Supplementary Information for

OCTOPUS: Operation **C**ontrol System for **T**ask **O**ptimization and Job **P**arallelization via a **U**ser-Optimal **S**cheduler

Hyuk Jun Yoo,^{1,2} Kwan-Young Lee,^{2*} Donghun Kim,^{1*} and Sang Soo Han^{1*}

¹Computational Science Research Center, Korea Institute of Science and Technology, Seoul 02792, Republic of Korea

²Department of Chemical and Biological Engineering, Korea University, Seoul 02841, Republic of Korea

*Correspondence to: sangsoo@kist.re.kr (S.S.H.); donghun@kist.re.kr (D.K.); kylee@korea.ac.kr (K.-Y.L.).

Table of contents

1. Terminology definition in OCTOPUS

Supplementary Figure S1. Definition of words

1. open storage

3. place vial

2. Job submission via the interface node and job scheduler of the

master node

Supplementary Figure S2. Detailed job script structure for manual and automated

experimentations

The "metadata" key represents information about the experiments, including the subject, group name and log level. The "algorithm" key represents the process recommendation, including the model name, the total number of experiments and the model hyperparameters. The "process" key represents the experimental process information, including the module and task sequences and the fixed experimental conditions for each module.

Supplementary Figure S2a. Job script for manual experimentations

```
 "metadata" : 
 "subject":"Manual Experiment", 
 "group":"KIST_CSRC", 
 "logLevel":"DEBUG"
 }, 
 "model": 
        "modelName":"Manual", 
        "totalExperimentNum":2, 
        "inputParams":[ 
                "AddSolution=AgNO3_Concentration" : 0.0125, 
                "AddSolution=AgNO3_Volume" : 1000, 
                "AddSolution=AgNO3_Injectionrate" : 100
                "AddSolution=AgNO3_Concentration" : 0.0175, 
                "AddSolution=AgNO3_Volume" : 800, 
                "AddSolution=AgNO3_Injectionrate" : 300
 }, 
     "process": 
        "Synthesis":{ 
            "BatchSynthesis":{ 
                "Sequence":["AddSolution_Citrate","AddSolution_H2O2","AddSolution_NaBH4","Stir", 
                "Heat","Mix", "AddSolution_AgNO3", "React"], 
 "fixedParams": 
                    "AddSolution=H2O2_Concentration" : 0.375, 
                   "AddSolution=H2O2_Volume" : 1100, 
                   "AddSolution=H2O2_Injectionrate" : 100, 
                   "AddSolution=Citrate_Concentration" : 0.04, 
                   "AddSolution=Citrate_Volume" : 1100, 
                   "AddSolution=Citrate_Injectionrate" : 100, 
                   "AddSolution=NaBH4_Concentration" : 0.01, 
                   "AddSolution=NaBH4_Volume" : 3000, 
                   "AddSolution=NaBH4_Injectionrate" : 100, 
                    "Stir=StirRate":1000, 
                   "Heat=Temperature":25, 
                   "Mix=Time":300, 
                   "React=Time":3600
```


Supplementary Figure S2b. Job script for automated experimentations

```
 "metadata" : 
       "subject":"Automated Experiment", 
       "group":"KIST_CSRC", 
       "logLevel":"DEBUG"
 }, 
 "model": 
       "modelName":"BayesianOptimization", 
       "batchSize":6, 
       "totalCycleNum":3, 
       "verbose":0, 
       "randomState":0, 
       "sampling":{ 
 "samplingMethod":"latin", 
 "samplingNum":20
 "acq":{ 
 "acqMethod":"ucb", 
 "acqSampler":"greedy", 
           "acqHyperparameter":{ 
               "kappa":10.0
, and \} , and \} , and \} "loss":{ 
           "lossMethod":"lambdamaxFWHMintensityLoss", 
           "lossTarget":{ 
               "GetAbs":{ 
                  "Property":{ 
                      "lambdamax":573
 "Ratio":{ 
 "lambdamax":0.9, 
                     "FWHM":0.03, 
                     "intensity":0.07
 "prange":{ 
 "AddSolution=AgNO3_Concentration" : [25, 375, 25], 
 "AddSolution=AgNO3_Volume" : [100, 1200, 50], 
 "AddSolution=AgNO3_Injectionrate" : [50, 200, 50] 
, and \} , and \} , and \} "initParameterList":[],
```

```
 "constraints":[] 
 }, 
 "process": 
        "Synthesis":{ 
            "BatchSynthesis": 
                "Sequence":["AddSolution_Citrate","AddSolution_H2O2","AddSolution_NaBH4", 
                "Stir","Heat","Mix", "AddSolution_AgNO3", "React"], 
 "fixedParams": 
                   "AddSolution=AgNO3_Concentration" : 1250, 
                  "AddSolution=H2O2_Concentration" : 375, 
                  "AddSolution=H2O2_Volume" : 1200, 
                  "AddSolution=H2O2_Injectionrate" : 200, 
                  "AddSolution=Citrate_Concentration" : 20, 
                  "AddSolution=Citrate_Volume" : 1200, 
                  "AddSolution=Citrate_Injectionrate" : 200, 
                  "AddSolution=NaBH4_Concentration" : 10, 
                  "AddSolution=NaBH4_Volume" : 3000, 
                  "AddSolution=NaBH4_Injectionrate" : 200, 
                   "Stir=StirRate":1000, 
                  "Heat=Temperature":25, 
                  "Mix=Time":300, 
                  "React=Time":7200
 }, 
 "FlowSynthesis":{} 
 }, 
 "Preprocess":{ 
            "Washing":{}, 
            "Ink":{} 
 "Characterization":{ 
 "UV-Vis": 
                 "Sequence":["GetAbs"], 
                "fixedParams": 
                    "UV=Hyperparameter_WavelengthMin":300, 
                  "UV=Hyperparameter_WavelengthMax":849, 
                  "UV=Hyperparameter_BoxCarSize":10, 
                  "UV=Hyperparameter_Prominence":0.01, 
                  "UV=Hyperparameter_PeakWidth":20
 }, 
        "Evaluation":{ 
            "RDE":{}, 
            "Electrode":{}
```
Examples of job scripts for manual and automated experiments. The job script is based on the JSON (JavaScript Object Notation) format. Other job script formats were uploaded to the GitHub repository. (https://github.com/KIST-CSRC/Octopus)

Supplementary Figure S3. Login process with Auth0 in the interface node of OCTOPUS

1. login process with Auth0

2. connect to OCTOPUS

Supplementary Figure S4. Commands definitions for clients and administrators

Supplementary Figure S5. Workflow of job submission in job scheduler

Supplementary Figure S6. Examples of job management via a command line interface

Supplementary Figure S6a. Command line interface examples of master node and module node

initializations

(Module, BatchSynthesis) Module node on \rightarrow heartbeat & update device information

When the master node is executed, it sends a heartbeat to the activated module nodes to check their connectivity. If all module nodes are successfully connected, the resource manager retrieves the device information from each module node. In the two examples mentioned above, this information includes the experimental device information associated with each module node.

Supplementary Figure S6b. Command line interface examples of module node updates

Jenu: DireartDeat, Status
2023-10-05 19:30:35,687 - UV-Vis::DEBUG -- [USB2000plus (heartbeat)]: Hello World!! Succeed to connection to main computer!
2023-10-05 19:30:35,687 - UV-Vis::DEBUG -- [Module Node (UV-Vis)]: {'Spe iend : b'heartbeat,status'

master node must be updated due to the allocation of device action via the new settings. The client can use the 'updateNode' command in the prompt interface to refresh the latest device information of the module node.

Supplementary Figure S6c. Command line interface examples of login process, job submission and job status

The client tries to log in to the master node via their ID and password. If the client ID or password does not match, the master node notifies the client of a login failure. After the login process is successful, the client can submit a job script using the 'qsub' command. The master then converts to an activated job and conducts the experiment based on the information in the job script. The client can monitor the status of their job using the 'qstat' command. The job status table includes the client name, job submission time, filename of job script, number of current experiments, number of total experiments, job status and mode type of the job.

Supplementary Figure S6d. Command line interface examples of job submission, hold, restart and deletion

(Client) $qsub \rightarrow qhold \rightarrow qrestart$

(Module) $qsub \rightarrow qhold \rightarrow qrestart$

(Client) $qsub \rightarrow qhold \rightarrow qdel$

(Module) $qsub \rightarrow qhold \rightarrow qdel$

|
|BatchSynthesis] : packet information list:
|Module Node (UV-Vis)] : qhold:jobID 0 is
|BatchSynthesis] : packet information list: 09-24 22:36:50,168
09-24 22:36:50,168 BatchSynthesis::INFO [BatchSynthesis] BatchSynthesis::DEBUG -09-24 22:36:54,042 - BatchSynthesis::INFO -- [BatchSynthesis] : pac ,
ation list:['0'. 'DS_B', 'storage_empty_to_stirrer', '0,1', 'virtual'] holded jobID : 0 joblD : 0
input_qhold_jobID_list : [0]
-24 22:37:02,296 - BatchSynthesis::INFO -- [BatchSynthesis] : packet information list:['qdel', '0']

The client can temporarily pause the submitted job or execute the job using the 'qhold' command. When the client enters 'qhold', the master node sends the job ID to the module node due to a job pause. Then, the module node stores the job ID and holds all the device actions of that job ID. If the client enters the 'qrestart' command, the paused job resumes. However, if the client enters the 'qdel' command, the paused job will be deleted.

3. Job executions in the master node

Supplementary Figure S7. Functions of the task generator

Supplementary Figure S7a. Detailed workflow of update device settings in resource manager

Supplementary Figure S7b. Task reflection depending on latest device information

Supplementary Figure S7c. Detailed workflow of the task generator

Supplementary Figure S7d. Examples of task templates

```
# Robot
self.MoveContainer template={
     "Task":"MoveContainer", 
     "Data":{ 
              "From":"", 
             "To":"", 
             "Container":"", 
             "Device":{} 
} 
self.PrepareContainer template={
     "Task":"PrepareContainer", 
     "Data":{ 
              "From":"", 
             "To":"", 
             "Container":"", 
             "Device":{} 
} 
self.AddSolution_template={ 
     "Task":"AddSolution", 
     "Data":{ 
         "Solution":"", 
         "Volume":{ 
              "Value":0,"Dimension":"μL"
             }, 
         "Concentration":{ 
              "Value":0,"Dimension":"mM"
         "Injectionrate":{ 
             "Value":0,"Dimension":"μL/s"
             }, 
         "Device":{} 
} 
self.Stir_template={ 
     "Task": "Stir", 
     "Data": { 
         "StirRate": { 
            "Value": 0,
             "Dimension": "rpm"
         "Device":{} 
} 
self.Heat_template={
```

```
 "Task": "Heat", 
     "Data": { 
         "Temperature": { 
              "Value": 0, 
              "Dimension": "ºC"
         }, 
         "Device":{} 
} 
self.Mix_template={ 
     "Task": "Mix", 
     "Data": { 
         "Time": { 
              "Value": 0, 
              "Dimension": "sec"
         }, 
         "Device":{} 
} 
self.React_template={ 
     "Task": "React", 
     "Data": { 
         "Time": { 
             "Value": 0,
              "Dimension": "sec"
         }, 
         "Device":{} 
}
```

```
self.GetAbs template={
    "Task":"GetAbs", 
    "Data":{ 
        "Device":{}, 
        "Hyperparameter":{ 
            "WavelengthMin":{ 
                    "Description":"WavelengthMin=300 (int): slice wavlength section 
depending on wavelength_min and wavelength max",
               "Value": 0,
                "Dimension": "nm"
            }, 
            "WavelengthMax":{ 
                    "Description":"WavelengthMax=849 (int): slice wavlength section 
depending on wavelength_min and wavelength_max",
               "Value": 0,
                "Dimension": "nm"
 },
```


The origin of task template has empty value. The examples of template uploaded in Github repository. (https://github.com/KIST-CSRC/Octopus)

Supplementary Figure S7e. Conversion of job script in terms of task sequences in the task

generator

A task generator allows clients to create process recipes based on their desired experimental conditions and process sequences. To execute various process sequences for each job script, we would present the generated recipe in the JSON data format shown in the GitHub repository.

Supplementary Figure S8. Examples of resource allocation in batch synthesis module

Supplementary Figure S9. The role of action translator for abstraction and digitalization

Supplementary Figure S10. Functions of the action executor

Supplementary Figure S10a. Predefined device commands of the action executor

Supplementary Figure S10b. Detailed workflow of the action executor

Action Translator

Action Executor

Supplementary Figure S11. Hierarchy structure of the generated material data

The second auxiliary function is to store the results of the experiments in individual JSON files, and is conducted during the job cycle. (Figure S12a, Supplementary Information) The reason for choosing JSON is its flexible, computer-readable structure. In the MAP for chemical experiments, there are various modules, each containing diverse tasks, device types and device settings. Given this variability, storing the material data in a structured format within a relational database would be challenging. In other words, tabular data with a relational database to store the task information in MAPs, inevitably face quite a sparse table due to the tremendous diversity of the attributes and methods from the various device settings and module components.

Therefore, we implement a nonstructured data format with inclusivity for the data structure of MAP. The most significant feature of JSON is its hierarchical structure. A well-defined JSON format enhances the flexibility of storing attributes/methods of diverse tasks and aids in data readability, making it easier for clients to understand the processes. Furthermore, the JSON format is a commonly used data specification on the web, so it helps to easily convert CLI to a web-based interface. We designate the standard material data structure with four main hierarchical keys—metadata, algorithms, processes and property/performances—as the highest-level categories. To enable the utilization of the accumulated data for AI-driven experimental planning in the future, we implement MongoDB, which stores and manages the JSON data.

4. Network-protocol-based modularization

Supplementary Figure S12. Process modularization for homogeneity via device

Г

Module Node

server

Supplementary Figure S12a. An actual example of heterogeneous environments

Module Node = UV-Vis

We configured the module node using Python based on the Windows OS. Communication between the environment of the module node and other devices is designed to set up a device server, enabling interaction with the module node. The table represents the configuration of our module node, as presented in our recent research.^[3]

Supplementary Figure S12b. Virtual workflow of the network protocol with a hierarchical structure

This workflow depends on device settings of modules.

Supplementary Figure S13. Process modularization for scalability via internal and external network

Supplementary Figure S13a. Utilization of the internal network protocol in the routing table

The third number represents the type of experimental process, and the fourth number represents the module information. The types of experimental processes include synthesis, preprocessing, evaluation, characterization and database, which are commonly defined in chemical experiments. These experimental processes were represented by numbers from 1 to 5, making it easy for researchers to identify which process they belong to via the third number alone. Starting from the fourth number, different modules are included in the same process type. For example, within the same preprocessing category, various modules might be grouped, such as processes for washing, ball milling, ink preparation, spray coating and sonication. In the fourth part of the internal network address, the presence of '1' represents the gateway. Consequently, modules should be sequentially recorded in the routing table starting from '11'.

In network protocols, broadcast refers to a method of sending messages or data packets across a network to multiple computers or network devices simultaneously via a single transmission. This means that the information is sent to all the devices on the network without specifying a specific target computer or device. Broadcast is typically used for network management, debugging, or when the same message needs to be sent to multiple devices. We set the internal network address of the emergency stop to "192.168.255.255". The term "255" enables broadcasting via an internal network protocol.

Supplementary Figure S13b. Virtual examples of scalable autonomous experiment platform based on internal and external network

Supplementary Figure S14. Process modularization and utilization for safety.

Supplementary Figure S14a. Broadcast-enabled module node shutdown

In network protocols, broadcast refers to a method of sending messages or data packets over network to multiple computers or network devices simultaneously via a single transmission. This means that the information is sent to all devices on the network without specifying a specific target computer or device. Broadcast is typically used for network management, debugging, or when the same message needs to be sent to multiple devices. We set internal network address of emergency stop to "192.168.255.255". "255" enables broadcast in internal network protocol.

Supplementary Figure S14b. Workflow of the safety alert system

Supplementary Figure S14c. Real messages of the alert system showing the experiment progress

Auto Lab (Main) BOT

######### [Automatic_synthesis_UV-HJ] Experiment 1 is running #########

오후 4:40

Auto Lab (Main) BOT

######### [Automatic_synthesis_UV-HJ] Experiment 1 is done #########

오후 4:50

This message is the result of communication through Dooray Messenger.[4]

Supplementary Figure S14d. Real messages of the alert system for device disconnection via heartbeat

 $D!$

Auto Lab (Main) BOT

[BatchSynthesis] LinearActuator is disconnected, please check status of connection

Auto Lab (Main) BOI

[UV-Vis] Pipetting Machine (uArm Swift Pro) is disconnected, please check status of connection

This message is the result of communication through Dooray Messenger.[4]

Supplementary Figure S14e. Real messages of the alert system via the restock function for chemical vessels

This message is the result of communication through Dooray Messenger.[4]

5. Job parallelization to address the module overlap challenge

Supplementary Figure S15. Schematic algorithm of job parallelization

1. Check the resources of the experimental equipment set on the modules.

2. Set an index *i* (the index *i* represents the position index of the stacked job ID in the waiting queue). 3. The name of the module that is to be executed is retrieved first according to the predefined module execution order from the job, with the job ID corresponding to the *i*-th position within the waiting queue. This is denoted as a *"module"* in the diagram.

- 4. Does the *"module"* own a device standby task?
	- 4-1. If yes, proceed to (5).
	- 4-2. If not, proceed to (8).
- 5. Is the *"module"* executing a device standby task?
	- 5-1. If yes, proceed to (8).
	- 5-2. If not, proceed to (6).
- 6. The index *i* is incremented by 1.
- 7. Determine if index *i* is the last.

7-1. If *i* is the last order, return to (2) \rightarrow Reset *i* to check the job with the job ID corresponding to the first order in the waiting queue.

7-2. If *i* is not the last order, proceed to (3) \rightarrow Search for the first module to execute in the job

with the job ID corresponding to the next order in the waiting queue.

8. Pop the job ID corresponding to index *i* from the waiting queue.

9. End the job trigger.

Supplementary Figure S16. The timeline of the modules used for catalyst development included device standby time

Supplementary Figure S17. Multiple jobs for catalyst application

Supplementary Figure S18. Definition of performance metrics

Supplementary Table S1. Performance metrics of job parallelization

Supplementary Table 1a. Job waiting time between serialization and parallelization

Supplementary Table 1b. Job turnaround time between serialization and parallelization

Job ID	Serialization (h)	Parallelization (h)

Supplementary Table 1c. Job total time between serialization and parallelization

6. Task optimization for preventing device overlaps.

Supplementary Figure S19. A bird's eye view image of modules including "BatchSynthesis" and "UV-Vis"

This image is a module used in a previous study $[3]$.

Supplementary Figure S20. Computing results of Boolean operation in python

Supplementary Figure S21. Real examples of masking table

The "BatchSynthesis" module has four resources which include the number of vials that can be processed in the magnetic stirrer. The "UV-Vis" module has also four resources which represents the number of vial holders storing vials for "UV-Vis" spectrum measurements. The image of hardware is a module used in a previous study $^{[3]}$.

7. The closed-packing schedule for optimizing module resource

Supplementary Figure S22. Definition of resource in realistic platform

The batch synthesis module has four resources which include the number of vials that can be processed in the magnetic stirrer. The UV-Vis module has also four resources which represents the number of vial holders storing vials for UV-Vis spectrum measurements. The image of hardware is a module used in a previous study^[3].

Supplementary Figure S23. Detailed workflow of closed-packing schedule in multi jobs

8. Performance test of user-optimal schedulers

Supplementary Figure S24. Schematic design of conventional scheduling algorithm (FCFS)

First Come First Serve (FCFS)

Times

Supplementary Figure S25. Job information for benchmark test of user-optimal schedulers

The bold text corresponds to "job waiting time" and the underline text corresponds to "job turnaround time".

Supplementary Figure S26. Residual resources-based job split via closed-packing

schedule in user-optimal schedulers

Supplementary Table S2. Performance test in realistic platform

Supplementary Table 2a. Result of job waiting time in realistic platform

Supplementary Table 2b. Result of job turnaround time in realistic platform

Job ID	FCFS (h)	User-Optimal Schedulers (h)
$\overline{0}$	9.75	9.75
	11.04	2.67
$\overline{2}$	11.34	0.3
3	11.72	0.68
$\overline{4}$	12.17	0.83
5	13.81	2.36
6	14.41	0.67
$\overline{7}$	14.86	0.45
8	15.92	1.89
9	15.16	0.3
10	16.84	1.39

Supplementary Table 2c. Result of job total time in realistic platform

Supplementary Figure S27. Results of closed packing schedule in UV-Vis module

(A) Job split by residual resource-based CPS.

(B) Job parallelization in characterization module without device standby times.

Supplementary Figure S28. Results of task optimization between batch synthesis and UV-Vis module

Supplementary Figure S29. Cause analysis for the delay time of CPS in batch synthesis module

9. Copilot of OCTOPUS

Supplementary Figure S30. Reusability comparison with and without Copilot of OCTOPUS

 $O =$ Reusable \triangle = Partial modification required
X = Complete modification required

Supplementary Figure S31. Example of action generation for each device in module

Supplementary Figure S31a. GPT prompt example of action generation for each device in module

- \checkmark name: module name
- \checkmark device str: all devices in module

Supplementary Figure S31b. Example of generated actions for each device via GPT recommendation and client feedback

Supplementary Figure S32. Example of task generation for module

Supplementary Figure S32a. GPT prompt example of task generation for module

- \checkmark input module name: module name
- \checkmark task str: need to include task name
- \checkmark description: module description

Supplementary Figure S32b. Example for generated tasks via GPT recommendation and client

Supplementary Figure S33. Example of the action sequence generation for task

execution

Supplementary Figure S33a. GPT prompt example of action sequence generation for task execution

 \checkmark input task list = generated tasks of module

 \checkmark input device action dict = generated actions of devices

Supplementary Figure S33b. Example of generated action sequences for task execution via GPT recommendation and client feedback

Supplementary Figure S34. Example of task template and type validation generation

Supplementary Figure S34a. Prompt engineering of task template generation for type validation using the OpenAI API

Role assignment

Supplementary Figure S34b. Prompt engineering of Pydantic class generation for type validation using the OpenAI API

Role assignment

Supplementary Figure S34c. Example of generated task template and type validation for each task

Supplementary Table S4. The result of automated code generation/customization via Copilot of OCTOPUS

* {}: the real name of module, ex) {module name}="SolidStateModule"

Purple boxes represent the core software of OCTOPUS. Green boxes indicate the automated code generation for module operation via Copilot of OCTOPUS, which includes functions for module operation. Red boxes represent the JSON file addressing new module information via Copilot of OCTOPUS.

Supplementary References

- [1] L. Parziale, W. Liu, C. Matthews, N. Rosselot, C. Davis, J. Forrester, D. T. Britt, others, *TCP/IP tutorial and technical overview*, IBM Redbooks, **2006**.
- [2] A. C. Vaucher, F. Zipoli, J. Geluykens, V. H. Nair, P. Schwaller, T. Laino, *Nat. Commun.* **2020**, *11*.
- [3] H. J. Yoo, N. Kim, H. Lee, D. Kim, L. T. C. Ow, H. Nam, C. Kim, S. Y. Lee, K.-Y. Lee, D. Kim, S. S. Han, Bespoke Metal Nanoparticle Synthesis at Room Temperature and Discovery of Chemical Knowledge on Nanoparticle Growth via Autonomous Experimentations. *Adv. Funct. Mater.* **2024**, 2312561. https://doi.org/10.1002/adfm.202312561
- [4] NHN Dooray! Corporation, *https://dooray.com/main/service/messenger*.