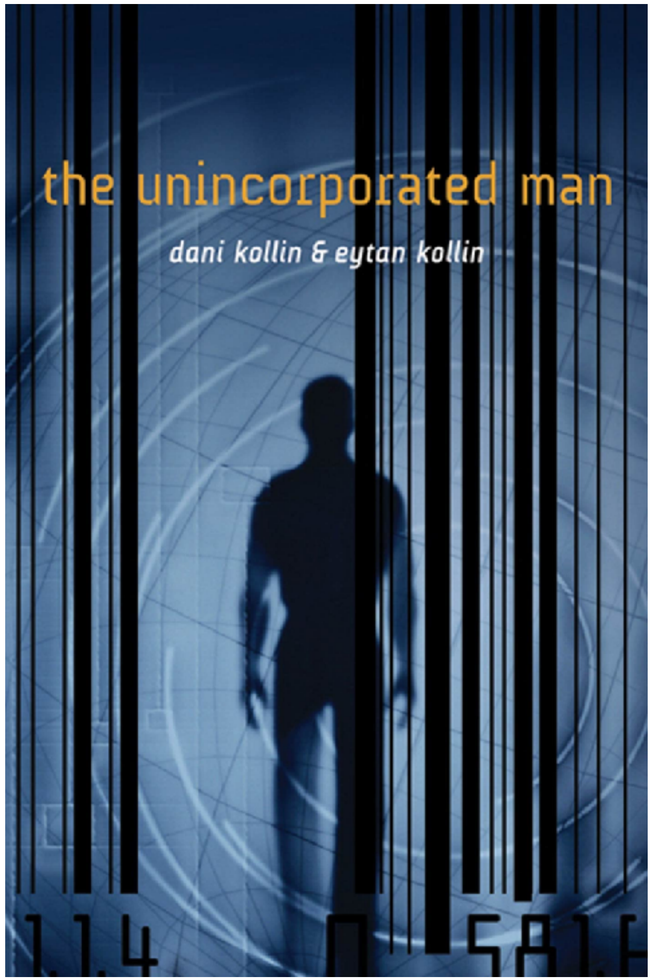
The future is private.

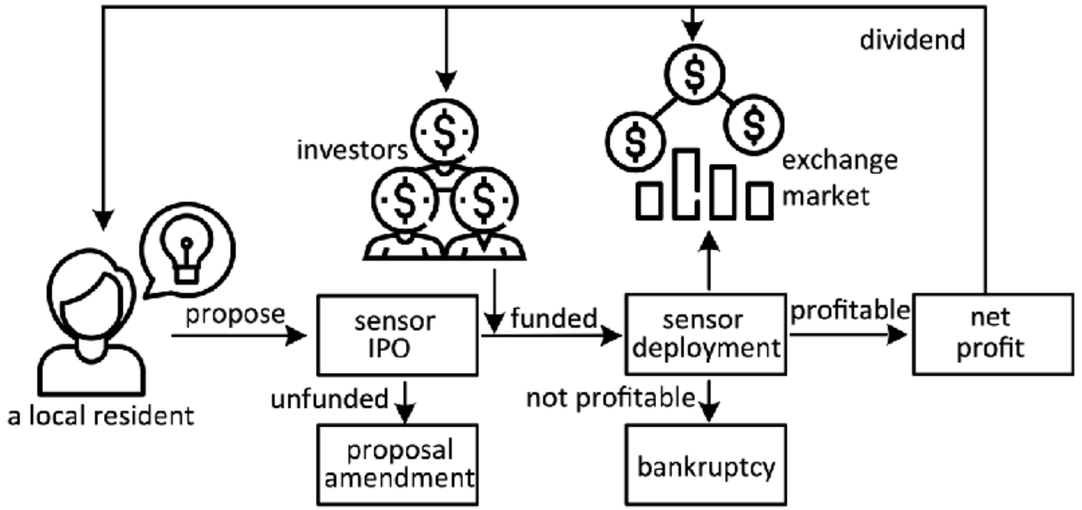
Sousveillance (inverse surveillance) - Transparency



- Auditable software/hardware

Purpose-based data economy

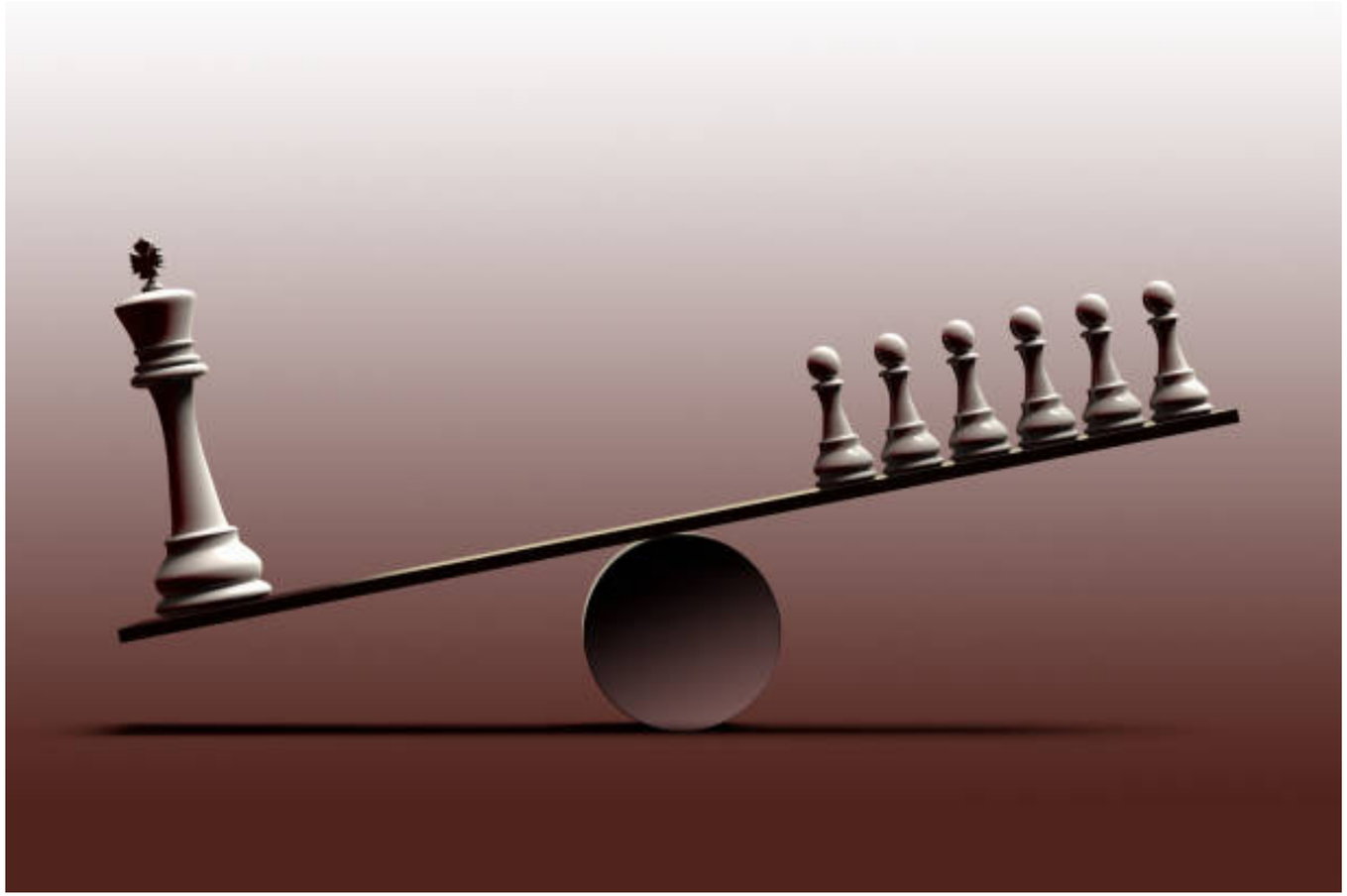




- New infrastructure to address data ownership/pricing

Jin et al., Sensor as a Company: On Self-Sustaining IoT Commons.

User-developer privacy negotiation

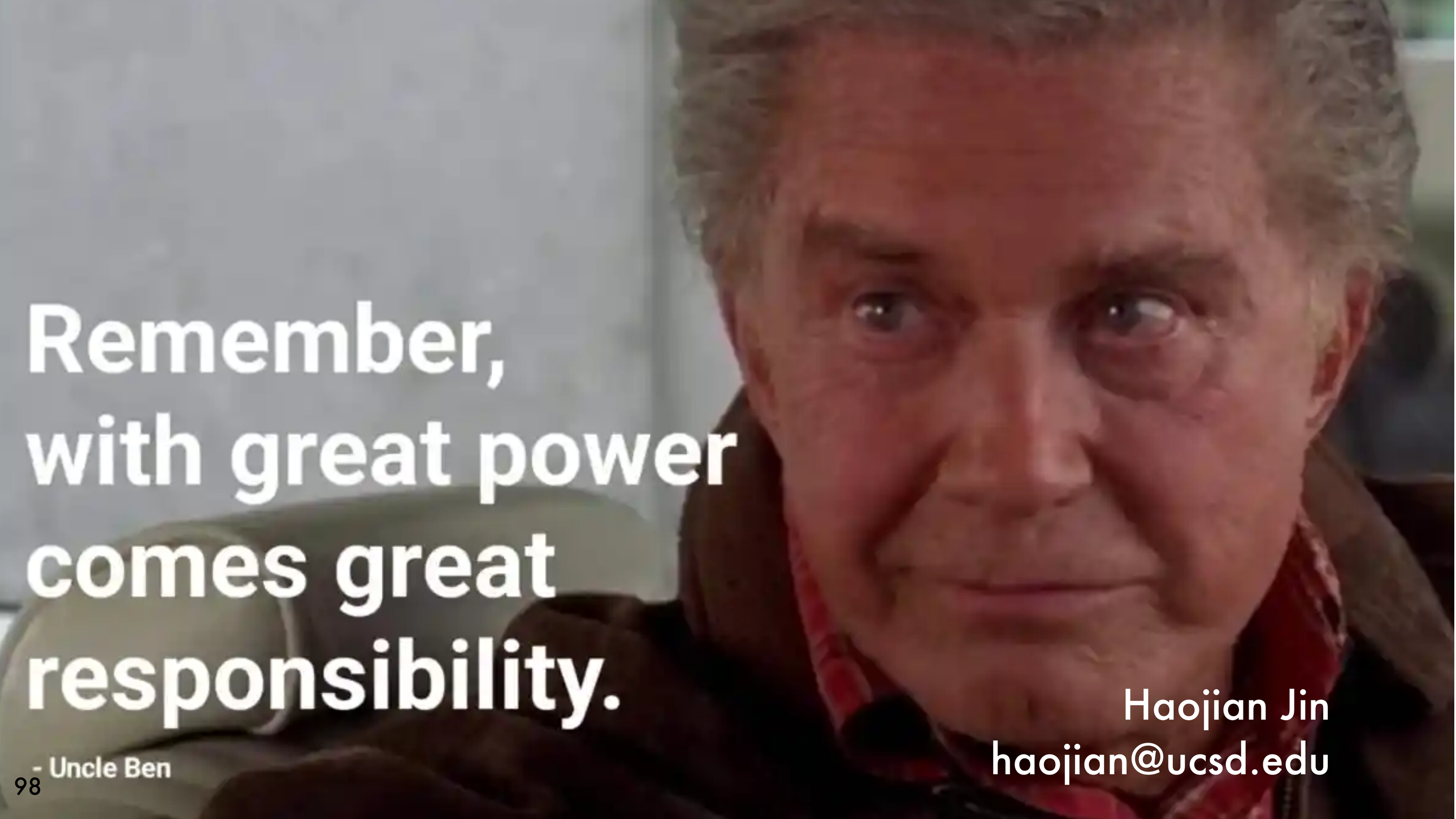


- *Empower users (e.g., Ad-blocker)*
- *Empower good developers*
- *Empower trusted third-parties*

Data Smith Lab

Safe, fair, cheap, accessible data economy

Casey Neistat's Guide to Filmmaking, https://www.youtube.com/watch?v=nLSUrTxquyE



Haojian Jin
haojian@ucsd.edu

Developers do not want to minimize data collection.

COMPLIANCE & ENFORCEMENT

A 14.5 Million Euro Fine for Failing to Get Rid of Old Files – Data Minimization Is Becoming a Stand-Alone Cybersecurity Obligation

by Avi Gesser, Matthew Kelly, Will Schildknecht, Dr. Vera Jungkind (Hengeler Mueller), and Dr. Carolin Raspé (Hengeler Mueller)

We have written several times here over the last few years about data minimization being an important part of an effective cybersecurity program. For most companies, the total amount of data that they control grows substantially each year, and more data

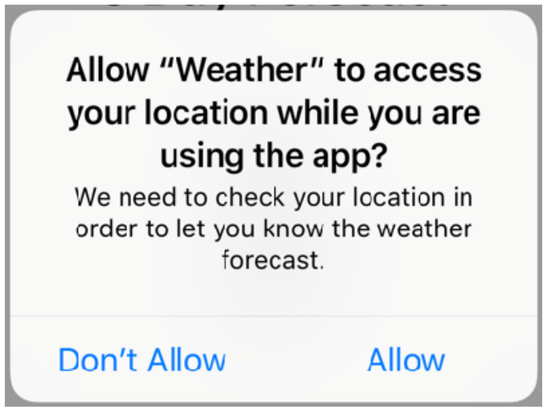
Strong policies on data minimization

New interaction paradigms and development supports

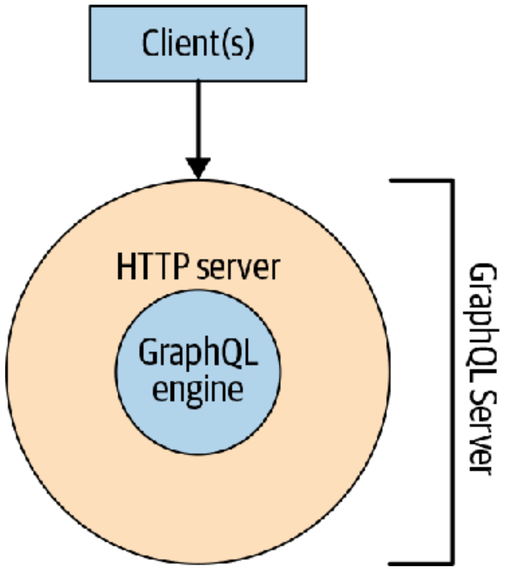
Adoption (2)

MPF is not the only way to implement data minimization.

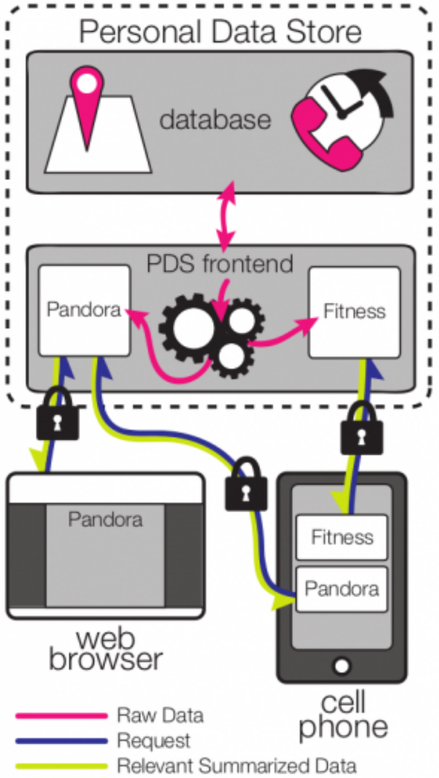
```
<manifest ...>
  <uses-permission android:name="android.permission.
       ACCESS_COARSE_LOCATION" />
</manifest>
```



Binary permissions



Database approach



Remote code execution