

Automated Machine Learning-based diagnosis of impaired consciousness: cross-center and protocol generalization of EEG biomarkers.

Federico Raimondo

fraimondo@uliege.be

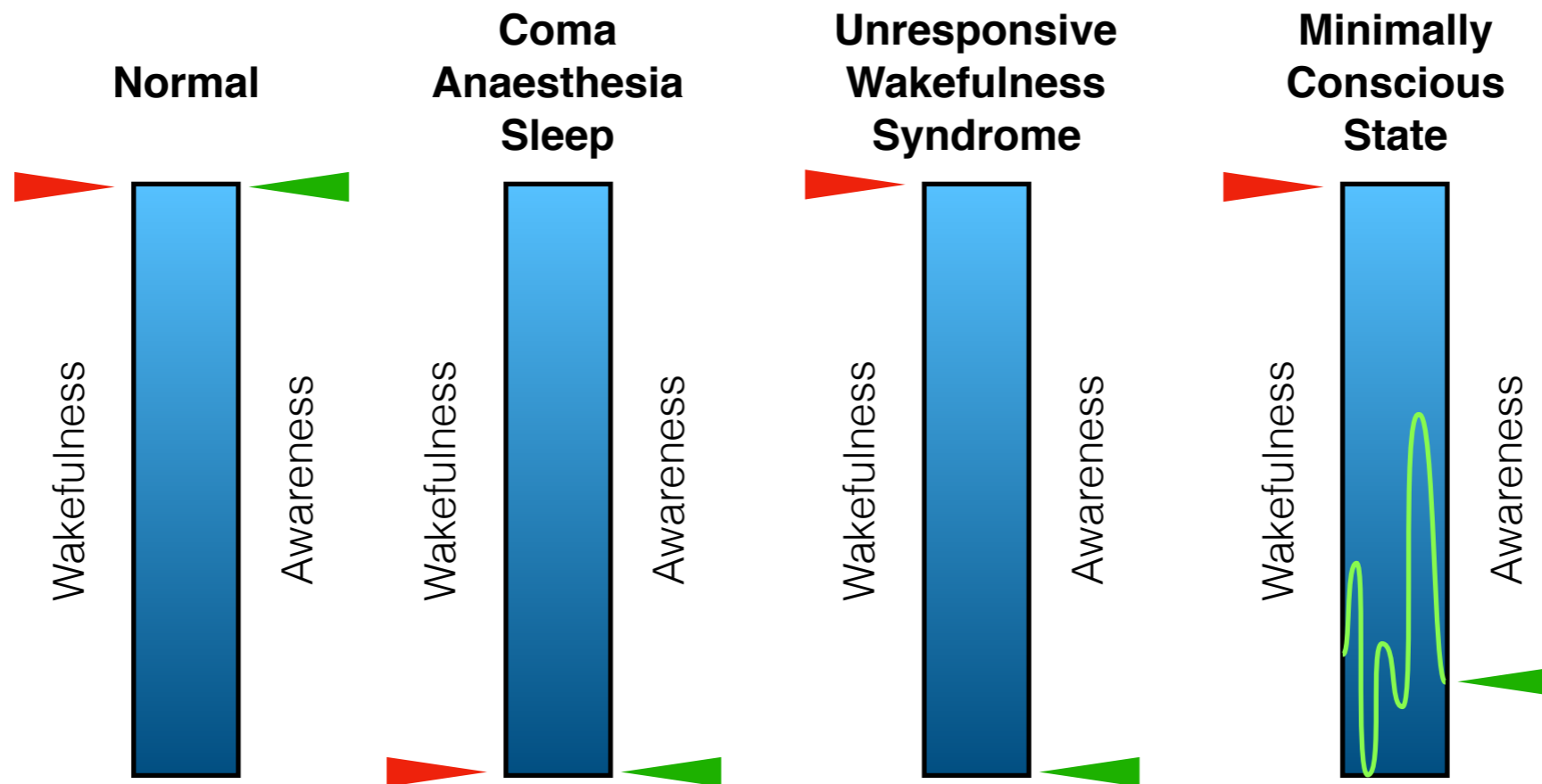


@fraimondo



@RaimondoFede

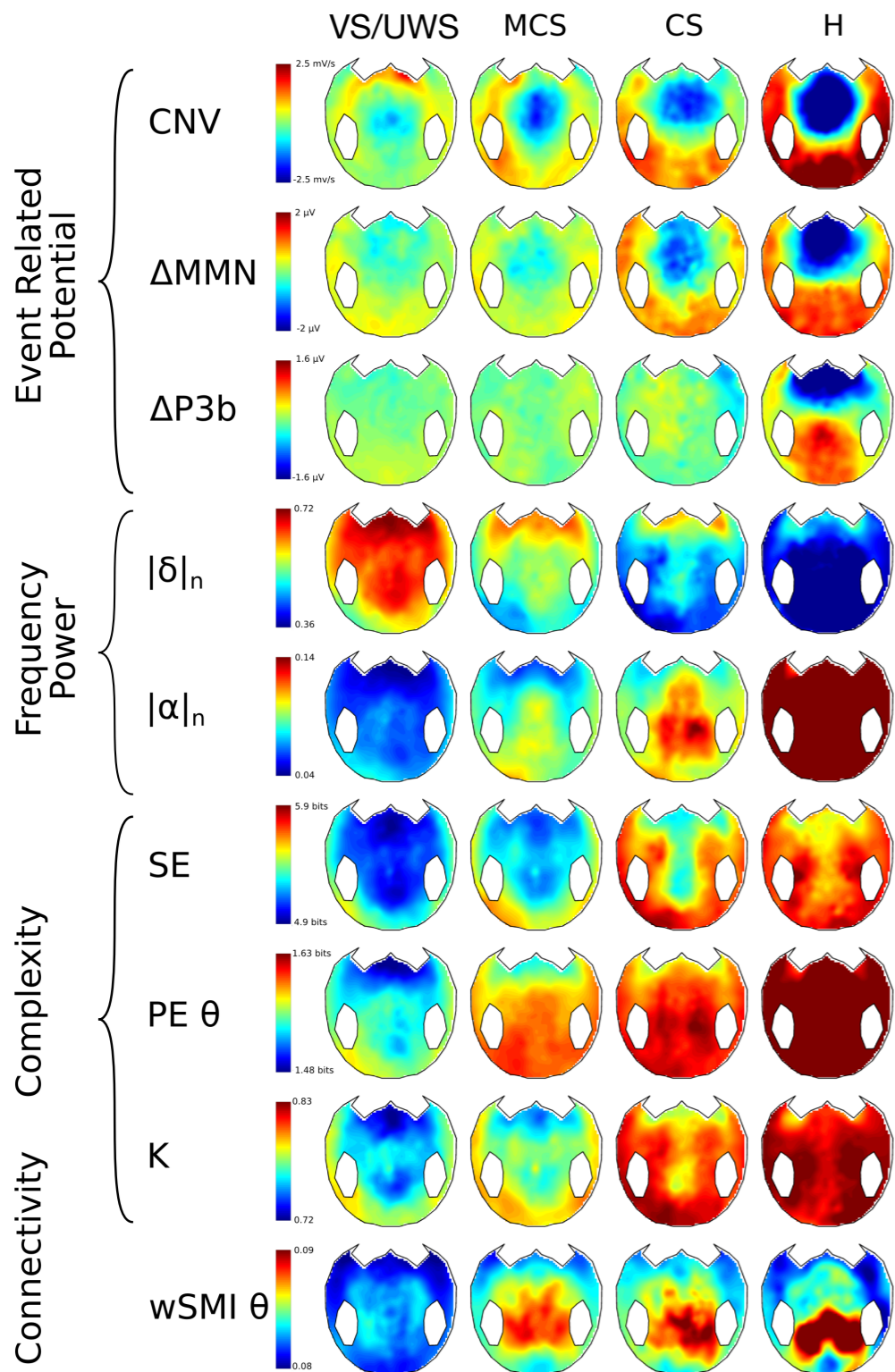
Disorders of Consciousness



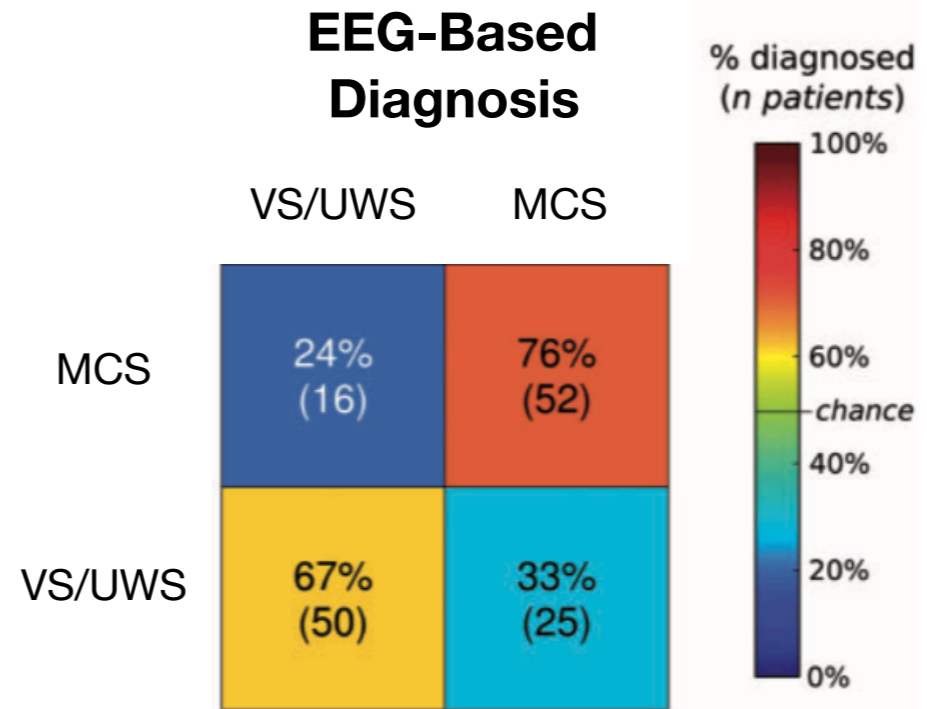
EEG-based diagnosis of DOC



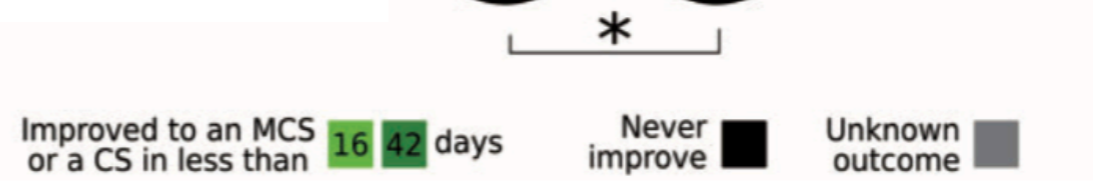
Sitt*, King* et al.
Brain 2014



**Behavioural
Diagnosis**



Outcome



Does this tool generalise?



Aim: assess the possibility of creating a universal tool for EEG-based diagnosis of DOC

Methods:

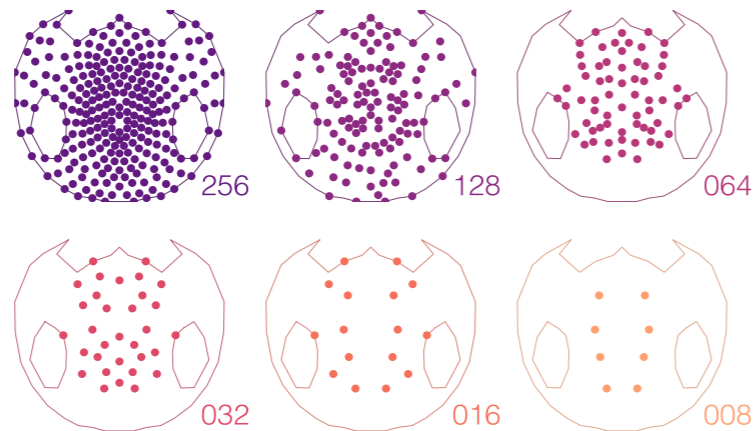
- Classifiers: SVM and Extra-Trees.
- Compare against a dummy classifier.
- Compare data spatially and temporally subsampled.
- Test on Resting State data (Liège).
- Bootstrap to estimate CI (2000 random samples).
- Measure AUC-ROC.

Dataset	Paris 1	Paris 2	Liège
N (EEG)	142	107	78
N (Patients)	98	92	78
N (UWS)	75	52	22
N (MCS)	66	56	57
Sex Ratio	2.0	1.9	1.2
Age (M [SD]) in years	46.5 [17.8]	45.4 [17.7]	38.0 [14.3]
Delay (M [SD]) in days	125.9 [372.9]	299.5 [823.6]	1040.5 [1227.6]
Delay (MD) in days	30.0	40.5	529
Delay (min to max)	6 to 2611	8 to 6570	11 to 5380
Anoxia (%)	29.6	30.4	N/A
Stroke (%)	29.6	15.2	N/A
TBI (%)	23.5	28.2	48.1
Other (%)	18.4	29.4	N/A

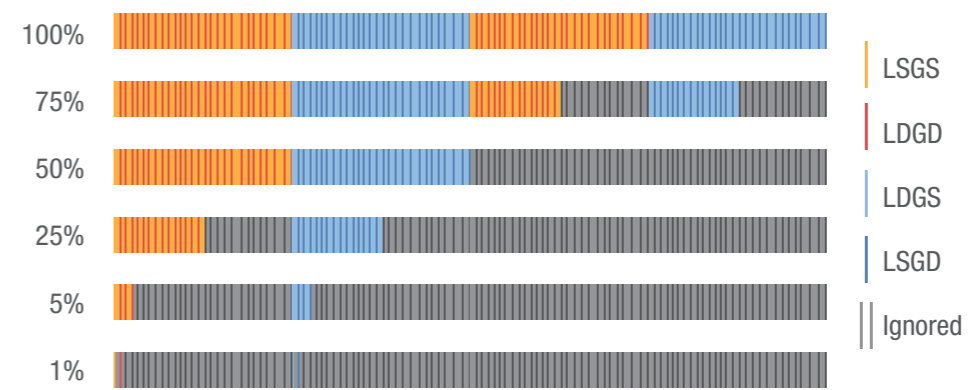
Generalisation: Alternative EEG configurations



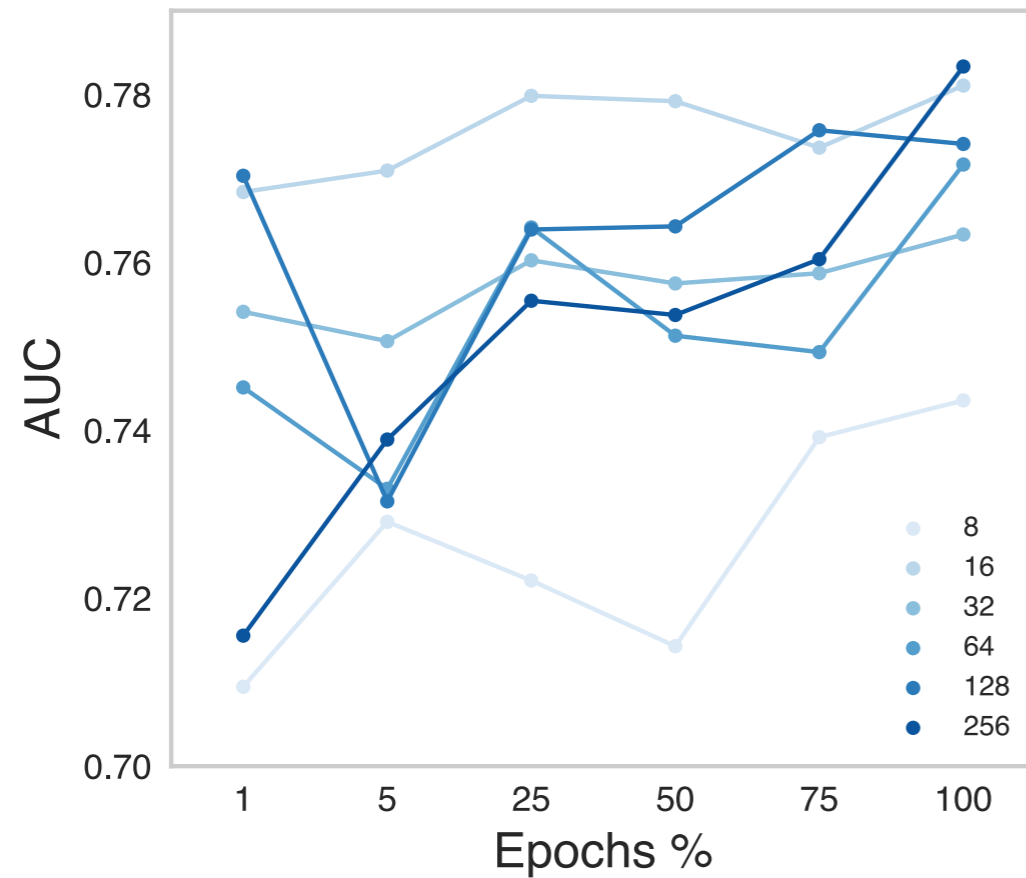
SPATIAL SUBSAMPLING



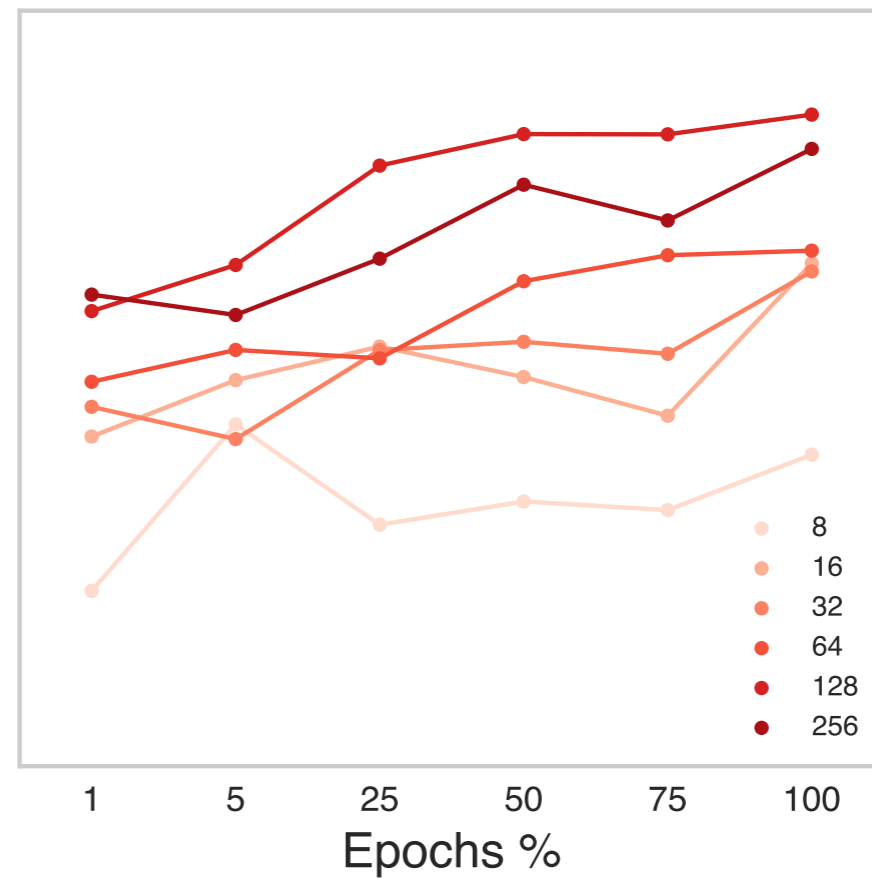
TEMPORAL SUBSAMPLING



SVM



ET

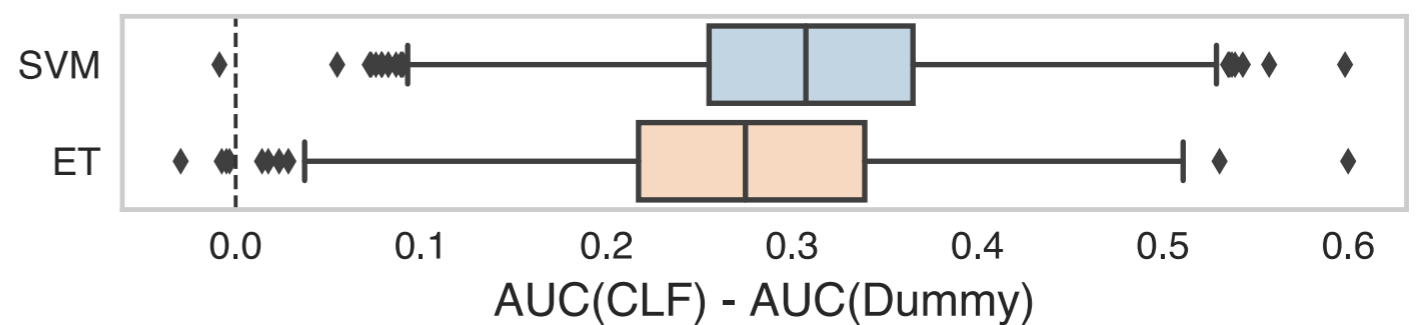
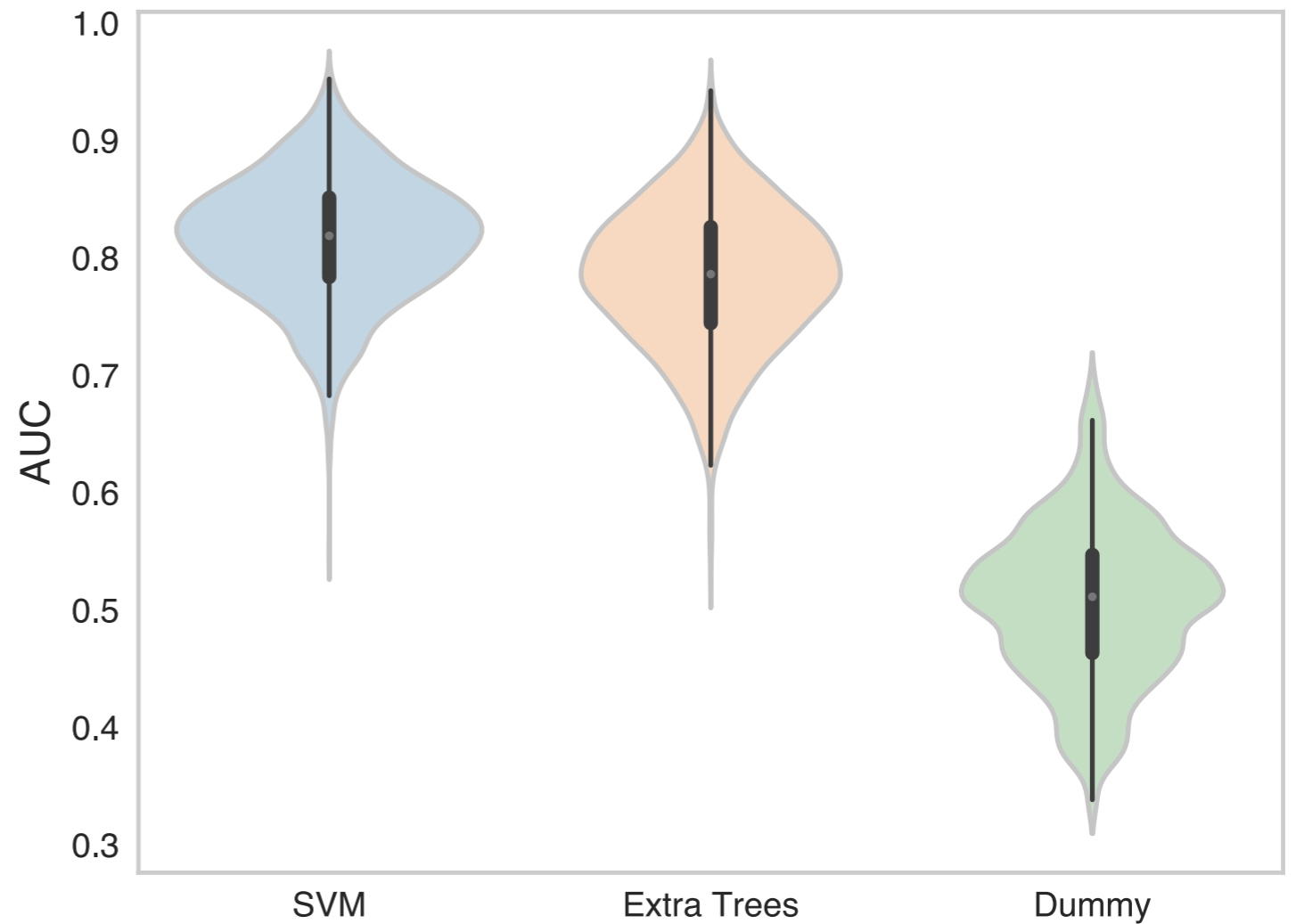
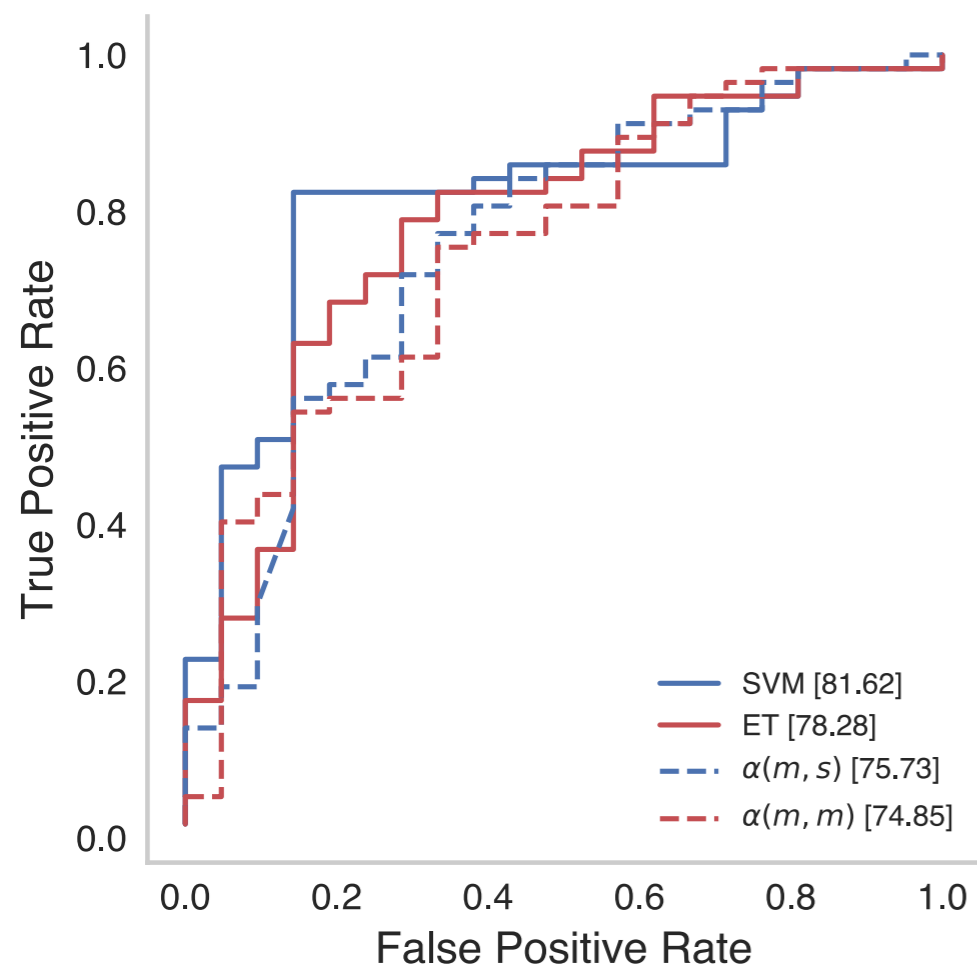


Generalisation: Resting State



Results:

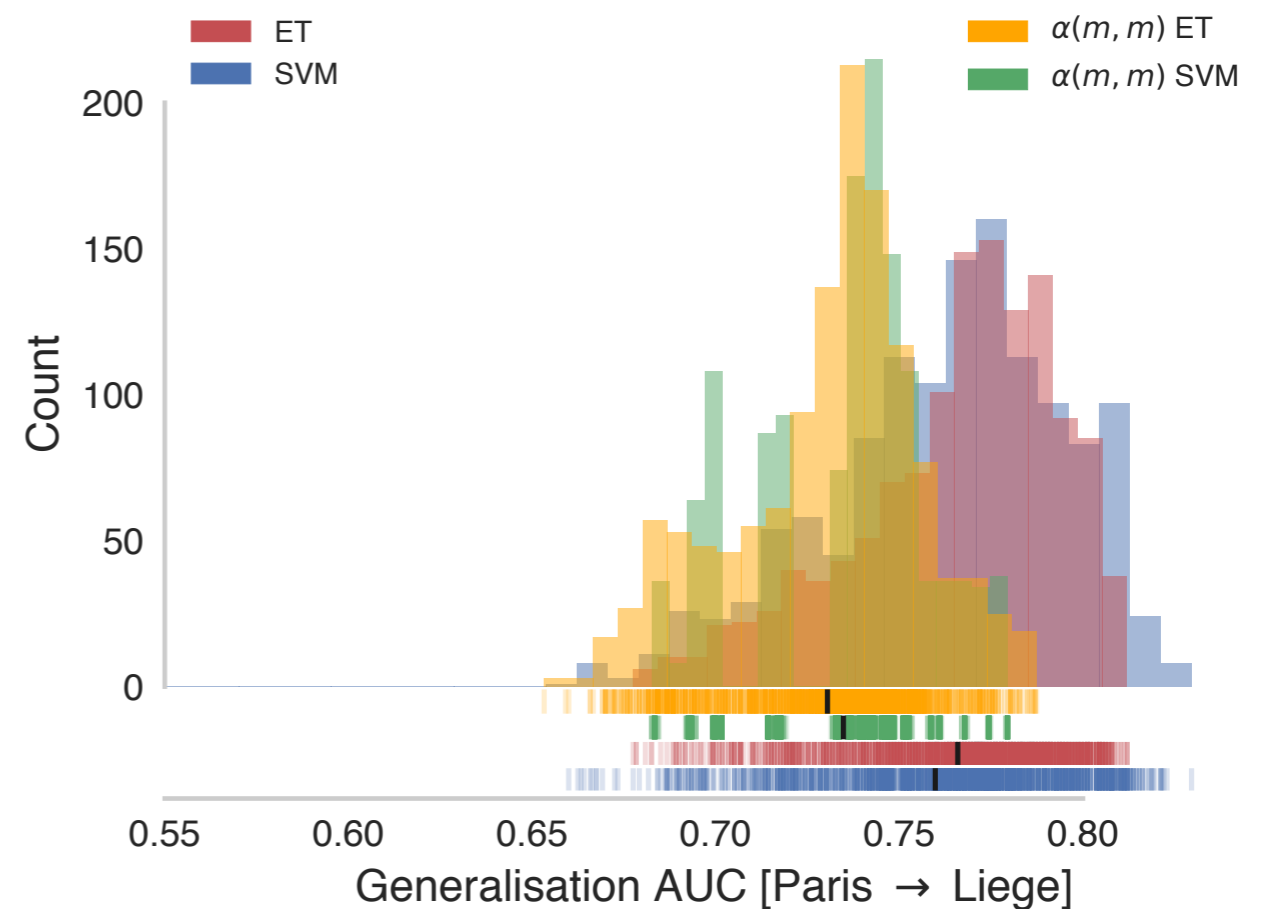
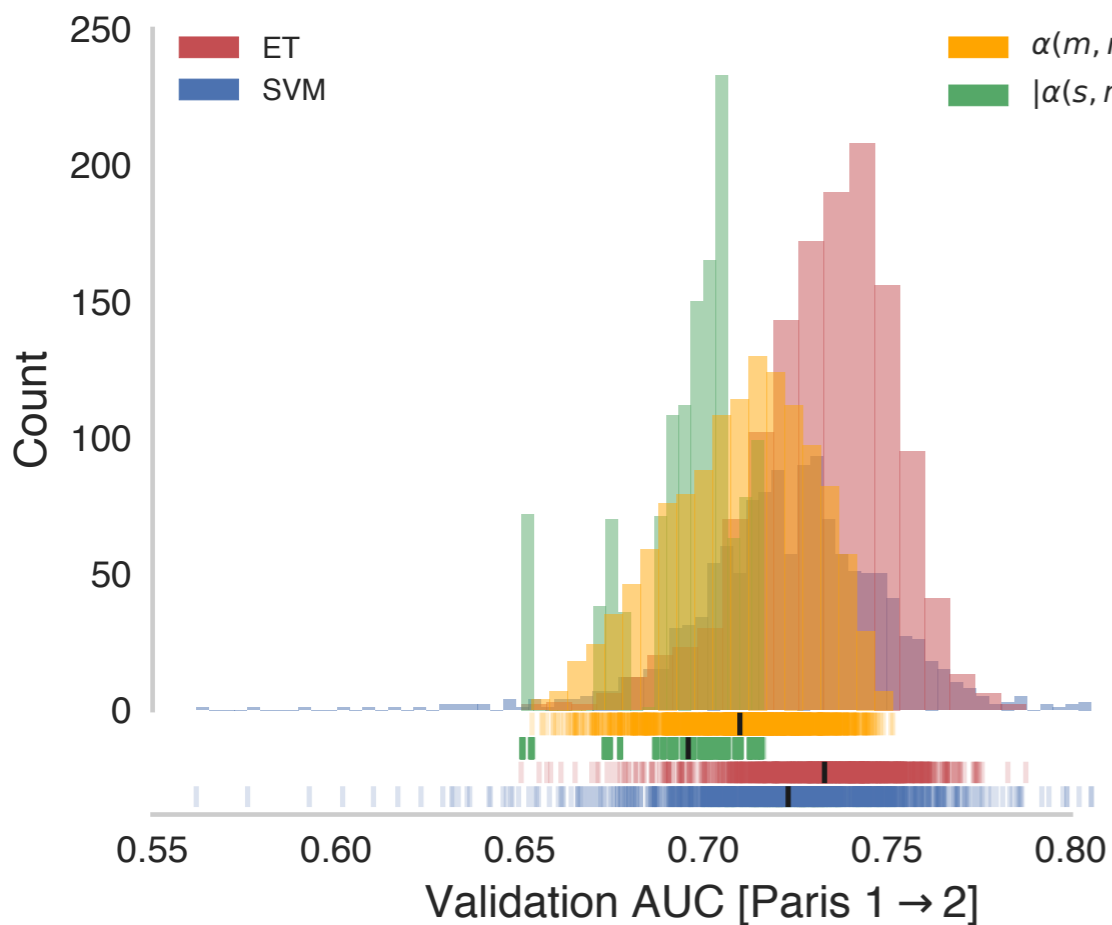
- SVM: 81.62%
- Extra-Trees: 78.28%



Generalisation: Mixed EEG configurations and acquisition protocols



All the different spatial and temporal configurations considered for training and testing



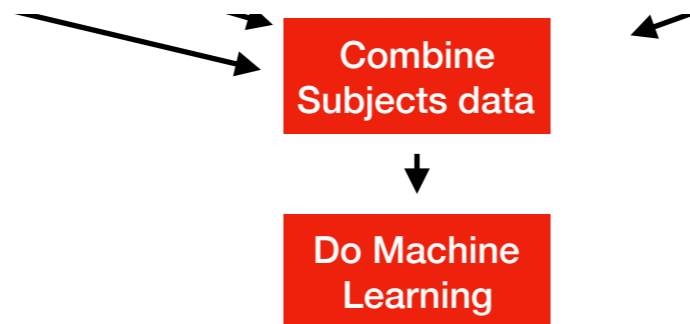
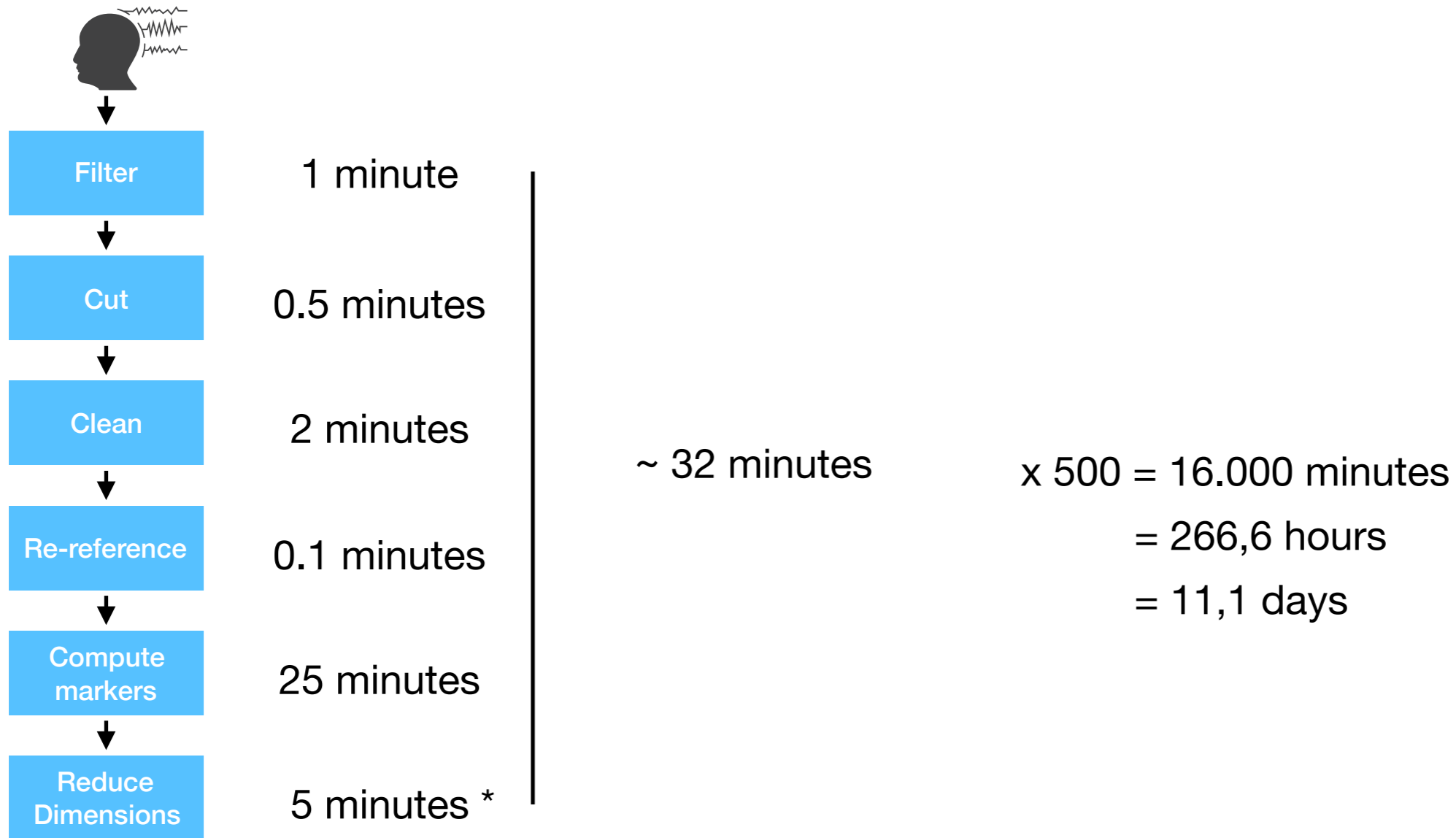


Conclusions

- Generalisation can be achieved despite changes in spatiotemporal configuration of the EEG and recording protocol.
- Multivariate overperforms univariate models.
 - Particularly with differences between training and testing sets.
- Choosing between ET and SVM classifier depends on the domain.
 - ET: model to Fit them All.
 - SVM: maximises AUC when data variability is low.
- Evoked potentials markers, while de-emphasized by the classifier, still contributes positively.
 - Validation Set: 72.91% with vs 71% without.
- While brain function during task and resting state differ, we could observe significant generalisation from task to resting state EEG.



Why is this impressive?



Processing

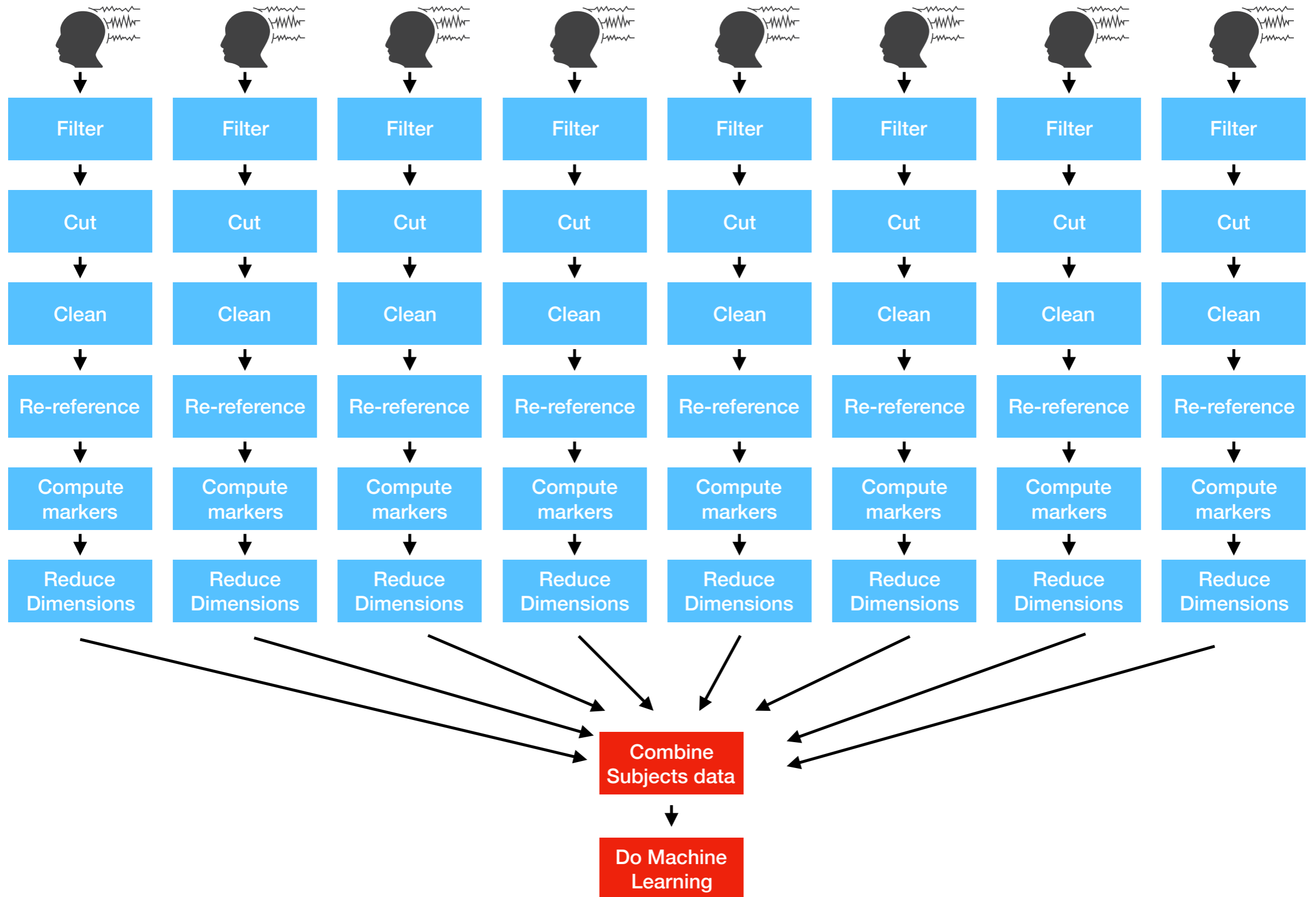


```
1 eegs % Array with EEG file names
2 results = zeros(500)
3
4 for idx=1:500
5     ... this_eeg = read_eeg(eegs(idx));
6     ... filtered_eeg = filter_eeg(this_eeg);
7     ... cut_eeg = epoch(filtered_eeg);
8     ... clean_eeg = clean(cut_eeg);
9     ... referenced_eeg = rereference(clean_eeg);
10    ... markers = compute_markers(referenced_eeg);
11    ... value = reduce_dimension(markers)
12    ... results(idx) = value
13 end
14
15 do_machine_learning(results)
```

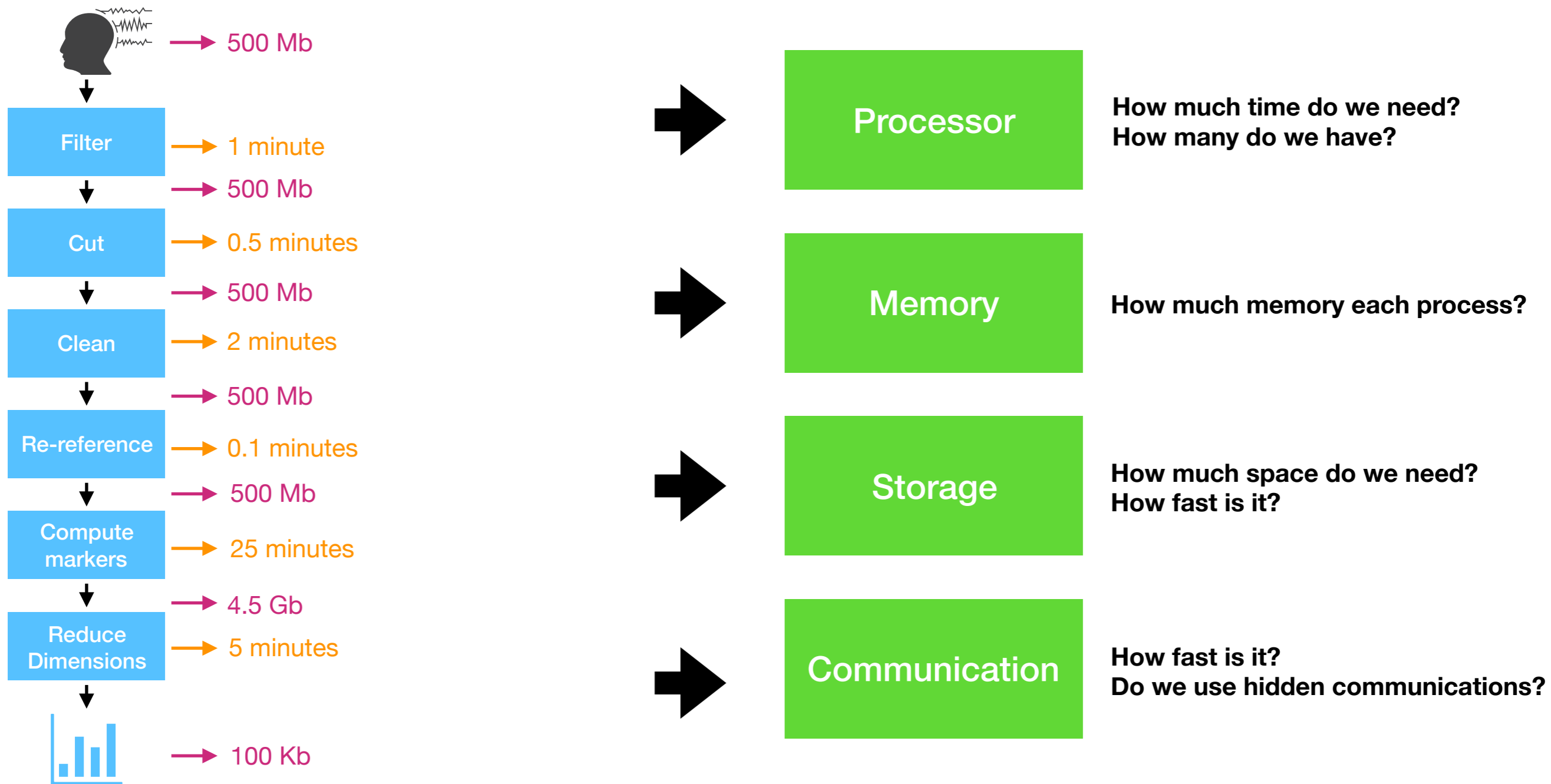
- Possible problems:
 - Defect → Expected Results \neq Actual Results
 - Bug → When we know about the defect!
 - Error/Fail → Manifestation of the defect
- Solution: Debug and correct
- Implies: > 11 days again!



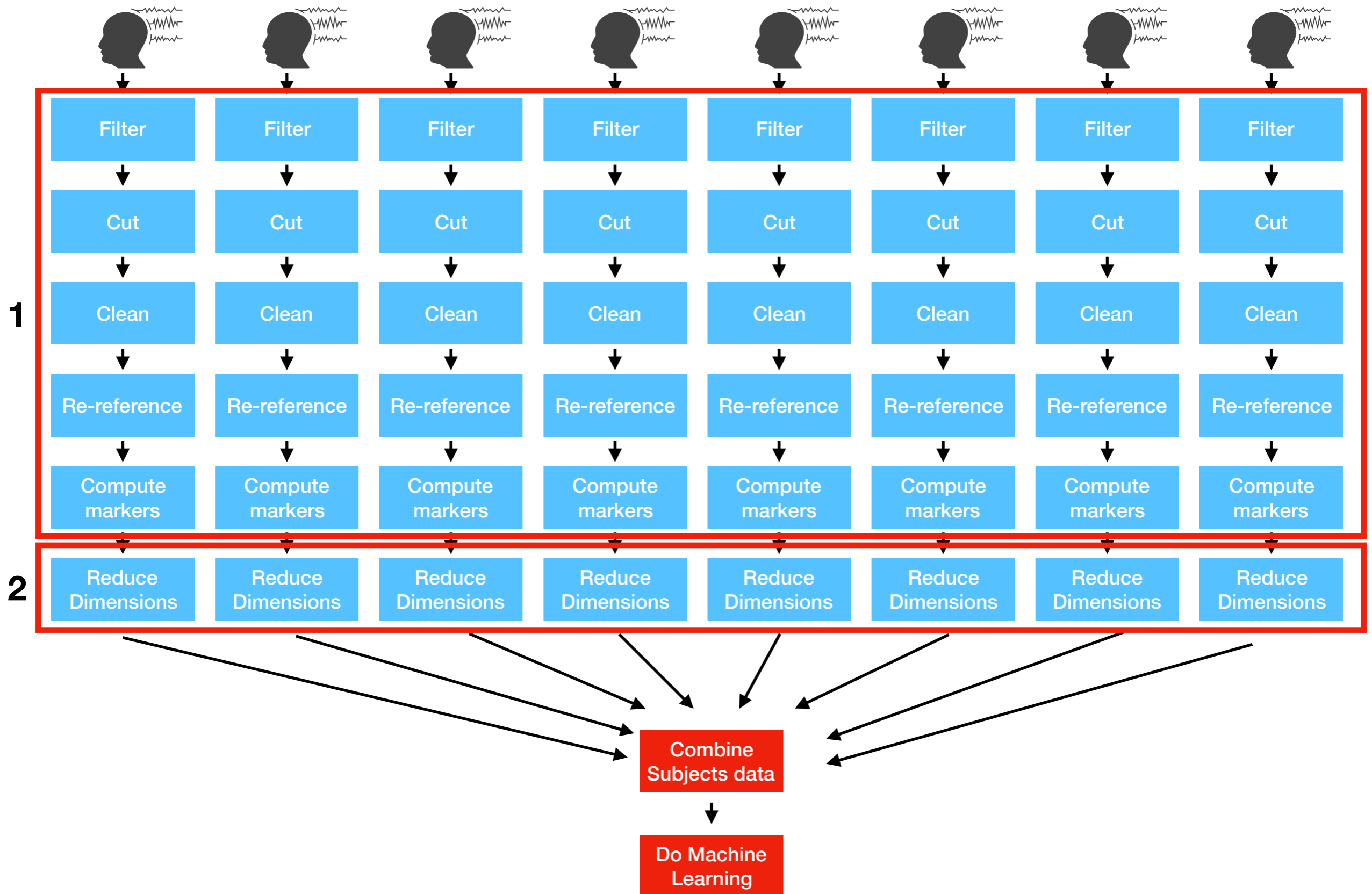
Parallel processing



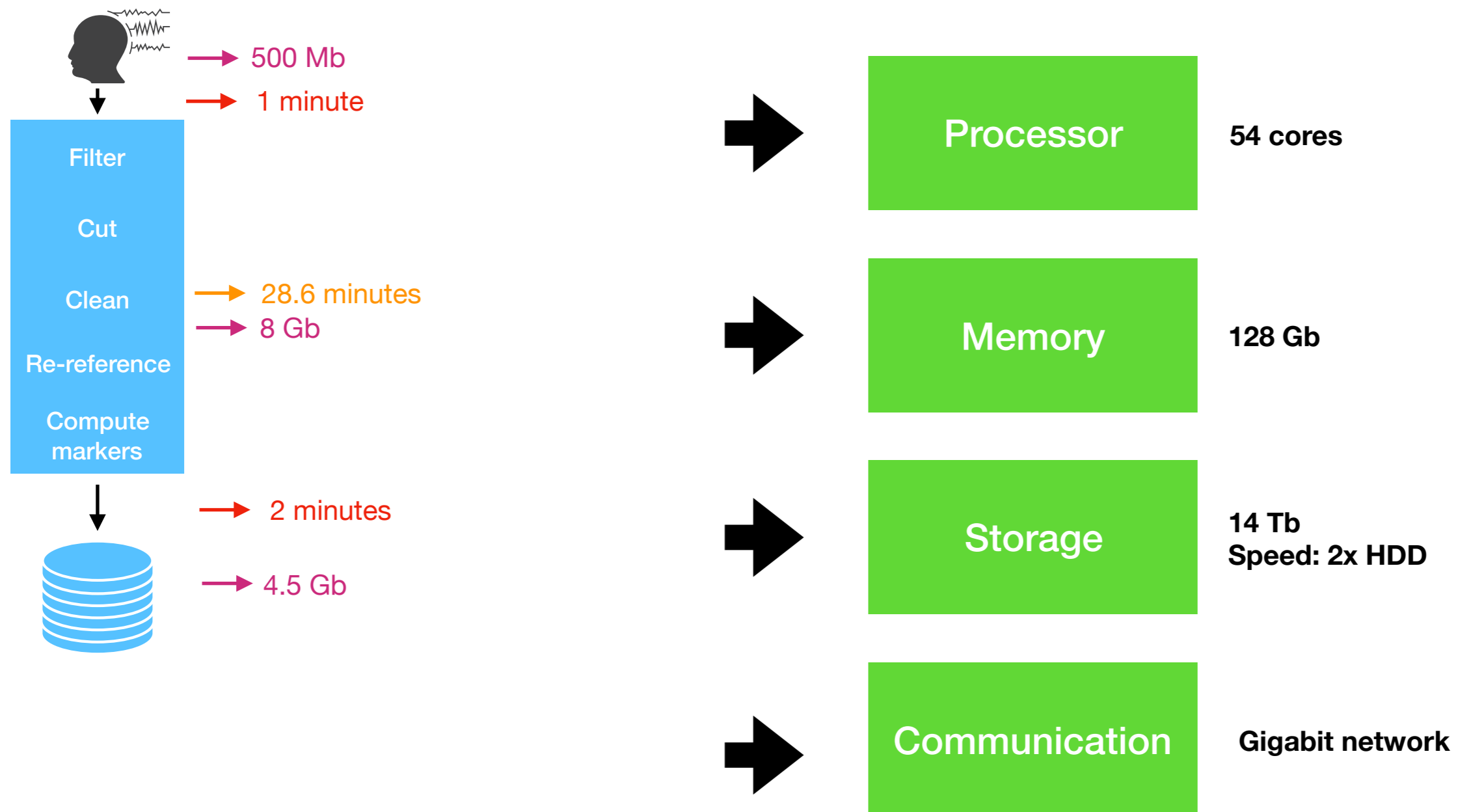
Parallel processing: considerations



Parallel processing



Parallel processing: considerations



How much can we parallelise this step?

Parallel processing: considerations



	Available	Required	Overflow?	# Jobs
Processor	54 cores	8 cores for a few seconds 1 most of the time	Some overhead time (not much)	6.75 - 54
Memory	128 Gb	8	Storage is used	16
Storage	14 Tb Speed: 2x HDD	4.5 Gb	Crash (and good luck!)	
Communication	Gigabit network	None	None	

10 Jobs → ~27 Hours

15 Jobs → ~18 Hours

Parallel processing: running



1. Create a script that given a subject/sample number, runs the whole analysis
2. Use GNU parallel to run the script

```
parallel -j10 --joblog job-analysis.log --resume --resume-failed --tag --delay 30 echo {} ::: `seq 1 10`
```

1 2 3 4 5 6 7 8

1. Number of Jobs to run in parallel
2. Keep a log of what happens
3. Resume where we left (if stopped)
4. Re-run the failed ones
5. Print a tag with the parameters before the output
6. Waiting time between processes
7. Command
8. Values to replace {} in the command

Parallel processing: running (example)



1. Create a script that given a subject/sample number, runs the whole analysis

```
run_analysis.m > ...
1 function run_analysis(sample_id)
2     disp(['Running sample ', num2str(sample_id)])
3 end
4
```

2. Use GNU parallel to run the script (within a bash script)

```
parallel_analysis.sh
1 #!/bin/bash
2
3 njobs=4
4 samples=`seq 1 20`
5 command='/Applications/MATLAB_R2017b.app/bin/matlab -nodisplay -nosplash -nodesktop -r "run_analysis({}); exit" '
6
7 parallel -j$njobs --joblog job-analysis.log --resume --resume-failed --tag --delay 0 $command ::: $samples
8
```

- Alternative declaration of samples

```
samples=(Name1 Name4 Sample23)
```

Parallel processing: running (example)



```
8
8
8      < M A T L A B (R) >
8      Copyright 1984-2017 The MathWorks, Inc.
8      R2017b (9.3.0.713579) 64-bit (maci64)
8      September 14, 2017
8
8      Warning: Function functions has the same name as a MATLAB builtin. We suggest
8      you rename the function to avoid a potential name conflict.
8
8      To get started, type one of these: helpwin, helpdesk, or demo.
8      For product information, visit www.mathworks.com.
8
8      Warning: Function functions has the same name as a MATLAB builtin. We suggest
8      you rename the function to avoid a potential name conflict.
8      > In path (line 109)
8         In addpath (line 94)
8         In startup (line 5)
8      Warning: Function functions has the same name as a MATLAB builtin. We suggest
8      you rename the function to avoid a potential name conflict.
8      > In path (line 109)
8         In startup (line 14)
8      Running sample 8
6
6
6      < M A T L A B (R) >
6      Copyright 1984-2017 The MathWorks, Inc.
6      R2017b (9.3.0.713579) 64-bit (maci64)
6      September 14, 2017
6
6      Warning: Function functions has the same name as a MATLAB builtin. We suggest
6      you rename the function to avoid a potential name conflict.
6
6      To get started, type one of these: helpwin, helpdesk, or demo.
6      For product information, visit www.mathworks.com.
6
6      Warning: Function functions has the same name as a MATLAB builtin. We suggest
6      you rename the function to avoid a potential name conflict.
6      > In path (line 109)
6         In addpath (line 94)
6         In startup (line 5)
6      Warning: Function functions has the same name as a MATLAB builtin. We suggest
6      you rename the function to avoid a potential name conflict.
6      > In path (line 109)
6         In startup (line 14)
```

Parallel processing: running (example)



```
mia → vcclass git:(master) ✖ cat job-analysis.log
```

Seq	Host	Starttime	JobRuntime	Send	Receive	Exitval	Signal	Command	Yes	No
1	:	1570777874.338	11.253	0	1039	0	0	/Applications/MATLAB_R2017b.app/bin/matlab -nodisplay -nosplash -nodesktop -r "run_analysis(1); exit"		
2	:	1570777874.340	11.296	0	1039	0	0	/Applications/MATLAB_R2017b.app/bin/matlab -nodisplay -nosplash -nodesktop -r "run_analysis(2); exit"		
4	:	1570777874.347	11.295	0	1039	0	0	/Applications/MATLAB_R2017b.app/bin/matlab -nodisplay -nosplash -nodesktop -r "run_analysis(4); exit"		
3	:	1570777874.343	11.398	0	1039	0	0	/Applications/MATLAB_R2017b.app/bin/matlab -nodisplay -nosplash -nodesktop -r "run_analysis(3); exit"		
5	:	1570777885.596	12.151	0	896	0	0	/Applications/MATLAB_R2017b.app/bin/matlab -nodisplay -nosplash -nodesktop -r "run_analysis(5); exit"		
7	:	1570777885.648	12.641	0	896	0	0	/Applications/MATLAB_R2017b.app/bin/matlab -nodisplay -nosplash -nodesktop -r "run_analysis(7); exit"		
8	:	1570777885.745	12.765	0	896	0	0	/Applications/MATLAB_R2017b.app/bin/matlab -nodisplay -nosplash -nodesktop -r "run_analysis(8); exit"		
6	:	1570777885.641	12.981	0	896	0	0	/Applications/MATLAB_R2017b.app/bin/matlab -nodisplay -nosplash -nodesktop -r "run_analysis(6); exit"		
9	:	1570777897.751	10.104	0	896	0	0	/Applications/MATLAB_R2017b.app/bin/matlab -nodisplay -nosplash -nodesktop -r "run_analysis(9); exit"		
10	:	1570777898.293	10.190	0	897	0	0	/Applications/MATLAB_R2017b.app/bin/matlab -nodisplay -nosplash -nodesktop -r "run_analysis(10); exit"		

Parallel processing: running (my example)



```
storing in 0x7f048604d010 and 0x7f0485d94010
5 DATA_SUR20080902 PY_DUMPSTE:: data dimensions => 381 samples, 256 channels and 232 trials
5 DATA_SUR20080902 STE:: Running @0x7f046e82a010) dims (256, 381, 232) nsym 6 and storing in 0x7f04
8ab1f010
5 DATA_SUR20080902 STE:: Will process 230 trials in 23 blocks of 10 trials (2 left out)
5 DATA_SUR20080902 STE:: Will delay 3 samples and make 5 random shuffles
5 DATA_SUR20080902 STE:: Using 12 threads to compute
5 DATA_SUR20080902 Elapsed ms in ste_timer = 69797
5 DATA_SUR20080902 PY_DUMPSTE:: data dimensions => 381 samples, 256 channels and 232 trials
5 DATA_SUR20080902 STE:: Running @0x7f041535b010) dims (256, 381, 232) nsym 6 and storing in 0x3935
6c0
5 DATA_SUR20080902 STE:: Will process 230 trials in 23 blocks of 10 trials (2 left out)
5 DATA_SUR20080902 STE:: Will delay 4 samples and make 5 random shuffles
5 DATA_SUR20080902 STE:: Using 12 threads to compute
5 DATA_SUR20080902 Elapsed ms in ste_timer = 68235
5 DATA_SUR20080902 PY_DUMPSTE:: data dimensions => 381 samples, 256 channels and 232 trials
5 DATA_SUR20080902 STE:: Running @0x7f040a6b6010) dims (256, 381, 232) nsym 6 and storing in 0x3935
6c0
5 DATA_SUR20080902 STE:: Will process 230 trials in 23 blocks of 10 trials (2 left out)
5 DATA_SUR20080902 STE:: Will delay 6 samples and make 5 random shuffles
5 DATA_SUR20080902 STE:: Using 12 threads to compute
5 DATA_SUR20080902 Elapsed ms in ste_timer = 70496
5 DATA_SUR20080902 PY_DUMPSTE:: data dimensions => 381 samples, 256 channels and 232 trials
5 DATA_SUR20080902 STE:: Running @0x7f03f4d6c010) dims (256, 381, 232) nsym 6 and storing in 0x3935
6c0
5 DATA_SUR20080902 STE:: Will process 230 trials in 23 blocks of 10 trials (2 left out)
5 DATA_SUR20080902 STE:: Will delay 7 samples and make 5 random shuffles
5 DATA_SUR20080902 STE:: Using 12 threads to compute
5 DATA_SUR20080902 Elapsed ms in ste_timer = 71055

1 [|||||||182 ] 9 [||||||| ] 17 [|||||||100.0%] 25 [|||||||186.2 ]
2 [|||||||186. ] 10 [||||||| ] 18 [|||||||100.0%] 26 [|||||||184. ]
3 [|||||||188. ] 11 [|||||||17 ] 19 [|||||||97.4%] 27 [||||||| ]
4 [|||||||195.4%] 12 [||||||| ] 20 [|||||||79 ] 28 [|||||||17 ]
5 [|||||||188.9 ] 13 [||||||| ] 21 [|||||||100.0%] 29 [||||||| ]
6 [|||||||191.4 ] 14 [|||||||182 ] 22 [|||||||100.0%] 30 [||||||| ]
7 [|||||||197.4%] 15 [||||||| ] 23 [|||||||96.1%] 31 [||||||| ]
8 [|||||||100.0%] 16 [|||||||17 ] 24 [|||||||96.7%] 32 [||||||| ]

Mem[||||||| ] Tasks: 127, 328 thr; 27 running
Swp[ ] Load average: 27.34 27.67
Uptime: 13:38:12

PID USER PRI NI VIRT RES SHR S CPU% MEM% TIME+ Command
29998 fraimondo 39 19 3427M 1734M 15520 R 1599 2.8 32:25.90 /opt/anaconda/bin/python ste_band.py --nth
29874 fraimondo 39 19 3427M 1734M 15520 R 899. 2.7 39:56.72 /opt/anaconda/bin/python ste_band.py --nth
29948 fraimondo 39 19 3427M 1734M 15520 R 100. 2.7 2:36.02 /opt/anaconda/bin/python ste_band.py --nth
29946 fraimondo 39 19 3427M 1734M 15520 R 100. 2.7 2:36.14 /opt/anaconda/bin/python ste_band.py --nth
29944 fraimondo 39 19 3427M 1734M 15520 R 100. 2.7 2:35.77 /opt/anaconda/bin/python ste_band.py --nth
29945 fraimondo 39 19 3427M 1734M 15520 R 100. 2.7 2:36.00 /opt/anaconda/bin/python ste_band.py --nth
29935 fraimondo 39 19 3427M 1734M 15520 R 100. 2.7 2:35.46 /opt/anaconda/bin/python ste_band.py --nth
30064 fraimondo 39 19 3464M 1770M 15512 R 100. 2.8 2:00.90 /opt/anaconda/bin/python ste_band.py --nth
30067 fraimondo 39 19 3464M 1770M 15512 R 100. 2.8 2:00.90 /opt/anaconda/bin/python ste_band.py --nth
30057 fraimondo 39 19 3464M 1770M 15512 R 100. 2.8 2:00.90 /opt/anaconda/bin/python ste_band.py --nth
29947 fraimondo 39 19 3427M 1734M 15520 R 99.7 2.7 2:36.09 /opt/anaconda/bin/python ste_band.py --nth
30071 fraimondo 39 19 3464M 1770M 15512 R 99.7 2.8 2:00.62 /opt/anaconda/bin/python ste_band.py --nth
30059 fraimondo 39 19 3464M 1770M 15512 R 99.7 2.8 2:00.90 /opt/anaconda/bin/python ste_band.py --nth
30062 fraimondo 39 19 3464M 1770M 15512 R 99.7 2.8 2:00.87 /opt/anaconda/bin/python ste_band.py --nth
30060 fraimondo 39 19 3464M 1770M 15512 R 99.7 2.8 2:00.90 /opt/anaconda/bin/python ste_band.py --nth
29934 fraimondo 39 19 3427M 1734M 15520 R 99.7 2.7 2:35.86 /opt/anaconda/bin/python ste_band.py --nth
30066 fraimondo 39 19 3464M 1770M 15512 R 99.7 2.8 2:00.87 /opt/anaconda/bin/python ste_band.py --nth
30061 fraimondo 39 19 3464M 1770M 15512 R 99.7 2.8 2:00.89 /opt/anaconda/bin/python ste_band.py --nth
30068 fraimondo 39 19 3464M 1770M 15512 R 99.7 2.8 2:00.90 /opt/anaconda/bin/python ste_band.py --nth
29942 fraimondo 39 19 3427M 1734M 15520 R 99.7 2.7 2:36.03 /opt/anaconda/bin/python ste_band.py --nth
30069 fraimondo 39 19 3464M 1770M 15512 R 99.7 2.8 2:00.90 /opt/anaconda/bin/python ste_band.py --nth
30070 fraimondo 39 19 3464M 1770M 15512 R 99.7 2.8 2:00.90 /opt/anaconda/bin/python ste_band.py --nth

F1|help F2|Setup F3|Search F4|Filter F5|Tree F6|SortBy F7|Nice F8|Nice + F9|Kill F10|Quit

1 [|||||||100.0%] 7 [|||||||98.7%] 13 [|||||||100.0%] 19 [|||||||98.7%]
2 [|||||||100.0%] 8 [|||||||100.0%] 14 [|||||||100.0%] 20 [|||||||98.1%]
3 [|||||||100.0%] 9 [|||||||94.2%] 15 [|||||||100.0%] 21 [|||||||88. ]
4 [|||||||100.0%] 10 [|||||||85.1%] 16 [|||||||100.0%] 22 [|||||||98.7%]
5 [|||||||100.0%] 11 [|||||||97.4%] 17 [|||||||100.0%] 23 [|||||||95.5%]
6 [|||||||100.0%] 12 [|||||||17 ] 18 [|||||||100.0%] 24 [|||||||77. ]

Mem[||||||| ] Tasks: 128, 457 thr; 27 running
Swp[ ] Load average: 23.18 22.71
Uptime: 13 days, 10:01:24

PID USER PRI NI VIRT RES SHR S CPU% MEM% TIME+ Command
123245 fraimondo 39 19 2955M 1564M 15384 R 997. 3.2 12:24.01 /opt/anaconda/bin/python ste_band.py --nth
123258 fraimondo 39 19 2955M 1564M 15384 R 100. 3.2 1:09.92 /opt/anaconda/bin/python ste_band.py --nth
123254 fraimondo 39 19 2955M 1564M 15384 R 99.8 3.2 1:09.92 /opt/anaconda/bin/python ste_band.py --nth
123256 fraimondo 39 19 2955M 1564M 15384 R 99.0 3.2 1:09.90 /opt/anaconda/bin/python ste_band.py --nth
123253 fraimondo 39 19 2955M 1564M 15384 R 99.0 3.2 1:09.68 /opt/anaconda/bin/python ste_band.py --nth
123252 fraimondo 39 19 2955M 1564M 15384 R 99.0 3.2 1:09.92 /opt/anaconda/bin/python ste_band.py --nth
123255 fraimondo 39 19 2955M 1564M 15384 R 99.0 3.2 1:09.90 /opt/anaconda/bin/python ste_band.py --nth
123250 fraimondo 39 19 2955M 1564M 15384 R 99.0 3.2 1:09.90 /opt/anaconda/bin/python ste_band.py --nth
123251 fraimondo 39 19 2955M 1564M 15384 R 99.0 3.2 1:09.90 /opt/anaconda/bin/python ste_band.py --nth
123257 fraimondo 39 19 2955M 1564M 15384 R 99.0 3.2 1:09.46 /opt/anaconda/bin/python ste_band.py --nth
18773 20 0 3298M 1197M 1143M S 2.0 2.5 30:52.46 /usr/lib/virtualbox/VirtualBox --comment M|
8338 fraimondo 20 0 30620 2772 1440 R 1.0 0.0 11:26.31 htop
7884 fraimondo 20 0 26096 2568 1448 R 2.0 0.0 12:06.09 htop
5593 20 0 3298M 1197M 1143M S 1.0 2.5 20:54.80 /usr/lib/virtualbox/VirtualBox --comment M|
25569 20 0 590M 25916 4476 S 0.0 0.1 6:28.81 /usr/lib/erlang/erts-5.8.5/bin/beam.smp -B|
18792 20 0 590M 25916 4476 S 0.0 0.1 6:11.16 /usr/lib/erlang/erts-5.8.5/bin/beam.smp -B|
25350 20 0 487M 12448 8752 S 0.0 0.0 3:33.62 indicator-multiloop |
18806 20 0 910M 56912 31200 S 0.0 0.1 4:54.04 /usr/lib/firefox/plugin-container /usr/lib|
2169 20 0 112.03 /home/inev/.dropbox-dist/dropbox-lnx.x86_64|
2718 20 0 1937M 301M 72516 S 0.0 0.6 11:34.15 /usr/bin/gnome-shell |
5792 20 0 2955M 1564M 15384 S 0.0 3.2 0:27.80 /opt/anaconda/bin/python ste_band.py --nth|
4414 20 0 800.39 /opt/anaconda/bin/python ste_band.py --nth|
4451 20 0 267M 9168 5744 S 0.0 0.0 19:22.58 /opt/teamviewer9/tv_bin/teamviewerd -f
25388 20 0 4971M 476M 92696 S 0.0 1.0 7:29.19 /usr/local/MATLAB/R2013a/bin/glnx64/MATLAB
4444 20 0 1513M 388M 54856 S 0.0 0.0 8:05.18 /usr/lib/firefox/firefox
1988 20 0 1513M 388M 54856 S 0.0 0.0 8:14.79 /usr/local/MATLAB/R2013a/bin/glnx64/MATLAB
4449 20 0 267M 9168 5744 S 0.0 0.0 20:07.55 /usr/lib/firefox/firefox
4446 20 0 1765M 41020 8156 S 0.0 0.1 17:20.40 c:\TeamViewer\TeamViewer.exe
2061 20 0 132M 4552 3764 S 0.0 0.0 0:45.96 /usr/lib/accountsservice/accounts-daemon

F1|help F2|Setup F3|Search F4|Filter F5|Tree F6|SortBy F7|Nice F8|Nice + F9|Kill F10|Quit

[24] 0:~$* "tee run.log" 00:01 02-Dec-14
```

Remarks



1. You will most probably need to learn a bit about Unix/Linux
2. Check local cluster computing service
 - Matlab on those settings is not that easy
3. Use more than one computer
 - GNU parallel supports that!
4. Don't be afraid of asking for help
5. Invest time in saving time

HOW LONG CAN YOU WORK ON MAKING A ROUTINE TASK MORE EFFICIENT BEFORE YOU'RE SPENDING MORE TIME THAN YOU SAVE?
(ACROSS FIVE YEARS)

	HOW OFTEN YOU DO THE TASK					
	50/DAY	5/DAY	DAILY	WEEKLY	MONTHLY	YEARLY
1 SECOND	1 DAY	2 HOURS	30 MINUTES	4 MINUTES	1 MINUTE	5 SECONDS
5 SECONDS	5 DAYS	12 HOURS	2 HOURS	21 MINUTES	5 MINUTES	25 SECONDS
30 SECONDS	4 WEEKS	3 DAYS	12 HOURS	2 HOURS	30 MINUTES	2 MINUTES
1 MINUTE	8 WEEKS	6 DAYS	1 DAY	4 HOURS	1 HOUR	5 MINUTES
5 MINUTES	9 MONTHS	4 WEEKS	6 DAYS	21 HOURS	5 HOURS	25 MINUTES
30 MINUTES		6 MONTHS	5 WEEKS	5 DAYS	1 DAY	2 HOURS
1 HOUR		10 MONTHS	2 MONTHS	10 DAYS	2 DAYS	5 HOURS
6 HOURS				2 MONTHS	2 WEEKS	1 DAY
1 DAY					8 WEEKS	5 DAYS



fraimondo@uliege.be



@fraimondo



@RaimondoFede