



**Maynooth
University**

National University
of Ireland Maynooth

Development and Evaluation of OSSMM, an Open-Source Sleep Monitor and Modulator

A dissertation submitted for the degree of
Doctor of Philosophy

By:

Jonny Giordano

Under the supervision of:

Barak A. Pearlmutter

Hamilton Institute
Maynooth University

October 2025

For:
C.H. (04/26/2026)
Pip (07/14/2024)
D.G.D. (01/24/2015)
M.G.K. (11/01/2012)

Declaration

I hereby declare that I have produced this manuscript without the prohibited assistance of any third parties and without making use of aids other than those specified.

The thesis work was conducted from September 2020 to October 2025 under the supervision of Barak Pearlmutter in the Hamilton Institute at Maynooth University.

Jonny Giordano

October 2025

Sponsor



This thesis has emanated from research conducted with the financial support of Taighde Éireann – Research Ireland under Grant numbers 18/CRT/6049. The opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the Research Ireland.

Abstract

This thesis investigated the abandonment of rapid eye movement (REM) sleep reduction as a therapeutic intervention for depression and developed an open-source solution to accessibility barriers in sleep research. A targeted review revealed that all three controlled human studies of REM sleep reduction reported symptom improvement, but contained significant methodological concerns. Even so, these limitations do not justify abandonment of this research. The primary contribution is the Open-Source Sleep Monitor and Modulator (OSSMM) system, with a component cost under €40, which captures putative frontal electroencephalography (EEG), electrooculography (EOG), photoplethysmography (PPG), plus movement data from an inertial measurement unit (IMU). Notably, EEG and EOG signals are captured using inexpensive conductive thermoplastic polyurethane (CTPU) electrodes sourced from consumer heart-rate monitor chest straps. A technical evaluation over a 17-night, single-participant dataset, with a silver-standard reference monitor demonstrated that machine learning classifiers achieved a Macro F1 of 0.810 and a Cohen’s kappa (κ) of 0.733, confirming the system captures physiologically relevant information for reliable sleep stage discrimination. Complete hardware designs, software, and assembly instructions are freely available, providing researchers worldwide with an accessible platform for sleep monitoring and modulation studies, thereby supporting renewed investigation of sleep interventions for depression.

Declaration of Novelty

This dissertation advances accessible sleep staging technology through the development and evaluation of an open-source, low-cost platform (under €40) that can be locally assembled using only commercially available components and 3D-prints. The technical evaluation demonstrates the system’s feasibility for capturing putative frontal EEG signals suitable for automated sleep stage classification. A key contribution is showing that inexpensive conductive Thermoplastic Polyurethane electrodes, sourced from consumer heart rate monitor chest straps, can capture these EEG sleep signatures with sufficient signal quality for machine learning classification. To enable replication and extension by the research community, complete hardware designs, firmware, software, and comprehensive build documentation are released as open-source deliverables.

Open-Source Repository

All hardware designs, firmware, and build documentation for the Open-Source Sleep Monitor and Modulator (OSSMM) are openly available at:

<https://github.com/jvgiordano/OSSMM>

With instructional material at:

<https://jvgiordano.github.io/OSSMM>.

Acknowledgements

I would like to give thanks to the following people and organizations that have helped me to undertake this research:

To SFI Centre for Research Training in Foundations of Data Science, for providing the opportunity to pursue this research.

To Professor Charles Markham, who at no benefit to himself, acted like a second thesis supervisor, introduced me to 3D printing, and benefited me with something like “advice,” but which I’m not allowed to call “advice.” Your Mount Vesuvius story will always remind me that it’s better to laugh.

To Professor David Malone, who always provided a willing ear, sound advice, and graciously supported my extension.

To Professor Barak Pearlmutter, who not only allowed but supported my curiosity and personal intuition throughout the thesis, a privilege rarely observed in modern academia.

To Patsy, for always believing in me.

To Dáire, for being my mentor, for the walks around Maynooth, the good long talks, and for keeping in touch.

To Shauna and Akash, for the lab chats and the laughs. This acknowledgment is too short for all the moral support, jokes, and help in year 4. I overestimated myself, and underestimated how long it would take to wrap things up.

To Chang and Darshana, for all the advice and always trying to help.

To Fergal, for being my third unofficial supervisor, and for the cat chats.

To Bill, for letting me use all of your desk.

To Nathan, for all the discussions, from American high school to rural Ireland.

To Ahmed, for the coffees, for the stories, and for letting me know when things were my problem to deal with.

To Anthony and Gabriel, for your patience and help with ML.

To Cormac, for the talks in the kitchen and for the help with ethics.

To Seán, for the extensive help with the thesis, being an engineering sound board, and the forever positive outlook. Amor fati.

To Connor, for letting me complain, and complaining with me.

To Fred and Giulia, for surviving Cluain Aoibhlin together.

Contents

List of Figures	xvi
-----------------	-----

List of Tables	xx
----------------	----

1 Introduction	1
1.1 The Depression Crisis and Sleep	2
1.2 What is Sleep, and What is It For?	3
1.3 An Introduction to Sleep Architecture	4
1.4 Measuring Sleep: Approaches and Accessibility Challenges	7
1.4.1 Polysomnography: The Gold Standard	7
1.4.2 The Inconvenience and Cost Barrier	8
1.4.3 The Accuracy-Cost Trade-off	8
1.5 The Critical Gap: From Monitoring to Modulation	10
1.6 The Sleep-Depression Connection	11
1.7 Research Problem and Opportunity	12
1.8 Research Aims and Objectives	12
1.9 Scope and Boundaries	13
1.10 Original Contributions	14
1.11 Thesis Structure	14
1.12 Ethical Considerations	15
2 REM Sleep Reduction and Depression: A Deep Dive Review	16
2.1 Introduction	17
2.1.1 REM Sleep Experiments	17
2.1.2 REM Reduction and Alleviation of Depressive Symptoms — Early Observations	18
2.1.3 Review Scope and Approach	19
2.2 REM Reduction and Depression — A Deep Dive	20
2.2.1 Vogel et al., 1975 — Endogenous Depression Cohort	20
2.2.2 Vogel et al., 1975 — Reactive Depression Cohort	27
2.2.3 Grözinger et al., 2002	29

2.2.4	Cartwright et al., 2003	34
2.2.5	Cross Study Synthesis	38
2.3	Hidden REM effects in NREM Interventions	41
2.3.1	Landsness et al., 2011	42
2.3.2	Assessment	44
2.4	Pharmacology, REM Sleep, and Depression	45
2.5	Discussion	51
2.5.1	Summary of Findings	51
2.5.2	Mechanistic Clues from Adjacent Research	54
2.5.3	Responders and Depression	56
2.5.4	Conclusion	59
3	Design Evolution of the OSSMM System	61
3.1	Introduction	62
3.2	Design Requirements and Goals	63
3.2.1	Design Constraints	63
3.2.2	Design Goals	65
3.3	System Conception and Architectural Decisions	65
3.3.1	Hardware Architecture Conceptualization	67
3.3.2	Hardware Safety	68
3.4	Foundational Investigations	68
3.4.1	Foundational Backbone - Analog Signal to Wireless Data Trans- mission Chain	68
3.4.2	Biosignal Selection and Component Foundations	72
3.4.3	EOG Electrode Placement Investigation	77
3.4.4	Microcontroller and Wireless Protocol Evaluation	81
3.5	Prototype #1: Minimum Viable Prototype	91
3.5.1	Design Objectives and Rationale	91
3.5.2	Component Integration	91
3.5.3	Battery Selection	92
3.5.4	Additional Sensors and Stimulator	93
3.5.5	Mechanical Design and Assembly	93
3.5.6	Cost Analysis	96
3.5.7	Prototype #1 — Performance Observations	97
3.5.8	Summary and Identified Limitations	99
3.6	Prototype #2	101
3.6.1	Design Evolution and Objectives	101
3.6.2	Integrated Redesign: From Sleep Mask to Headband	101
3.6.3	Electronic Case Redesign and Integration	103

3.6.4	Manufacturing Simplification and Modularity Improvements	105
3.6.5	Implementation Challenges	106
3.6.6	Prototype #2 — Weight and Cost Analysis	107
3.6.7	Prototype #2 — Performance Evaluation	109
3.6.8	Critical Design Flaws	110
3.6.9	Lessons and Path Forward	111
3.7	Prototype #3: Release Version	111
3.7.1	Design Evolution and Objectives	111
3.7.2	Mistakes and Progress	112
3.7.3	Microcontroller Migration - Seeed Xiao Sense	114
3.7.4	Mechanical Design Optimization	115
3.7.5	Assembly, Manufacturing, and Accessibility	118
3.7.6	Prototype #3 — Weight and Cost Analysis	120
3.7.7	Prototype #3 — Performance Observations	122
3.7.8	Identified Flaws and Future Work	127
3.8	Comparative Analysis and Design Evolution	127
3.8.1	Evolution of Key Metrics	127
3.8.2	The Four Horsemen of Progress	130
3.8.3	Revisiting the Design Requirements and Goals	131
3.8.4	Summary Design Evolution	134
3.9	Software Ecosystem Evolution	135
3.9.1	MCU Firmware	135
3.9.2	Companion Android App	135
3.9.3	Future Software Work	137
3.10	Conclusions	137
4	Technical Specification of the Open-Source Release	139
4.1	Introduction	140
4.1.1	Chapter Scope and Objectives	140
4.1.2	Design Philosophy and Open-Source Principles	140
4.2	System Architecture	140
4.3	Hardware Implementation: The Recorder	142
4.3.1	Mechanical Design and Materials	142
4.3.2	Electronic Hardware	144
4.4	Firmware and Software Ecosystem	146
4.4.1	Recorder Firmware (V1.0.4)	146
4.4.2	Base Station Companion App (Android)	152
4.4.3	Bluetooth Sampling Cadence and Timestamping	152
4.5	Safety, Regulatory, and Ethical Considerations	154

4.5.1	Electrical and Battery Safety	154
4.5.2	Material Biocompatibility and Hygiene	155
4.5.3	Data Security and Privacy	155
4.5.4	Risk Analysis and Mitigation	155
4.6	The Open-Source Deliverable: Dissemination and Reproducibility . .	156
4.6.1	Open-Source Repository	156
4.6.2	Comprehensive Build and Configuration Guide	157
4.7	Discussion	158
4.7.1	Design Trade-offs and Justifications	158
4.7.2	Limitations and Critique	159
4.7.3	Contribution to Open-Source Biomedical Hardware	160
4.7.4	Constraints and Innovation	160
5	Proof-of-Concept Evaluation: Classical Machine Learning	161
5.1	Introduction	162
5.2	Methods	164
5.2.1	Experimental Design	164
5.2.2	Ethics, Health and Safety	165
5.2.3	Experimental Equipment Setup	166
5.2.4	Data Methods	166
5.2.5	OSSMM System Hardware and Software	167
5.2.6	Signal Processing and Feature Selection	167
5.2.7	Classifiers	168
5.2.8	Performance Metrics	169
5.3	Results	170
5.3.1	Durability, Wearability, and Power	170
5.3.2	Connection Stability	170
5.3.3	Data	171
5.3.4	Classifier Configuration Selection — Group K-Fold	172
5.3.5	Classifier Performance on Test Data	173
5.3.6	Night-to-Night Classifier Performance	175
5.3.7	Feature Selection Importance by Ablation	177
5.4	Discussion	179
5.4.1	System Reliability	180
5.4.2	Data Quality for Automated Sleep Staging	180
5.4.3	Feature and Sensor Contribution	183
5.4.4	Limitations and Proof-of-Concept Scope	185
5.4.5	Conclusions	186

6	Proof-of-Concept Evaluation: Deep Learning Approaches	188
6.1	Introduction	189
6.1.1	Deep Learning for Sleep Stage Classification	189
6.1.2	Study Objectives	190
6.2	Methods	191
6.2.1	Experimental Data	191
6.2.2	Spectrogram Generation	191
6.2.3	Convolutional Neural Network Architecture	195
6.2.4	Hyperparameter Selection and Regularization	197
6.2.5	Training Protocol	200
6.2.6	Performance Metrics	200
6.2.7	Implementation and Reproducibility	201
6.3	Results	202
6.3.1	Visual Inspection of EEG/EOG Spectrogram	202
6.3.2	Classification Performance on Test Set	203
6.3.3	Sensor Modality Hierarchy	205
6.3.4	Night-to-Night Performance Variability	206
6.4	Discussion	208
6.4.1	CNN Performance Relative to Random Forest Classifier	208
6.4.2	Persistence of Sensor Hierarchy Across Methodologies	209
6.4.3	Reference Device Ceiling Question	210
6.4.4	Implications for CTPU Electrode Performance	212
6.4.5	Limitations	213
6.4.6	Conclusion	214
7	Future Work and Conclusion	216
7.1	Overview	217
7.2	Contributions	217
7.2.1	Literature Review	217
7.2.2	System Development	218
7.2.3	System Evaluation	218
7.3	Future Directions	219
7.3.1	Project 1 — Validation and Generalisation of the OSSMM Platform	219
7.3.2	Project 2 — Closing the Loop: Real-Time Staging and Selective Modulation	220
7.3.3	Project 3 — Clinical Investigation of REM-R for Depression	221
7.4	Context and Scope	223
7.5	Conclusion	224

A Appendix — REM Sleep Reduction and Depression: A Deep Dive Review	226
A.1 Normality and Homogeneity Tests for Parametric Testing for Endogenous Patient Data	226
A.1.1 Normality	226
A.1.2 Homogeneity of Variance	227
A.2 Additional Methodological Considerations for Cartwright et al. 2003 .	229
A.2.1 A Note on Heterogeneity and the Specificity Question	229
B Appendix — Design Evolution of the OSSMM System	232
B.1 Foundational Prototyping Detail	232
B.1.1 MIT App Inventor and the Foundational Smartphone Application	232
B.1.2 Kivy Detour	233
B.1.3 Transition to Dart and Flutter	233
B.2 Arduino Nano 33 IoT — Full Exploration of the Bluetooth Classic Limitation	234
B.3 Bluetooth Classic and Bluetooth Low Energy — Technical Description	235
B.3.1 Conceptual Models	236
B.3.2 Interoperability, Hardware, and Libraries	236
B.3.3 Achieving Sufficient BLE Throughput	237
B.4 HRMCS Modification Investigations	238
B.4.1 Snap-Electrode Relocation	238
B.4.2 Direct CTPU Material and Conductive 3D-Printing Filament	239
B.5 Prototype #2 — Additional Trial Observations	239
B.6 Component Sources and Pricing Receipts	242
C Appendix — Technical Specification of the Open-Source Release	251
C.1 OSSMM Recorder (headband) Arduino Code	251
D Appendix — Proof-of-Concept Evaluation: Classical Machine Learning	268
D.1 Feature Selections	268
D.2 Hyperparameter Configurations test using Group K-Fold on Training Data	273
D.3 Confusion Matrices for Selected SVM, RF, and XGB Configurations on Test Data	275
D.4 ROC Plots for SVM, RF, and XGB on Test Set	277
D.5 Per-Class Metrics	279
D.6 Feature Importance of Random Forest and XGBoost on Test Data . .	281

D.7	Random Forest — Comprehensive Ablation Study Results	283
D.8	Complete Ablation Investigation Metrics	285
E	Appendix — Proof-of-Concept Evaluation: Deep Learning Ap- proaches	288
E.1	Hyperparameter Exploration	288
E.2	Per Class Metrics on Withheld Data	291
E.3	ROC Plots on Withheld Data	291
E.4	Normalized Confusion Matrices	291
F	Appendix — Ethics	296
F.1	Approval for OSSMM Evaluation on Awake Participants	296
F.2	Approval for OSSMM Technical Evaluation	298
	Bibliography	300

List of Figures

1.1	Typical Hypnogram	4
1.2	Example of a K-Complex, Sleep Spindle, and Delta Waves	6
1.3	State-of-the-Art (2022) of Affordable Sleep Monitoring Devices	9
2.1	Vogel et al. 1975 — Participant Progression History	22
2.2	Grözinger et al. 2002 — Participant Progression History	30
2.3	Cartwright et al. (2003) Figure 1	37
3.1	Display of 3D Prints During OSSMM Development	62
3.2	Early OSSMM System Conceptualization	67
3.3	Foundational Setup	69
3.4	Read out of voltage over an adjustable potentiometer shown in 3.3 on Arduino Serial Monitor (left) and Arduino Serial Plotter (right).	70
3.5	Display of momentary voltages on an Android app built with MIT App Inventor.	70
3.6	Multi-Channel Analog Capture on the Foundational Setup	71
3.7	Flutter App with CSV Saving Capability.	72
3.8	Heart Rate Chest Monitor Strap Breakdown	74
3.9	Exploded View of CTPU Electrode from a Heart Rate Monitor Chest Strap	75
3.10	HRMCS CTPU Electrode and EOG Detection	76
3.11	Investigated EOG Electrode Placements	78
3.12	EOG Response: Placement 3	79
3.13	EOG Response: Placement 2	80
3.14	EOG Response: Placement 1	80
3.15	M5StickC Plus - Taken from (M5Stack, 2025)	84
3.16	M5StickC Plus and M5 Atom Matrix Comparison	86
3.17	Top and Bottom View of Adafruit Feather Sense	89
3.18	Front View of Prototype #1	91
3.19	CAD and Interior View of Electronic Case V0.1.0	94
3.20	Views of Prototype #1 Assembly	95
3.21	Prototype #1 — Intentional Eye Movements	97

3.22	Prototype #1 — REM-like Sleep	98
3.23	Prototype #1 — Sleep Movement Artifacts	99
3.24	Rear View of Prototype #1	100
3.25	Isometric View of Prototype #2	101
3.26	Early Version of Prototype #2 Showing Strap Fold	103
3.27	Rear View of Electronic Case and Inside View of Electronic Case Slot of Prototype #2	104
3.28	Prototype #2 CAD Annotated Views	104
3.29	Prototype #2 — Intentional Eye Movement	109
3.30	Prototype #2 — REM-like Sleep	110
3.31	Isometric View of Prototype #3	112
3.32	AD8232 PCB 3.5mm Jack Quick Removal Method	113
3.33	Exploded CAD View of Prototype #3 with Annotation	115
3.34	Component Set for Prototype #3 Electronic Case	116
3.35	Annotated Rear View of Prototype #3	118
3.36	Annotated View of Prototype #3 Separated in Major Components	119
3.37	Prototype #3 — Intentional Eye Movements	122
3.38	Prototype #3 — NREM Light Sleep	123
3.39	Prototype #3 — NREM Deep Sleep	124
3.40	Prototype #3 — REM Sleep	125
3.41	Prototype #3 — REM Sleep (Waking from Dream)	126
3.42	Prototype #3 — Sleep Movement Artifact	126
3.43	Top View of All Major Prototypes	128
3.44	OSSMM Application Screenshots	136
4.1	Block Diagram of the OSSMM System Architecture	141
4.2	Sliced Views of Receiver/Chassis	143
4.3	Front and Rear Isometric Views of Prototype #3	144
4.4	OSSMM BLE Packet Layout	148
4.5	Current Consumption vs Sampling Frequency	150
4.6	OSSMM V1.0.4 Power Usage and BLE Transmission	151
4.7	OSSMM Sample Count for 30 Second Epochs	154
4.8	OSSMM GitHub Landing Page	157
5.1	Breakdown of Experimental Nights — Train, Test, Discarded	166
5.2	Night 16 — REM Sleep Observation	172
5.3	Conventional Classifier Performance Comparison — SVM, RF, XGBoost	173
5.4	Normalized Confusion Matrix for RF Classifier	174
5.5	Top 10 Feature Importance Ranking for RF Classifier	175

5.6	Somnofy and Predicted Hypnogram Overlay for Test Nights (13 and 17)	177
5.7	Weighted F1-Score Comparison by Feature Combination	178
6.1	EOG/EEG Spectrogram Observations from Night 17 for 4 Sleep-Wake States	192
6.2	PPG Spectrogram Observations from Night 17 for 4 Sleep-Wake States	194
6.3	CNN Architecture for EEG/EOG and PPG Spectrograms	196
6.4	Multi-Modal Fusion Architecture	197
6.5	Sleep Spindle Identification in Light Sleep	202
6.6	Reproduction of Fig. 1 and 2 from (Li et al., 2022)	203
6.7	Normalized Confusion Matrix for EEG+PPG CNN Classifier	205
6.8	Macro F1-Score Comparison Across Test Nights and CNN Models . . .	206
6.9	Conceptual Diagram of reference device ceiling Dilemma	210
7.1	Gantt chart of PhD timeline	223
A.1	Q-Q Plot of Residuals for Endogenous Patients in (Vogel et al., 1975) .	228
B.1	Kivy Android App Screens	233
B.2	Early Flutter Android App with Bluetooth Classic	234
B.3	NINA W102 on Arduino Nano 33 IoT	235
B.4	BLE Bulletin Board Model	236
B.5	Prototype #2 — NREM-like Sleep	240
B.6	Prototype #2 — Sleep Movement Artifacts	241
D.1	Normalized Confusion Matrix for SVM Classifier on Test Data.	275
D.2	Normalized Confusion Matrix for Random Forest Classifier on Test Data.	276
D.3	Normalized Confusion Matrix for XGBoost Classifier on Test Data. . .	276
D.4	SVM ROC Plot	277
D.5	Random Forest ROC Plot	278
D.6	XGBoost ROC Plot	278
D.7	SVM Per-Class Metrics	279
D.8	Random Forest Per-Class Metrics	280
D.9	XGBoost Per-Class Metrics	280
D.10	Top 10 Feature Importance Rankings for Random Forest Classifier using Mean Decrease in Impurity.	281
D.11	Top 10 Feature Importance Rankings for XGBoost Classifier using Mean Decrease in Impurity.	282
D.12	Random Forest Ablation Study — F1 Scores	283
D.13	Random Forest Ablation Study — Accuracy	283
D.14	Random Forest Ablation Study — Feature Counts	284

E.1	Precision, Recall, and F1 for the PPG CNN Model	291
E.2	Precision, Recall, and F1 for the EEG CNN Model	292
E.3	Precision, Recall, and F1 for the EEG+PPG CNN Model	292
E.4	ROC Plot for the PPG CNN Model	293
E.5	ROC Plot for the EEG CNN Model	293
E.6	ROC Plot for the EEG+PPG CNN Model	294
E.7	Normalized Confusion Matric for PPG CNN Model	294
E.8	Normalized Confusion Matric for EEG CNN Model	295
E.9	Normalized Confusion Matric for EEG+PPG CNN Model	295

List of Tables

1.1	Translation Table of Sleep Staging Standards	5
2.1	Vogel et al. 1975 — Outcomes Comparison of Endogenous Participants on HAMD-17 Scores	25
2.2	Vogel et al. 1975 — Outcomes Comparison of Reactive Participants on HAMD-17 Scores	28
2.3	Summary of Differences Between REM Reduction Studies for Treat- ment of Depression	39
2.4	Antidepressant effects on REM sleep latency and REM sleep. Based on Staner et al. (2010).	47
3.1	System Design Requirements for the OSSMM System	64
3.2	Design Goals for the OSSMM System.	65
3.3	M5StickC Plus Baseline Current Consumption	85
3.4	M5StickC Plus Battery Life Projections	85
3.5	Feature Comparison of nRF52840 Development Boards	88
3.6	Adafruit Feather Sense Current Consumption Across Sketches	89
3.7	Current Consumption: M5StickC Plus vs. Feather Sense	90
3.8	Feather Sense Current Draw with AD8232 Boards	90
3.9	Parts Cost for Prototype #1	96
3.10	Vibration Motor Disc Current Measurements	107
3.11	Parts Cost for Prototype #2.	108
3.12	Parts Cost of Prototype #3	120
3.13	Comparison of Prototypes #1, #2, and #3	129
3.14	Achievement Status of Design Constraints — (Part 1: Design and Per- formance)	132
3.15	Achievement Status of Design Constraints — (Part 2: Implementation and Safety)	133
3.16	Achievement Status of Design Goals	134
4.1	OSSMM BLE Packet Layout	147

5.1	Sleep Stage Distribution for Training and Test Sets	171
5.2	Conventional Classifier Selection from Hyperparameter Search	172
5.3	RF Performance Across Test Nights	176
5.4	Performance Metrics for Random Forest Classification Across Feature Combinations	179
6.1	EEG Spectrogram Generation Parameters	192
6.2	PPG Spectrogram Generation Parameters	194
6.3	CNN Architecture for EEG/EOG and PPG Spectrograms	195
6.4	Multi-Modal Fusion Architecture	197
6.5	Hyperparameter Configurations Tested for Regularization Sensitivity Analysis	198
6.6	Four-fold Cross-Validation Performance Across Hyperparameter Con- figurations	199
6.7	Compute Environment and Training Configuration	201
6.8	Results Comparison Between RF and CNN Classifiers	204
6.9	Performance Metrics Across Nights and CNN Models	207
D.1	PPG-derived Heart-rate and Hear-rate Variability Features	269
D.2	IMU-Derived Features from Accelerometer and Gyroscope	270
D.3	EOG-Derived Features for Eye-Movement	271
D.4	EOG-Derived Features for EEG Frequencies	272
D.5	Hyperparameter Configurations for Support Vector Machine	273
D.6	Hyperparameter Configurations for Random Forest	273
D.7	Hyperparameter Configurations for XGBoost	274
D.8	Performance Metrics for Random Forest Classification Across Multiple Sensor Combinations	286
D.9	Performance Metrics for Random Forest Classification Across Single Sensor Features	287
E.1	EEG-Only CNN Performance Across Configurations and Cross-Validation Folds	289
E.2	PPG-Only CNN Performance Across Configurations and Cross-Validation Folds	290

Introduction

“...I’ve tried to assemble only my original projects. However lacking the full knowledge of humankind I cannot be quite sure of the originality of my own ideas.”

— *Maxim Sergeevich Egorov, January 15, 2021*

This chapter introduces the motivations behind the work presented in this thesis and provides an overview of the material contained within the following chapters.

1.1 The Depression Crisis and Sleep

Depression is a psychiatric disorder characterized by persistent negative feelings such as sadness, hopelessness, and anhedonia (inability to experience pleasure). According to the World Health Organization, depression is the leading cause of disability worldwide and affects an estimated 300 million people, or 4.4% of the global population ([World Health Organization, 2017](#)). Depression is also a major contributor to suicide, which in 2015 accounted for nearly 1.5% of all deaths, and was the second largest cause of death for persons aged 15–29 ([World Health Organization, 2017](#)).

Current standard treatments for depression are antidepressant medication and therapy, but both have significant caveats. Typical antidepressants include Monoamine Oxidase Inhibitors (MAOIs), Tricyclic Antidepressants (TCAs), and Selective Serotonin Reuptake Inhibitors (SSRIs). Antidepressants can cause unwanted side-effects, and these affect the majority of users ([Bet et al., 2013](#)). Side effects range from minor inconveniences (dry mouth, headaches) to more serious concerns (hypertension, disturbed sleep, serotonin syndrome) ([Uher et al., 2009](#)). Furthermore, treatment with antidepressants requires a titration period and takes at least 4–6 weeks before an improvement in condition is clinically noticeable for most users ([Beitinger and Fulda, 2010](#); [Nielsen et al., 2000](#); [Volpi-Abadie et al., 2013](#)). Of particular concern, there is mounting evidence of the potential for antidepressants to increase the risk of suicidal behavior, especially for those below the age of 25 ([Hengartner, 2017](#); [Levenson and Holland, 2006](#)).

In comparison, therapy avoids many of the pitfalls associated with pharmacology, but is not without its own challenges. Standard therapies for treatment of depression include Cognitive Behavioral Therapy (CBT) and Interpersonal Therapy (IPT). Difficulties with therapies include cost, accessibility, and time commitment. Therapy requires a long treatment duration, at minimum on the order of months ([National Institute for Health and Care Excellence, 2022](#)). Both treatment types, therapy and medication, are associated with stigmas as some individuals may feel shame or embarrassment in seeking help for treatment of depression ([Pyne et al., 2004](#); [Sirey et al., 2001](#)).

However, one of the notable findings in depression research is the connection between sleep disturbances and mood. It is well-established that Total Sleep Deprivation (TSD) is an efficacious and rapid anti-depressant treatment, albeit not long lasting ([Borbély and Wirz-Justice, 1982](#); [Giedke and Schwärzler, 2002](#); [Hu et al., 2021](#); [Minkel et al., 2017](#); [Moretto and Palagini, 2019](#); [Wirz-Justice and Van den Hoofdakker, 1999](#)). More specifically, Rapid Eye Movement (REM) sleep reduction has shown promise in producing antidepressant effects. Follow-up studies have

supported the findings of the anti-depressive properties of REM sleep reduction (Cartwright et al., 2003; Vogel et al., 1975). Animal studies in both cats and rodents have also paralleled these findings, with REM sleep deprived animals exhibiting increased behavioral activity, lower reward thresholds and greater neuronal excitability (Cohen and Dement, 1965; Dement, 1965b; Steiner and Ellman, 1972).

This presents an unresolved question, one which is the inspiration and driver behind the broader aims of this thesis. If reducing REM sleep can alleviate depression, why has this potential avenue been largely abandoned? The arduous experimental requirements for such sleep manipulations are one of the main reasons they have been attempted so rarely, the last being published more than twenty years ago (Cartwright et al., 2003; Grözing et al., 2002). This thesis explores whether modern technology could lower barriers to this research and enable proper investigations of this phenomenon. In doing so, this thesis addresses a larger issue facing sleep research: the prohibitive entry costs that restrict access to the field.

1.2 What is Sleep, and What is It For?

Despite significant advances in sleep research, the specific functions of sleep and its stages remain incompletely understood, though numerous theories have been proposed (McNamara, 2019; Nayak and Anilkumar, 2023). Sleep is a complex biological phenomenon characterized by distinctive neural, muscular, and ocular activity. It involves reduced sensory awareness, physical quiescence, altered digestive processes, and occurs in predictable circadian cycles (Kryger et al., 2017; McNamara, 2019; Vaccaro et al., 2020). Sleep is a reversible state; organisms can be roused with sufficient stimulation. The near-universal occurrence of sleep across animal species underscores its biological necessity, and total sleep deprivation has been demonstrated to be fatal in multiple species (de Manaceine, 1894; Everson et al., 1989; McNamara, 2019; Rechtschaffen et al., 1983).

Sleep duration and intensity are homeostatically regulated: when sleep is restricted, subsequent sleep becomes longer or deeper to compensate for the deficit (Borbély et al., 1982; McNamara, 2019). This homeostatic regulation extends to specific sleep states (e.g., REM sleep) or signatures (e.g., slow wave activity), with selective restriction leading to corresponding rebounds (Borbély et al., 1982; Landsness et al., 2011; Vogel et al., 1975).

The dual nature of mammalian sleep, comprising non-Rapid Eye Movement (NREM) and REM sleep states, suggests that sleep serves multiple, potentially distinct functions (McNamara, 2019). Notably, while total sleep deprivation is fatal, adult humans can survive with severely suppressed REM sleep for extended periods, even

years (Beitinger and Fulda, 2010; Vogel, 1975). This dissociation between REM sleep and survival, combined with REM’s homeostatic regulation, suggests that REM sleep serves specialized functions that are not necessary for adult survival. Understanding the distinct functions of NREM and REM sleep, particularly REM’s role in mood regulation, remains critical for developing novel therapeutic interventions.

1.3 An Introduction to Sleep Architecture

Sleep is not a homogeneous state, but rather a dynamic process characterized by periods of distinct neural patterns. This was first discovered by Loomis, Harvey, and Hobart in 1937 (Loomis et al., 1937). It would not be until 1953, the same year DNA was discovered, that REM sleep was identified (Aserinsky and Kleitman, 1953; Watson and Crick, 1953). Since then sleep research has expanded considerably, and numerous associations have been uncovered between sleep states and facets of cognition and health (Kryger et al., 2017; McNamara, 2019; Palagini et al., 2013; Shepard et al., 2005; Sikka et al., 2019; Vogel et al., 1975).

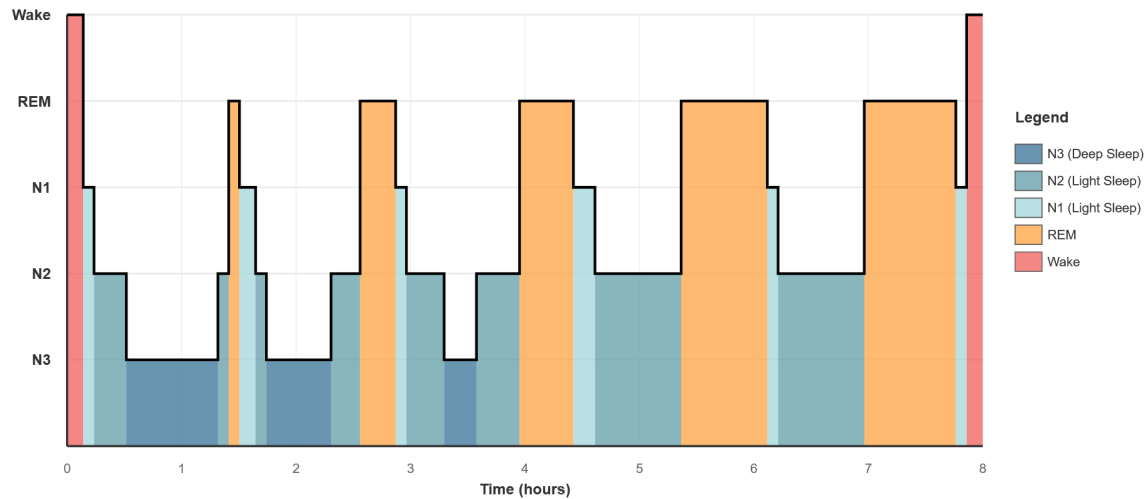


Figure 1.1: A typical hypnogram showing 5 sleep cycles during an 8 hour sleep period. N3 sleep is concentrated during the early cycles, while REM sleep periods lengthen in later cycles. Produced by the author.

Sleep is a multi-staged behavioral state of quiescence, divided in humans into two main stages: NREM sleep and REM sleep. NREM sleep is further subdivided into N1, N2, and N3 (AASM, 2023; McNamara, 2019; Nayak and Anilkumar, 2023). Sleep is structured with repeating cycles and transitions through these states. A typical sleep cycle lasts 90–120 minutes, progressing through N1, N2, N3, often returning to N2, and then into REM sleep. This cycle repeats 4–6 times over the sleep period. The proportion of NREM and REM sleep evolves throughout the night: NREM stages predominate earlier in the sleep period, while REM periods

lengthen and dominate later portions (Kryger et al., 2017; McNamara, 2019; Nayak and Anilkumar, 2023). A hypnogram, as shown in Figure 1.1, illustrates this cyclical nature by plotting sleep stages over time.

Sleep stages are defined primarily by brain wave characteristics measured through electroencephalography (EEG), with additional information from eye movements (electrooculography, EOG) and muscle tone (electromyography, EMG) (AASM, 2023; McNamara, 2019; Nayak and Anilkumar, 2023). The American Academy of Sleep Medicine (AASM) provides the standard sleep stage classification used in modern research and clinical practice. Under this system, continuous sleep data is divided into 30-second intervals, known as epochs. Each epoch is classified as N1, N2, N3, REM, or Wake based on the characteristics within that interval or observed in surrounding epochs (AASM, 2023). The AASM standard, introduced in 2007, refined the earlier Rechtschaffen & Kales (R&K) criteria from 1968 (Nayak and Anilkumar, 2023). The primary change was combining R&K Stages 3 and 4 into a single stage, AASM N3 (Nayak and Anilkumar, 2023). Table 1.1 presents the equivalent categories between these standards.

Rechtschaffen & Kales		AASM Standard	4-Stage Classification
Wake (W)	≡	Wake (W)	Wake
Stage 1	≡	N1	Light Sleep
Stage 2	≡	N2	Light Sleep
Stage 3 + Stage 4	≡	N3	Deep Sleep
REM	≡	REM (R)	REM

Table 1.1: Comparison of Rechtschaffen & Kales, AASM, and simplified 4-stage sleep classification

Under the AASM framework, N1 sleep represents the transitional state between wake and sleep. N1 is identified by the absence of alpha rhythms (8–13 Hz), the presence of theta rhythms (4–7 Hz), and slow eye movements. N2 is characterized by K-complexes and sleep spindles. K-complexes are EEG signatures lasting approximately 0.5 seconds, with a slight negative voltage peak followed by larger positive and negative peaks (Kryger et al., 2017; McNamara, 2019; Nayak and Anilkumar, 2023). Sleep spindles are similarly brief (0.5–2 seconds) bursts of activity with frequencies of 11–16 Hz (Li et al., 2022; Nayak and Anilkumar, 2023). N1 and N2 are sometimes categorized together as Light Sleep, while N3 is referred to as Deep Sleep. N3 sleep is characterized by high-amplitude, slow-wave activity (0.5–4 Hz), known as delta waves, and predominantly occurs during the first third of the sleep period (McNamara, 2019; Nayak and Anilkumar, 2023).

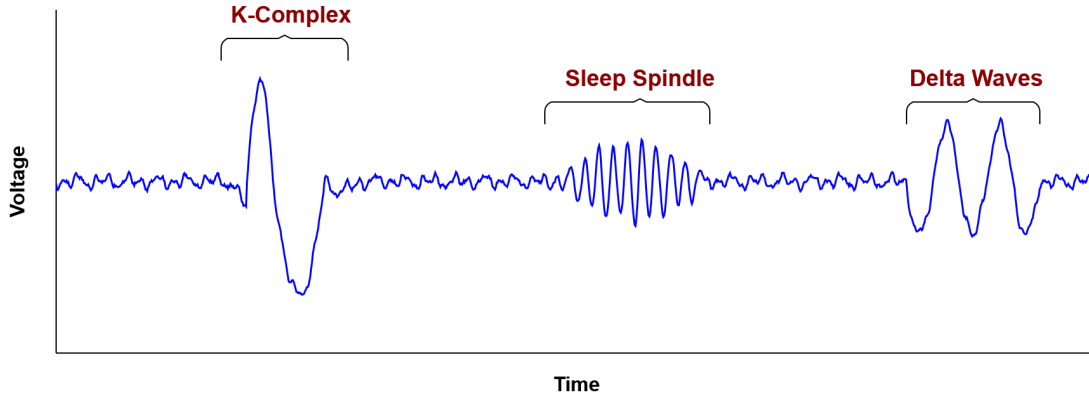


Figure 1.2: A simulated EEG channel demonstrating an example of a K-Complex, Sleep Spindle, and Delta Waves. The simulation occurs over 6 seconds, with a sampling frequency of 200 Hz. The sleep spindle has a frequency of 12 Hz, and the Delta wave of 3.25 Hz.

REM sleep is distinguished from NREM stages by its unique physiological characteristics and can be further categorized into tonic and phasic events. Tonic characteristics persist throughout most of the REM period, while phasic events occur intermittently. Phasic events include rapid eye movements, muscle twitches, and increased variability in heart rate and breathing (McNamara, 2019; Nayak and Anilkumar, 2023). Sawtooth waves, sharp triangular waveforms between 2–6 Hz, are a characteristic phasic EEG feature of REM sleep (Nayak and Anilkumar, 2023). Tonic characteristics include desynchronized, low-amplitude EEG activity and muscle atonia. Although muscle tone decreases during NREM sleep, the lowest muscle tone occurs during REM sleep (McNamara, 2019; Nayak and Anilkumar, 2023). REM sleep is also strongly associated with vivid, story-like dreaming, though dreaming can occur during all sleep stages (McNamara, 2019).

Although sleep has been presented here as five distinct stages (N1, N2, N3, REM, Wake), these stages represent points along a continuum of brain activity rather than discrete categories (Borbély et al., 1982). Characteristic features of one stage often appear in others. For example, while sleep spindles are most characteristic of N2, they can also occur in other NREM stages (McNamara, 2019). Similarly, slow waves appear in N2 sleep, with an epoch classified as N3 only when slow waves occupy at least 20% of the epoch (AASM, 2023). Despite these soft boundaries, categorical sleep staging remains valuable for advancing understanding of sleep physiology and its associations with health outcomes.

1.4 Measuring Sleep: Approaches and Accessibility Challenges

Accurate identification of sleep stages is essential for sleep research. A sleep study, called a polysomnography (PSG), captures multiple physiological signals during sleep to assess sleep architecture and diagnose disorders. Traditionally, sleep staging, the categorization of sleep into distinct classes (i.e., Wake, REM, N1, N2, N3) relies on manual scoring of PSG by qualified experts ([AASM, 2023](#)).

1.4.1 Polysomnography: The Gold Standard

An attended in-lab PSG (Type I) captures EEG from multiple scalp electrodes, combined with EOG and EMG signals to differentiate sleep stages. Additional measurements may include electrocardiogram (ECG), respiratory effort belts, airflow sensors, oxygen saturation, and leg EMG ([AASM, 2023](#); [Kryger et al., 2017](#)). While sleep stages are primarily defined by EEG signatures, this multi-modal approach enables better classification accuracy and a more comprehensive sleep assessment ([AASM, 2023](#)).

Beyond in-lab PSG, portable take-home units exist with varying measurement capabilities. AASM standard utilizes the Type categorization (Type I–IV) ([Kapur et al., 2017](#)). This classification includes: Type I (attended in-lab PSG), Type II (unattended full PSG with EEG/EOG/EMG for staging), Type III (unattended limited-channel cardiorespiratory polygraphy, typically ≥ 4 channels without EEG), and Type IV (one or two parameters, often oximetry $\pm$ heart rate). These classifications were developed primarily for obstructive sleep apnea (OSA) evaluation and insurance reimbursement. As such, each Type category technically requires cardiac and/or respiratory measurements.

This thesis uses Type nomenclature (I, II, III, IV) to contextualize device capabilities and cost structure in the sleep monitoring landscape. This framework distinguishes systems that capture neurophysiological signals for sleep staging (Type I and II) from those focused on cardiorespiratory monitoring (Type III and IV).

1.4.2 The Inconvenience and Cost Barrier

While considered the gold standard, attended in-lab PSG (Type I) has significant trade-offs: it is not ecologically valid, requiring attendance at a sleep lab, necessitates trained staff for set-up and monitoring, while demanding the participant tolerate uncomfortable, restrictive equipment throughout the sleep period.

Portable and at-home devices (Type II and Type III) were developed to address these limitations, allowing unattended sleep studies in more natural settings. However, substantial barriers remain. In the United States, a single in-laboratory PSG costs on average \$1,840, while an at-home evaluation costs \$1,575 (Kim et al., 2015). The upfront cost of acquiring portable PSG equipment is even greater: Type III systems cost approximately \$3,500 to \$5,500, while Type II systems with full EEG capability range from \$5,500 to \$9,100 (CPAP Europa, 2024a,b; Tikkanen, 2025a,b). Beyond equipment costs, manual sleep staging requires specialized training and is time-intensive. Portable devices, though allowing at-home use, often still require conductive gels or specialized electrode pads, presenting usability challenges for participants. These combined barriers of high equipment costs, labor-intensive analysis, and usability limitations create significant obstacles for research groups seeking to conduct longitudinal or large-scale sleep studies.

To illustrate, a four-week study with $N = 30$ participants and only four Type-II PSG units would require over seven months to complete, assuming constant use of the devices, and would represent an equipment cost of at least €20,000 (assuming each PSG unit costs €5,000, and not yet accounting for the cost of consumables). Furthermore, each night requires manual set-up by trained lab personnel, an additional labor and time cost. Such constraints force research groups to limit either participant numbers or observation duration.

1.4.3 The Accuracy-Cost Trade-off

Both consumer sleep technology and research efforts have emerged as potential solutions to these accessibility challenges. Reduced equipment setups requiring fewer sensor modalities or electrode placements have been employed, lowering costs and reducing movement restrictions (Arnal et al., 2020; Grözinger et al., 1999; Miller et al., 2022; Vitalthings / Somnofy). Automatic sleep scoring using machine learning eliminates costly manual analysis and reduces variability from human error (Grözinger et al., 2002; Koushik et al., 2019; Li et al., 2022; Mousavi et al., 2019; Supratak et al., 2017). However, most low-cost consumer sleep monitoring devices show poor performance when compared to PSG (Miller et al., 2022; Robbins et al., 2024; Schyvens et al., 2025). Moreover, consumer devices typically employ proprietary algorithms

that prevent researchers from understanding and validating the classification methods, limiting scientific understanding and reproducibility.

Figure 1.3 illustrates the state-of-the-art in affordable sleep monitoring solutions that were PSG-validated by 2022. The figure shows epoch-level accuracy versus price for platforms performing 4- or 5-stage sleep staging. The market landscape since 2022 reveals a concerning trend affecting the most accurate devices. The Dreem headband (and its successor, Dreem 2) was pulled from direct consumer sale before the end of 2022 as the company shifted toward clinical validation and the health industry (Dreem / Business Wire, 2021). The Somnify SM-100 is no longer available as a consumer product, with pricing increasing from approximately €270 (\$333 USD) in December 2020 to €1,778 in May 2025. Somfit, not included in the original comparison due to unavailable listing price data, currently retails for approximately €1,785 as a clinically validated headband device (Remserve Medical, 2025). Meanwhile, affordable smartwearables lack the accuracy required for research applications demanding reliable sleep staging.

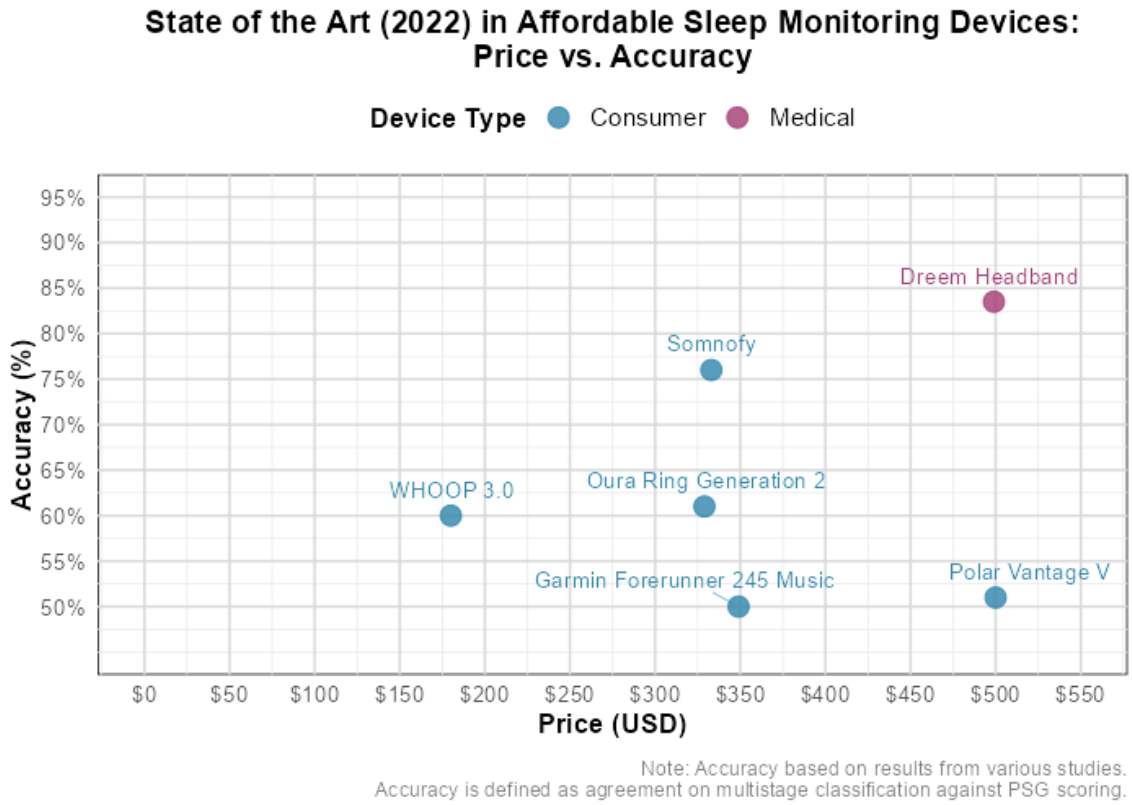


Figure 1.3: Marketplace survey displaying accuracy versus cost of low-cost platforms validated with PSG. Accuracy is epoch-by-epoch agreement with PSG scoring. Costs represent initial listing prices plus 6 months of subscription fees where required (Arnal et al., 2020; Maker, 2019; Miller et al., 2022; Sawers, 2017; So, 2021; Toften et al., 2020; Vincent, 2018; Vitalthings / Somnify).

The combination of expensive portable PSG systems (\$3,500–9,100) and inaccurate consumer devices creates a significant accessibility barrier for sleep research. Without substantial funding, researchers face a difficult trade-off; using low-cost devices with poor accuracy, or investing in high-cost systems that may limit sample sizes and study durations. This cost barrier significantly restricts both the scope of sleep research that can be conducted and access to the field, particularly for laboratories with limited budgets.

1.5 The Critical Gap: From Monitoring to Modulation

Beyond the challenges of monitoring sleep, an even greater gap exists for sleep modulation research. Sleep modulation refers to actively intervening in or manipulating sleep processes, such as enhancing specific sleep stages or suppressing unwanted states. Unlike passive monitoring, modulation investigations require real-time sleep classification capabilities for timely interventions. When the goal is to intervene during a specific sleep state or in response to a neural signature, the system must detect and respond in real-time. For example, to selectively suppress REM sleep through auditory disruption, the system must identify REM onset and deliver stimuli before the REM period has progressed too far, a capability that requires both accurate classification and minimal latency.

Sleep modulation encompasses various intervention techniques, for example, acoustic stimulation to enhance or suppress slow-wave activity, physical awakenings to prevent entry into particular sleep states, and closed-loop stimulation synchronized to enhance specific neural oscillations (Gröninger et al., 2002; Ketz et al., 2018; Landsness et al., 2011; Vogel et al., 1975). Each of these approaches requires not only the monitoring capabilities discussed previously, but also the hardware and software infrastructure to act on that detection in real-time. The technical demands are therefore greater, requiring not only signal acquisition and classification methods, but also decision logic, delivery mechanisms, and timing control.

The market for sleep modulation devices in 2020–2022 was exceptionally limited. No general-purpose platforms existed for the task. Lucid dreaming devices represented the closest commercial option, as they attempt to detect REM sleep and deliver stimuli, but these were designed for consumer use rather than research applications. A 2019 review found that only one of these devices, DreamLight, had been PSG validated, however it was no longer available for purchase. Another device, ZMax, was research-grade and showed promise, but lacked published PSG validation and

had a price point of \$900 USD (approximately €800) (Hypnodyne Corp., 2024; Mota-Rolim et al., 2019).

No viable commercial options were available for sleep modulation research during this period. This gap was particularly problematic for investigating therapeutic interventions like REM sleep reduction for depression, which requires precisely timed detection and intervention. Combined with the monitoring challenges of cost, accuracy, and proprietary algorithms, the absence of accessible modulation platforms created a substantial barrier for research groups without significant funding.

1.6 The Sleep-Depression Connection

There is a strong bidirectional association between sleep and depression (Minkel et al., 2017). Sleep complaints are nearly universal among individuals with major depression, with insomnia affecting between 70–80% of depressed patients (Ahmadi et al., 2010; Moretto and Palagini, 2019). However, the relationship extends beyond depression causing sleep problems. Longitudinal studies have demonstrated that persistent insomnia increases the risk of developing major depression by nearly fourfold, and some research has found that early treatment for insomnia may prevent the recurrence of a depressive episode, indicating a bidirectional relationship (Ahmadi et al., 2010; Minkel et al., 2017).

Polysomnographic studies have identified measurable differences in sleep architecture between healthy and depressed individuals, with REM sleep abnormalities garnering particular attention. Depressed individuals consistently display shortened REM latency, increased REM density, and a higher proportion of REM sleep, among other changes (Ahmadi et al., 2010; Minkel et al., 2017; Moretto and Palagini, 2019). These abnormalities, and the broader relationship between REM sleep and depression, are examined in detail in Chapter 2.

Inspired by REM sleep deprivation experiments in animals which resulted in increased behavioral activity, and the observed suppression of REM sleep by antidepressant medication, selective REM sleep reduction was tested in humans for treatment of depressive symptoms (Cartwright et al., 2003; Grözinger et al., 2002; Vogel et al., 1968, 1975). Three controlled studies demonstrated consistent patterns of mood improvement, yet substantial methodological differences between them raise more questions than answers. Despite these promising findings, no modern replication has been attempted in over twenty years (Cartwright et al., 2003; Grözinger et al., 2002). The counterintuitive finding that reducing an already disturbed sleep period can alleviate depressive symptoms raises fundamental questions about the role REM sleep may play in depression. This presents a compelling therapeutic pos-

sibility that has remained largely unexplored. A critical analysis of these studies, along with convergent pharmacological evidence, is presented in Chapter 2.

1.7 Research Problem and Opportunity

The connection between REM sleep and depression requires further investigation, particularly regarding if and how REM sleep modulation may be used as both a treatment and preventative measure for the onset of depressive symptoms. Research into this area has been challenging for several reasons. From an ethical standpoint, working with vulnerable participants, such as individuals suffering from depressive symptoms, requires careful consideration and appropriate safeguards. Due to the nature of this work, it must be carried out in dedicated facilities with properly trained clinicians. From a technical standpoint, sleep modulation, in particular, reduction or deprivation experiments, presents unique challenges. Sleep is self-regulating, and therefore, altering one aspect of sleep has knock-on effects to other sleep stages. The timing of interventions may be critical, even among control participants. From a financial standpoint, as has already been discussed, sleep research is expensive. However, this burden can be reduced. For example, over two decades ago, [Grözing et al. \(2002\)](#) in their follow-up to Vogel et al.'s first REM reduction experiment, utilized a neural network connected to a single EEG channel to automate sleep stage detection, resulting in a reduction of the labor required for the experiment. However, this work was limited and had significant performance drawbacks. Despite this advancement in 2002, later REM reduction and sleep modulation experiments completed nearly a decade later reverted to manual monitoring approaches ([Landsness et al., 2011](#); [Rosales-Lagarde et al., 2012](#)).

Technology has progressed substantially since then, presenting a clear research opportunity. Can modern technology enable the creation of a low-cost platform with open algorithms for sleep research and modulation experiments? Specifically, is it possible to produce an accurate monitor at an affordable price point that matches or exceeds validated affordable consumer devices in performance while undercutting them in cost?

1.8 Research Aims and Objectives

This research addresses a dual challenge in sleep science: the prohibitive cost barriers that restrict access to quality sleep monitoring technology, and the abandonment of potentially therapeutic sleep interventions, in part due to experimental challenges. The connection between REM sleep reduction and antidepressant effects, demonstrated in earlier studies but largely underexplored in modern research, represents a

compelling but technically and financially demanding area of investigation. Simultaneously, broader sleep research faces fundamental accessibility challenges, where low-cost consumer devices perform poorly compared to PSG, while research-grade systems cost thousands of euros per unit, forcing researchers to choose between participant scale and data quality.

The primary aim of this thesis is to develop a low-cost, open-source sleep monitoring and modulation system capable of automated sleep staging, thereby providing the technical basis for democratizing sleep research and enabling renewed investigation of REM sleep reduction for depression. To achieve this aim, three objectives guided this research:

- **Literature Review:** Systematically review evidence on REM sleep reduction interventions for depression to establish the evidence base and identify opportunities for contemporary investigation.
- **System Development:** Develop a low-cost ($<€130$) open-source sleep monitoring and modulation system with automated sleep staging capability.
- **System Evaluation:** Assess technical feasibility through multi-night evaluation of wearability, hardware reliability, signal fidelity, and discriminant information capture for sleep staging.

1.9 Scope and Boundaries

The initial project plan anticipated completing both system development and clinical intervention testing within a standard PhD time frame. However, hardware development required substantially more time than anticipated. While broader clinical aims remain important future goals, establishing a validated, reliable technical platform is a necessary prerequisite. Attempting clinical trials with an inadequately validated system would be ethically inappropriate. It is important to note, however, that this thesis presents a technical evaluation of the Open-Source Sleep Monitor and Modulator (OSSMM) system, not a full validation. Validation, in the formal sense, would require standardized performance metrics, multi-participant testing, and direct comparison against gold-standard PSG. This work establishes proof-of-concept and technical feasibility through single-participant evaluation against a silver-standard reference device, providing a technical foundation upon which researchers in future work can build for use in research and clinical investigations. Full validation remains an important direction for future work.

1.10 Original Contributions

Primary Contribution. The main contribution of this thesis is the production of the first open-source platform, both open hardware and open software, that is purpose-built for both sleep monitoring and modulation. This system captures putative frontal EEG, EOG, and PPG (photoplethysmogram) signals, functionally positioning it between Type II and Type III devices. All the assets and a detailed build guide for the platform have been available since May 2025 in a GitHub repository, located at: <https://github.com/jvigiordano/OSSMM>. The associated GitHub Pages site with instructional material can be found at: <https://jvigiordano.github.io/OSSMM/>. The investigations conducted during this thesis provide preliminary evidence supporting the viability of conductive Thermoplastic Polyurethane electrodes from low-cost heart rate monitor straps for EEG signal collection. A differential frontal EEG montage, approximately Avg(F7, Fp1)–Avg(Fp2, F8), was found to capture sufficient information for sleep stage discrimination when evaluated against a silver-standard research device (Somnify SM-100). Machine learning analysis demonstrated that conventional classifiers (XG-Boost, Random Forest) achieved Macro F1 scores of 0.741 and 0.770 for sleep staging from the captured signals, with a lightweight Convolutional Neural Network architecture reaching 0.810, supporting the system’s viability for automated sleep staging applications. While the OSSMM system is not yet complete for interventional studies, the work herein establishes a proof-of-concept that affordable sleep staging suitable for research applications is technically feasible at a price point below €40. Beyond its technical contributions, this work advances the broader goal of democratizing sleep research by releasing a complete, reproducible platform that reduces cost and lowers barriers to entry.

Secondary Contribution. The secondary contribution of this thesis is an in-depth review of REM-R as a treatment for depression, critically analyzing the three existing controlled human trials and highlighting important methodological differences that have obscured conclusions. It examines an NREM intervention revealing hidden REM correlations, and reviews adjacent REM-R research in healthy controls suggesting mechanisms of emotional reactivity and cortical excitability. Finally, it examines pharmacological evidence showing that most traditional antidepressants consistently suppress REM sleep, pointing to convergent behavioral and pharmacological pathways.

1.11 Thesis Structure

This thesis is structured into five main chapters, followed by a concluding one. Chapter 2 reviews the limited human studies investigating REM sleep reduction as

a treatment for depressive symptoms. The design evolution of the OSSMM system developed in this work is documented in Chapter 3. Technical specifications of the final OSSMM release, including the open-source repository and build instructions, are presented in Chapter 4. Chapter 5 evaluates the final prototype against a silver-standard reference monitor, demonstrating the system’s robustness over a multi-week experiment and its ability to capture signals necessary for sleep stage classification. Chapter 6 further investigates these captured signals using spectrograms and deep learning techniques. Finally, Chapter 7 concludes the thesis, summarising its contributions, proposing directions for future work, and reflecting on the scope and limitations of the project.

1.12 Ethical Considerations

The technical evaluation completed as part of this thesis required ethics approval. This approval was given by Maynooth University’s Biomedical & Life Sciences Research Ethics Subcommittee. A copy of this approval is found in Section F.2 of Appendix F.

Initially, a more extensive study involving system evaluation on an awake cohort was planned, and ethics approval for this was obtained (Section F.1). However, time and resource constraints necessitated a narrower scope, resulting in the single-participant technical evaluation presented in this thesis, where the researcher served as the participant. The inclusion of this additional ethics approval demonstrates the originally intended scale of experimental validation.

REM Sleep Reduction and Depression: A Deep Dive Review

“Wrong hypothesis, rightly worked from, have produced more useful results than unguided observation.”

— Augustus De Morgan, 1872

Abstract

More than twenty years have passed since the last human study investigating REM Sleep Reduction (REM-R) as a treatment for depression. Despite promising findings, no modern scaled replication has been conducted. Here I review the only three controlled REM-R experiments for depressive symptoms, examine an NREM-focused intervention, and consider pharmacological associations between REM sleep and depression. Despite abandonment of this line of research following claims that REM-R lacks specificity as a treatment for depressive symptoms, this dismissal lacks sufficient support. Rather, reanalysis of the existing research identifies methodological concerns and converging evidence supports reinvestigation into REM-R as a potential treatment. Additional material is presented in Supplementary Appendix [A](#).

2.1 Introduction

Chapter 1 established the dual challenge confronted in this thesis. First, the abandonment of REM sleep reduction (REM-R) as a potential therapeutic intervention despite suggestive findings of clinical benefit, and second, the prohibitive costs that restrict access to quality sleep monitoring and modulation technology. This chapter addresses the first research objective, evaluating whether the dismissal of REM-R research as a treatment was justified, or whether methodological limitations and incomplete analyses may have led to premature conclusions. A careful re-examination of the existing evidence argues for renewed investigation. This in turn further supports the work towards democratizing sleep research developed in subsequent chapters.

2.1.1 REM Sleep Experiments

Although observations of REM sleep were recorded as early as 1936, it would not be until Aserinsky and Kleitman’s landmark study in 1953 that REM sleep was identified as a reoccurring, natural sleep state ([Aserinsky and Kleitman, 1953](#); [Derbyshire et al., 1936](#)). Despite decades of experimental manipulation, our understanding of REM sleep has remained largely observational, though this has begun to change with techniques like optogenetics ([Boyce et al., 2016](#); [Gottesmann, 2001](#)). The discovery of REM sleep led to a flurry of non-pharmacological REM sleep manipulation experiments in the following decades, the first decade being described as the Golden Era of REM sleep discoveries ([Gottesmann, 2001](#)). A variety of terms have been used to describe these manipulations and methods. For clarity, the following definitions are used throughout this review: REM sleep reduction (REM-R) refers to selective restriction of REM sleep. REM sleep deprivation (REM-D) is a more severe subset of REM-R, defined as a reduction of 50% or more compared with baseline levels. Concerning intervention methods, REM sleep disruption is any form of perturbation to sleep during a REM period. REM sleep awakenings (REM-A) are a specific type of disruption where the subject is woken from REM sleep. Unless otherwise specified, all REM sleep manipulations referred to are non-pharmacological.

Contrary to standard practice, REM-R experiments were carried out in humans before they were in animals ([Dement, 1958, 1960](#); [Siegel and Gordon, 1965](#)). Though the earliest REM-D studies in humans indicated harmful psychological consequences, they suffered from significant methodological flaws and were self-admittedly influenced by the authors’ preconceptions ([Dement, 1960, 1965a](#); [Dement and Fisher, 1963](#); [Fisher, 1965](#)). Despite initial concerns, further REM-R investigations continued in both animals and humans with more rigorous measures, and subsequent appropriately controlled studies have consistently failed to replicate any significant

psychological disturbance or physical malaise. While there are occasional reports of treatable anxiety and insomnia in some individuals, REM-R through REM-A has been found safe even in vulnerable patients including depressives and schizophrenics (Cartwright et al., 2003; Dement, 1965b; Grözinger et al., 2002; Landsness et al., 2011; Vogel and Traub, 1968b; Vogel et al., 1968, 1975).

Beyond establishing safety, these studies revealed an unexpected therapeutic benefit. Several studies have demonstrated REM-R to have an antidepressant effect (Cartwright et al., 2003; Dement, 1965b; Grözinger et al., 2002; Steiner and Ellman, 1972; Vogel et al., 1968, 1975).

2.1.2 REM Reduction and Alleviation of Depressive Symptoms — Early Observations

This effect was first observed in humans during a pilot study in 1968 by Vogel et al. (Vogel et al., 1968). The authors were inspired by several earlier observations. REM sleep deprived cats displayed increased activity levels, and both typical pharmacological antidepressants (i.e., Tricyclic antidepressants (TCA) and Monoamine Oxidase Inhibitors (MAOI)) and electroconvulsive therapy (ECT) deprive REM sleep. Of the five psychotically depressed patients who underwent REM-D for 7–14 days, two showed clear signs of improvement and one showed moderate signs. Importantly, improvement was generally in accordance with reactivity to REM-D. Patients who experienced greater amelioration also demonstrated elevated REM pressure, while those who had negligible improvement exhibited minimal if any REM pressure.

REM pressure is the homeostatic drive for REM sleep that increases during and after REM-R. This elevated pressure manifests as REM sleep rebound, a compensatory increase in REM sleep duration and percentage of Total Sleep Time (TST) on recovery nights (nights post-manipulation) compared with baseline (nights pre-manipulation). Additional markers of elevated REM pressure include a shortened REM latency (earlier onset of the first REM period), increased number of REM cycles, and elevated awakening threshold during REM sleep compared to baseline.

Results from the first pilot study were replicated by a second, non-controlled pilot study with 4 additional patients who underwent REM-D for 10–35 days (Vogel and Traub, 1968a).

However, the specificity of REM-R as a treatment for depressive symptoms remains uncertain. Total Sleep Deprivation (TSD) produces rapid but transient antidepressant effects (Borbély and Wirz-Justice, 1982; Giedke and Schwärzler, 2002; Hu et al., 2021; Minkel et al., 2017; Moretto and Palagini, 2019; Vogel et al., 1975; Wirz-Justice

and Van den Hoofdakker, 1999), operating on a different timescale than selective stage interventions which require repeated application over days to weeks, and whose effects are persistent beyond recovery sleep periods. Evidence that NREM sleep reduction (NREM-R) can also produce mood improvements led researchers to question whether REM-R has specific therapeutic properties or whether any form of sleep disruption produces similar effects (Giedke and Schwärzler, 2002; Grözing et al., 2002; Palagini et al., 2013; Vogel et al., 1975). Following this reasoning, REM-R research was largely abandoned (Giedke and Schwärzler, 2002; Palagini et al., 2013). This abandonment may have been premature.

2.1.3 Review Scope and Approach

To the author’s knowledge, only three human studies with some form of control group have been completed to assess REM-R as a treatment for depressive symptoms (DS). The arduous experimental requirements for such sleep manipulations are one of the main reasons they have been attempted so rarely. In the last decade there has been a resurgence of interest in the REM-R and depression link, with several reviews published (Crişan et al., 2023; Lia, 2019; Palagini et al., 2013). However, these reviews have overlooked some important distinctions between the three controlled studies. The question of specificity, whether REM-R produces distinct antidepressant effects compared to other forms of sleep deprivation, remains unresolved. Understanding this is critical for both theoretical knowledge and development of practical interventions.

The three studies were identified through a targeted narrative search rather than a formal systematic review. Vogel et al. (1975) was first identified through general engagement with the sleep literature as the foundational controlled trial of REM-R for depressive symptoms. PubMed and Google Scholar were then searched using combinations of terms such as “REM deprivation and depression” and “REM reduction depressive symptoms” to locate any subsequent replications. Inclusion required: (i) human participants, (ii) some form of control group or control condition, and (iii) non-pharmacological REM-R as the primary intervention. Grözing et al. (2002) represents a partial exception to the third criterion, as a subset of participants remained on antidepressant medication during the trial; it is retained here as the only direct replication of Vogel’s design. Animal studies, total sleep deprivation studies, pharmacological REM-suppression studies, and pilot studies lacking a control group were excluded from the core analysis, though they are referenced where relevant for context. This process yielded the three controlled human trials examined in detail below: Vogel et al. (1975), Grözing et al. (2002), and Cartwright et al. (2003) (Cartwright et al., 2003; Grözing et al., 2002; Vogel et al., 1975). No additional

controlled human trials meeting these criteria have been identified in the recent or older review literature, consistent with the claim that this set is exhaustive.

This review evaluates whether the dismissal of REM-R was justified and addresses the unresolved specificity question through three complementary approaches. First, with a critical analysis of the controlled REM-R trials: two Randomized Control Trials (RCT) utilizing REM-R via awakenings: Vogel (1975) and Grözinger (2002), and one Non-Randomized Controlled Trial (NRCT) of REM-R via awakenings: Cartwright (2003) (Cartwright et al., 2003; Grözinger et al., 2002; Vogel et al., 1975). For each, previous criticisms are considered and additional concerns are noted. This section also highlights confounds and findings which have yet to be further explored. A cross-study perspective shows what these three studies jointly reveal about REM-R as a potential therapeutic intervention.

Second, the challenge of establishing specificity is examined through Landsness et al. (2011), an NREM intervention that unexpectedly revealed correlations between REM suppression and mood improvement, highlighting how the interdependent nature of sleep architecture complicates efforts to isolate stage-specific effects (Landsness et al., 2011).

Third, pharmacological evidence is reviewed, showing that multiple classes of traditional antidepressants consistently suppress REM sleep (Beitinger and Fulda, 2010; Staner et al., 2010; Vogel et al., 1990). This convergence across behavioral and pharmacological interventions suggests REM suppression may be mechanistically relevant to antidepressant action.

Finally, insights from adjacent research in healthy controls are examined, revealing that REM-R increases emotional reactivity and cortical excitability, offering mechanistic clues about how REM suppression might influence mood. The chapter concludes by discussing responder heterogeneity, the need for biomarkers to identify depression subgroups, and the methodological barriers that have prevented systematic investigation. These barriers motivate the open-source platform development described in subsequent chapters.

2.2 REM Reduction and Depression — A Deep Dive

2.2.1 Vogel et al., 1975 — Endogenous Depression Cohort

In order to determine if REM-D was responsible for the antidepressant effects observed in the earlier pilot experiments, Vogel et al. (1975) completed a double-blind,

crossover study comparing the effectiveness of REM-D and NREM sleep interruptions (NREM-I) as treatment methods (Vogel et al., 1975). Published in 1975, the three-year study remains the largest human REM-R experiment yet completed. Beginning with 89 hospitalized patients, and ending with 52 who would complete the study in its entirety, a staggering 4,620 polysomnographs (PSG) were recorded, most of which required active observation (Vogel, 1975; Vogel et al., 1975). Except where specified, this review focuses on the completers. Given the size and complexity of this study, the two depression cohorts are examined in separate subsections. This subsection addresses the endogenous cohort ($N = 34$); the reactive cohort is examined in Section 2.2.2.

Study Design and Participants. A participant history flowchart of the study as described in Vogel et al. (1975) is shown in Figure 2.1. Participants were classified as having endogenous ($N = 34$), “internally caused,” or reactive ($N = 18$), “situationally caused,” depression by two independent clinicians. These two cohorts were further split, and participants randomly assigned to the experimental (REM-D) or control condition (NREM-I) such that there were 4 groups overall: reactive control, reactive experimental, endogenous control, and endogenous experimental. After a minimum 7-day baseline observation, participants entered a three-week pre-crossover where groups experienced their assigned treatment. After this time, participants in the NREM-I control groups were crossed over to the REM-D experimental condition if they had not sufficiently recovered.

Patients were screened for moderate-severe depression and no other psychological or neurological conditions. They were free of any antidepressant and psychotropic medication, except some specified instances of chloral hydrate for insomnia, and chlorpromazine and phenobarbital for agitation. Dosages are noted, but not which patients received medication and with what frequency. Patients had their own private room in the inpatient unit. Napping was not permitted. At baseline, there were no significant differences between experimental and control groups within either depression type (endogenous or reactive) for age, sex, previous depression, duration of current depression, scores on the five measures used to evaluate depression, percentage of total sleep time spent in REM sleep (REM %), or TST.

Procedure. Participants in the experimental condition were REM deprived via awakenings over an intercom at the start of each REM period, and were required to sustain wakefulness for 3 minutes as monitored through EEG measurements. Patients were REM deprived each night until they either experienced 30 forced awakenings in one night, or 6 consecutive nights of REM-D. Once either of these limits was reached, the following night would be a “recovery night” of uninterrupted

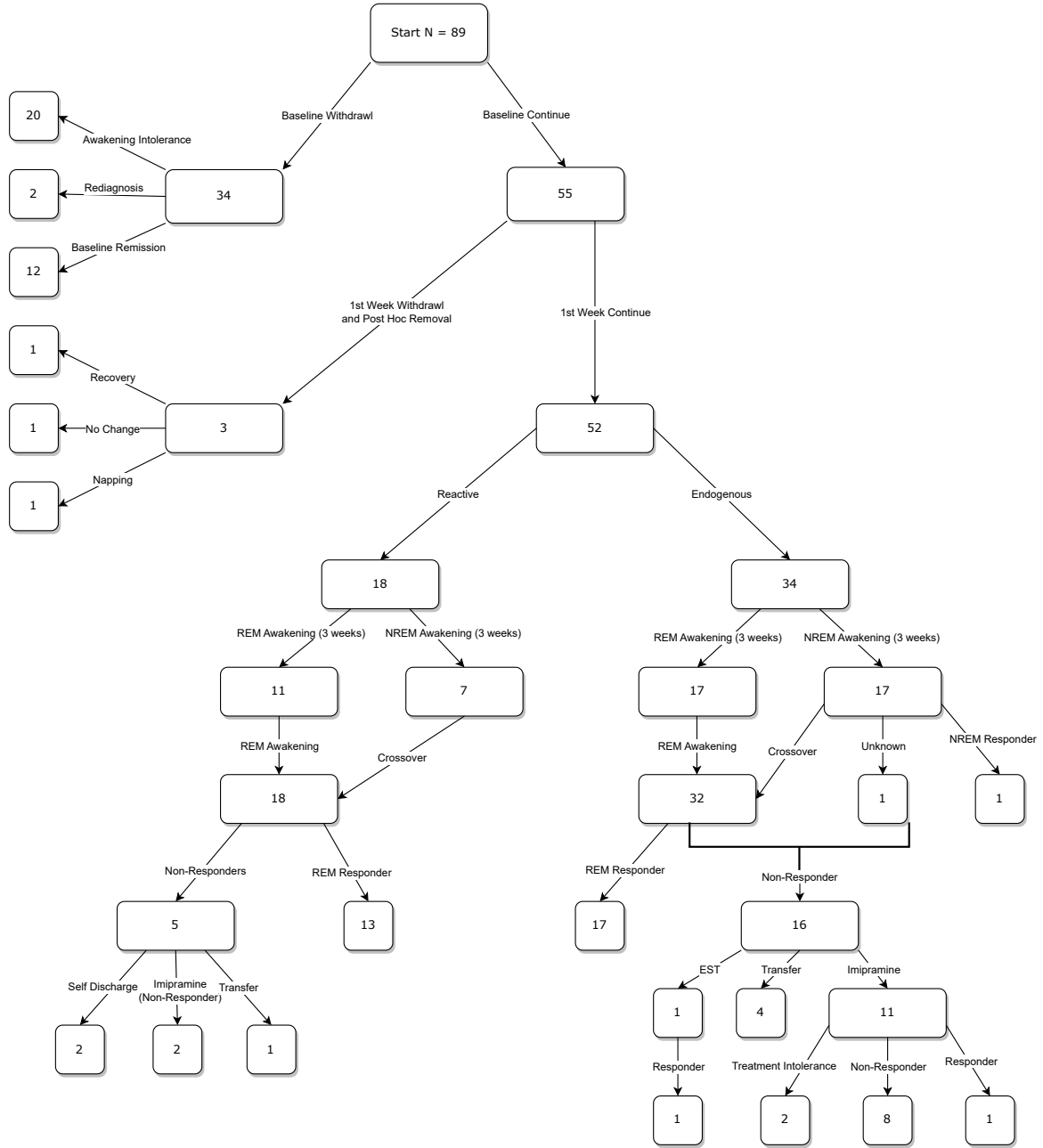


Figure 2.1: Progression history of participants in Vogel et al. 1975. Here ‘Non-Responder’ refers to the directly preceding treatment. ‘Treatment Intolerance’ is any form of withdrawal, including choice, from the preceding treatment. Not shown are further treatment outcomes involving ECT in the endogenous depression cohort. Division of patients into reactive and endogenous categories may not be chronologically accurate because it is not noted from which group the first week dropouts belong.

sleep. This procedure would cycle until either the patient sufficiently recovered, chose to withdraw, or nine weeks post baseline passed. Controls were matched with a REM participant in order to equate both the number of NREM and REM awakenings and recovery nights between groups. Importantly, the timing of the awakenings was yoked to ensure a similar distribution between groups. Control participants were awakened 10 minutes into an NREM sleep period preceded by a REM sleep period. If participants did not improve sufficiently in 9 weeks, they were typically either transferred or treated with imipramine hydrochloride, a TCA, and/or ECT.

Outcome Measures. Outcome measures for the study included both sleep and depression scores. Patients were monitored nightly by PSG. Depressive symptoms were assessed by five different measures, three times a week. Three were symptom rating scales: the Hamilton Depression Rating (HAMD-17), 7-point Global disability scale, and the Zung Self-Rating Depression scale. Two assessed psychomotor speed: the Letter Cancellation Test (LCT) and a subsection of the Wechsler Adult Intelligence Scale (WAIS).

Endogenous Group Results: Sleep and Depression Outcomes Pre-Crossover.

Outcomes for the endogenous cohort can be summarized as follows. During the control period, the experimental group experienced a significant reduction in REM % ($P < 0.001$, two-tailed paired t -test), while the control group experienced a significant increase ($P < 0.01$, two-tailed paired t -test). Respectively, this was a 63.2% reduction and a 14.1% increase. Any percentage changes displayed in this review are to indicate the direction and amplitude of changes, and do not concern percent change as a statistical measure (Vickers, 2001). Table 2.1 summarizes the endogenous cohort's longitudinal results on the HAMD-17. The control group did not significantly improve during this period, but the statistical test only accounted for 15 of the 17 control patients. One control patient recovered sufficiently for discharge, and the other control is not accounted for. This last control participant may be the individual who is noted as prematurely crossing over to the REM-D condition after 13 days.

The original text included individual Hamilton and Global Scores, at baseline and crossover for the endogenous cohort, making it possible to reanalyze the data. Therefore this review verified that there was no significant improvement when considering all 17 controls ($p > 0.10$, two-tailed paired t -test). During the pre-crossover period, the original text does not report whether the experimental group significantly improved, but a reanalysis of the data shows significant improvement from baseline ($p < 0.001$, two-tailed paired t -test). Between-group testing at the end of the pre-crossover period showed mixed results. Vogel et al. (1975) reported that parametric testing showed that the experimental group improved more than the control on both the HAMD-17 ($p = 0.029$, one-tailed ANCOVA test) and Global ($p = 0.018$, one-tailed ANCOVA test). Non-parametric Mann-Whitney U -test showed significance on the Global Score ($p \leq 0.025$), but not on the HAMD-17 ($p = 0.09$).

Endogenous Group Results: Sleep and Depression Outcomes Post-Crossover.

Three weeks after crossover the control group was significantly less depressed than at crossover time on both the Hamilton ($P < 0.007$, two-tailed paired t -test) and Global ($P < 0.04$, two-tailed paired t -test). The rate of improvement before crossover was similar to the rate after crossover ($P > 0.10$). By the end of the study 17 endogenous (50%) patients recovered sufficiently for discharge. Of these remitters, eight were from the control group and nine from the experimental group. 11 participants who did not recover received imipramine hydrochloride (TCA) treatment. Two withdrew from this treatment, and of the remaining nine participants who completed 4 weeks of medication, only one responded, an atypically low response rate (Stewart et al., 1993).

Methodological Considerations. This study has several limitations and concerns, some of which have been discussed in previous literature. The most significant for the endogenous cohort are addressed here; concerns specific to the reactive cohort are addressed in Section 2.2.2.

The Total Sleep Time Confound. Giedke & Schwärzler (2002) questioned the results of Vogel et al. (1975) for several reasons (Giedke and Schwärzler, 2002). Although the authors misinterpreted some of the results and methodology, the most important point to address is their observation of the TST differences between treatment groups. REM-D patients slept approximately 40 minutes less than controls during the pre-crossover period in both cohorts. Since the antidepressant effects of sleep deprivation can be attributed to the amount of sleep reduction, this is a crucial confounding factor to address. In this case the concern is unfounded because the endogenous cohort, which showed differences in depression improvement at crossover, already exhibited the same approximate 40 minute difference in TST at baseline.

Table 2.1: Outcomes Comparison of Endogenous Participants on HAMD-17 Scores to Week 6

	Baseline	Control Period	Crossover	Post Crossover	Week 6 End
Control		No Improvement ¹ ($P > 0.10$)		Improvement ($P < 0.007$)	
Experimental		Improvement ² ($P < 0.001$)		?	
	$\underbrace{\hspace{1cm}}$ n		$\underbrace{\hspace{1cm}}$ */n ³		$\underbrace{\hspace{1cm}}$?

Note: Rows show within group changes; columns show between group differences. "n" = non-significant; "?" = not reported in the original text. Between group tests: Baseline, Crossover (Week 3), End (Week 6) are one-tailed ANCOVA; Within group tests: Control Period (Weeks 1–3), Post Crossover (Weeks 4–6) are two-tailed paired t-tests. Constructed by the present author from data reported in Vogel et al. (1975).

¹This measure only accounts for 15/17 of the control participants in the original text, however reanalysis of all 17 controls also resulted in non-significant change.

²Significance was calculated by two-tailed *t*-test with data from Table 3 in the original study

³Parametric and non-parametric tests disagreed; see body text for p-values and reanalysis.

Explicitly, the endogenous experimentals slept 43.4 minutes less than controls before any experimental procedure, therefore one can not conclude that this difference necessarily confounds the results.

Statistical Approach: Parametric vs. Non-parametric Testing. Another raised concern is that the significance level of outcome scores at the end of pre-crossover period between treatment groups for the endogenous cohort differed based on parametric (ANCOVA) or non-parametric testing (Mann-Whitney U). This calls into question if the ANCOVA was an appropriate statistical test. Of the requirements necessary for an ANCOVA, two were not addressed in the original study: normality and homogeneity of variance. Using the data provided in the original text, both the Shapiro-Wilk and D’Agostino-Pearson tests were applied to test for normality, and Levene’s test for homogeneity of variance. None of the tests were significant (see Appendix A, Section A.1), therefore satisfying the requirements for the use of ANCOVA, which when its assumptions are met, is known to be a more powerful and sensitive method than non-parametric alternatives (Knoke, 1991). Both the covariate (baseline scores, $p = 0.002$) and the group effect ($p = 0.03$, one-tailed) were significant in the ANCOVA model. Although baseline depression scores did not differ significantly between groups ($p=0.62$), the experimental group’s slightly higher baseline scores (49.9 vs 47.8) were statistically controlled for by ANCOVA. As a third independent check, one-tailed permutation testing was also conducted

and returned non-significance ($p=0.08$), consistent with the Mann-Whitney result.

The REM Pressure Claim. This review notes, as did Giedke & Schwärzler (2002), that Vogel et al. (1975) falsely concluded that, “The results then supported the hypothesis that REM sleep deprivation with increased REM pressure has antidepressant activity in endogenous but not reactive depression.” This is not what the presented data and analyses show. First, no clear association between a given group’s or individual’s change in REM pressure (i.e., increase in REM sleep periods, recovery REM rebound, etc.) due solely to REM-D and a change in depressive symptoms is presented. Only group means for sleep variables during the pre-crossover period are presented, but no analyses are shown which link these sleep outcomes to depression scores.

This first problem is partly addressed in later articles by Vogel et al., which found positive correlations for percentage change in REM pressure (defined as REM % on 1st Recovery Night / REM % on Last Two Baseline Nights) and improvement in depression scores. Hamilton and Global scale improvement correlated with REM pressure at $\rho = 0.69$ ($p < 0.003$) and $\rho = 0.75$ ($p < 0.001$) respectively. However, this analysis included only 14 of the 34 endogenous patients (Vogel et al., 1977). Further analysis of these same 14 patients revealed more specific predictors of improvement (Vogel et al., 1980).

Recovery night late REM % (the proportion of REM sleep occurring after the first REM period) correlated more strongly with improvement than total REM rebound: $\rho = 0.71$ (Hamilton, $p < 0.01$) and $\rho = 0.81$ (Global, $p < 0.001$). When late REM % was held constant, the partial correlation between total REM % increase and improvement became non-significant, indicating that the extension of late REM sleep, rather than total REM rebound, was the specific correlate of clinical improvement. Additionally, this study found that improvement correlated with the rate at which late REM % normalized across successive recovery nights, with stronger responders showing faster increases in late-night REM sleep. These findings also revealed that the same depressed patients exhibited abnormal temporal distribution of REM sleep at baseline, with disproportionately long first REM periods and shortened later REM periods, a pattern that was exaggerated after REM-D.

Summary. Despite methodological limitations and incomplete reporting, several patterns emerge for the endogenous cohort. REM-D produced clinically meaningful improvements, with 50% of endogenous patients recovering sufficiently for discharge, on par with antidepressant treatment (Stewart et al., 1993). The crossover design revealed a notable pattern: endogenous patients showed no improvement under NREM-I, but significant improvement after switching to REM-D. Critical questions

remain unanswered. Incomplete reporting of experimental group outcomes at the end of the study prevents direct comparison of REM-D versus NREM-I efficacy, and the relevance of REM pressure is unclear in the main 1975 report, though partly addressed in Vogel's later 1977 and 1980 follow-up analyses.

2.2.2 Vogel et al., 1975 — Reactive Depression Cohort

The study design, procedure, and outcome measures for the reactive cohort were identical to those of the endogenous cohort described above in Section 2.2.1. This subsection focuses on outcomes for the reactive cohort ($N = 18$) and methodological considerations specific to it.

Reactive Group Results: Sleep and Depression Outcomes Pre-Crossover.

Outcomes for the reactive cohort ($N = 18$) can be summarized as follows. During the control period, the experimental REM-D condition ($N = 11$) experienced a significant reduction in REM % ($P < 0.001$, two-tailed paired t -test), and the NREM-I controls ($N = 7$) did not. This represented a 64.3% reduction in REM % for the experimental group and an 11.8% increase for controls. Table 2.2 summarizes the reactive cohort's longitudinal results on the HAMD-17. The control group improved significantly on both the Hamilton ($P < 0.001$, two-tailed paired t -test) and Global ($P < 0.01$, two-tailed paired t -test) depression scores. It is not noted if this improvement was mirrored in the experimental group. However, at the time of crossover no significant differences were found between groups, and no participants sufficiently recovered for discharge.

Reactive Group Results: Sleep and Depression Outcomes Post-Crossover.

All seven controls were then crossed over to the REM-D treatment. Three weeks after crossover (Week 6), the control group was significantly less depressed than at the crossover period on both the Hamilton ($P = 0.0005$, two-tailed paired t -test) and Global ($P = 0.005$, two-tailed paired t -test). The rate of improvement was similar before and after crossover ($P > 0.40$). No information is provided about the experimental group at three weeks post-crossover or about between-group differences at Week 6. By the end of the study, 13 (72%) of the reactive patients recovered sufficiently for discharge. It is not noted how many remitters (i.e., those who recovered for discharge) belonged to the experimental or control group. Table 2.2 highlights the lack of information for the reactive experimental group, and for a final comparison between the controls and experimentals.

Table 2.2: Outcomes Comparison of Reactive Participants on HAM-D-17 Scores from Baseline to Week 6

	Baseline	Control Period	Crossover	Post Crossover	Week 6 End
Control		Improvement ($P < 0.001$)		Improvement ($P < 0.005$)	
Experimental		?		?	
	$\underbrace{\hspace{1.5cm}}$ n		$\underbrace{\hspace{1.5cm}}$ n		$\underbrace{\hspace{1.5cm}}$?

Note: Rows show within group changes; columns show between group differences. "n" = non-significant; "?" = not reported in the original text. Between group tests: Baseline, Crossover (Week 3), End (Week 6) are one-tailed ANCOVA; Within group tests: Control Period (Weeks 1–3), Post Crossover (Weeks 4–6) are two-tailed paired t-tests. Constructed by the present author from data reported in Vogel et al. (1975).

Methodological Considerations. One concern is particularly pertinent to the reactive cohort and is addressed here.

Selective Reporting and Study Modifications. One final concern to highlight might be called “reverse p-hacking.” Vogel et al. (1975) state that upon observing no difference in REM and NREM deprivation for reactive patients, they “preferred not to admit reactive depressive patients to the project.” The lack of reported statistics for the experimental reactive group’s improvement leaves the reasoning open to interpretation. However, Vogel et al. (1975) stated this within the context of the cost and difficulty of running further participants who may not show differences in results.

Summary. For the reactive cohort, REM-D and NREM-I produced similar effects. 72% of reactive patients recovered sufficiently for discharge, and the rate of improvement did not differ meaningfully before and after crossover. The reactive cohort’s higher response rate and faster recovery (mean 5.6 weeks of REM-D procedures, compared with 7.3 weeks for the endogenous cohort) challenges the original authors’ conclusion that REM-D has antidepressant effects only for endogenous depression. The mid-study decision to halt recruitment of reactive patients constitutes a form of selection bias that limits the strength of conclusions drawn from this cohort.

Cross-Cohort Synthesis. Taken together, the two cohorts complicate a simple reading of Vogel et al.’s findings. The endogenous cohort provides the study’s strongest case for a REM-specific effect, with the crossover design showing improvement only after REM-D. The reactive cohort, which the authors de-emphasized, shows that NREM-I produced outcomes comparable to REM-D. Based on the data and analyses presented, the results support the following hypotheses: for reactive

depression, both REM and NREM deprivation are effective, although the association between increased REM pressure and this effect remains unknown; for endogenous depression, REM-D is most effective, and symptom improvement correlates with increased REM pressure. The scale of this study, some 4,620 polysomnographs across three years, has not been replicated, leaving many of these questions unresolved.

2.2.3 Grözinger et al., 2002

In 2002 Grözinger et al. published the first and only replication of Vogel et al. (1975), although there were important differences between the studies (Grözinger et al., 2002). The replication was a 10-day, double-blind, parallel RCT where the experimental condition was REM-R, and the control condition NREM-I. The study started with 37 hospitalized, depressed patients, and finished with 21 completers.

Study Design and Participants. Patients were rigorously examined for entry to the study. They were vetted by two psychiatrists to ensure they had no other psychiatric or neurologic disorders and qualified as having unipolar depression (per the Diagnostic and Statistical Manual of Mental Disorders, 4th ed.; DSM-IV). Medical histories, blood tests, EKG and EEG were completed for each participant, and regular benzodiazepine use was exclusionary. Sleep disorders including sleep apnea and restless leg movements were also exclusionary. Unlike Vogel et al. (1975), patients were allowed antidepressant medication if they had taken the medication without improvement at a constant dosage for more than 2 weeks prior to the study.

Figure 2.2 shows a participant progression history. Participants were randomly assigned to the control or experimental condition. After the first treatment night, ten participants withdrew, five from each condition. The remaining participants were described as the “intent-to-treat”, with ($N = 13$) controls and ($N = 14$) experimentals. Throughout the study six more participants withdrew, three from each condition. Therefore ($N = 10$) controls completed the study as did ($N = 11$) experimental participants. To comply with ethics requirements and address insomnia, trimipramine (a TCA) was offered to all participants. Patients already taking trimipramine for more than two weeks continued unchanged. Those on other antidepressants for more than two weeks switched to trimipramine (minimum 50 mg daily). Drug-free participants or those off medication for more than two weeks could elect to begin trimipramine (Grözinger et al., 2002). Four patients in the experimental group and six controls were on antidepressants at the “intent-to-treat” stage. Four participants in the REM-D condition received lorazepam or chloral hydrate for anxiety, and two controls received lorazepam. Of the “intent-to-treat” participants who were on antidepressants and/or anti-anxiety medication, it is not noted which finished the study and which were partial-completers.

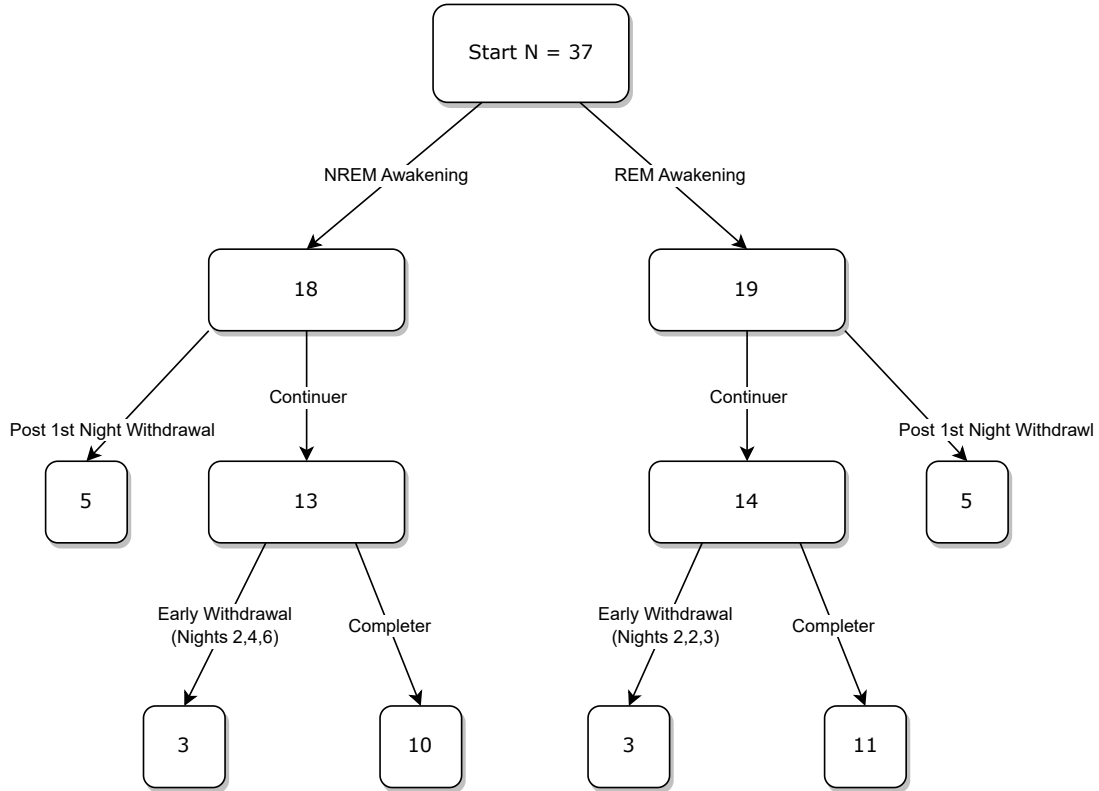


Figure 2.2: Participant progression in Grözinger (2002). Patients were randomly assigned to REM/NREM groups prior to the 1st treatment night

Procedure. Patients’ sleep was actively monitored through a single EEG channel in combination with an Artificial Neural Network (ANN) which had been trained and validated to recognize REM or NREM sleep. Previous research showed the ANN had a classification accuracy of 90% by epoch between NREM and REM detection (Grözinger et al., 1997). Participants in the experimental condition underwent REM-R via awakenings at the detection of each REM period via a strong light source and an alarm. After pressing a button to cancel the alarm, they were allowed to return to sleep. Control patients were awakened in the same manner, but during NREM sleep. Importantly, awakenings in NREM sleep were equally distributed among NREM stages II, III, and IV, unlike in Vogel et al. (1975). The number of control awakenings was generally matched to REM awakenings. All patients underwent one adaptation night, followed by 10 consecutive treatment nights. There were no recovery nights, and napping was not permitted.

Outcome Measures. Outcomes for the study included both sleep and depression scores. Complete PSG recordings were obtained each night and the records later scored by two independent raters. Depression was measured through Hamilton De-

pression Scale 21 (HAMD-21), HAMD-17 (for comparison with Vogel et al. (1975)), Bech-Rafaelson-Melancholia Scale (BRMS), and Clinical Global Impressions (CGI). Tests were administered on mornings of the 1st day (pre-treatment), 6th day, and 11th day (post last treatment night). Subjects also rated themselves daily on the Befindlichkeits Skala (Bfs), a self-report inventory, and rated their tiredness upon waking and their tiredness from the previous day

Results. Baseline sleep records were not taken, preventing within-group comparisons. However, REM-R patients showed approximately 45% less REM sleep than NREM-I controls. Additionally, the REM-R group’s absolute REM sleep values were substantially below normative levels reported for healthy adults, supporting an estimated 45% reduction from expected baseline (Floyd et al., 2007). The experimental participants also exhibited a significant decrease in REM latency from the first treatment night to the eighth treatment night ($p < 0.001$, two sided paired t -test), nearly halving the time of REM sleep onset. NREM-I patients experienced an increase in REM % throughout the experiment. On treatment days 1–2 controls’ REM % constituted 18.4% of their TST, 21.5% on days 5–6, and 23.4% on days 9–10. Changes were also observed in the duration of the first three sleep cycles. Participants in the experimental condition displayed shortened NREM-REM sleep cycles, whereas controls displayed lengthened ones. Without a baseline comparison, it is not clear if this difference is due to treatment or pre-existing conditions. However, similar shortening of NREM-REM rhythms has been observed in depression and high-risk groups (Lauer et al., 1995).

Over the course of the study both groups displayed a decline in depressive scores on the HAMD-21 ($p < 0.001$, paired t -test), HAMD-17 ($p < 0.001$, paired t -test), and BRMS ($p < 0.001$, paired t -test). Only controls had a significant improvement on the CGI ($p < 0.001$, paired t -test), though the REM-D condition neared significance ($p = 0.07$). There was no change on the Bfs. On the last day of measurement there was no significant difference between control or experimental participants on any of the scales. This was true for completers and for the “intent-to-treat” group with partial completers.

Although omitted from the main publication, an earlier abstract reported a detectable effect several days after treatment. Depression scores were collected five days after the last treatment night, and showed that NREM-I controls had significantly improved compared with experimental participants (Grözinger et al., 1999). It is not known (even to the main author himself) why this result was not included in the main paper later (Grözinger, 2023). Despite raising further questions, this result was later cited as evidence that REM-D is not an effective antidepressant.

Methodological Considerations. There are several points of consideration, but the most relevant are highlighted here.

Antidepressant Usage. For ethical reasons, participants in both conditions were permitted antidepressants. Trimipramine was chosen as the antidepressant if participants elected to take medication because earlier studies had shown it to be one of the few antidepressants to not affect REM %. This has been supported by several studies (Feuillade et al., 1992; Holsboer-Trachsler et al., 1994; Ware et al., 1989). However, one study found a significant increase in REM % for depressed patients, and another found in healthy controls that a single dose of trimipramine will not affect % REM sleep during treatment, but upon withdrawal will significantly increase REM sleep, suggesting an increase of REM pressure (Feuillade et al., 1992; Sonntag et al., 1996). Although there are no indications in any of the follow-up analyses that the antidepressants had any effect, as it was present for both groups, it is not clear how it affects the study’s outcome as a whole. The combination of sleep deprivation and antidepressants has had mixed results, and seems partly to depend on the specific medication used (Giedke and Schwärzler, 2002; Leibenluft et al., 1993).

Awakenings and NREM-REM Rhythms. The difference in awakenings distribution between groups is notable. Although the number of awakenings per night for each group was not significantly different (REM-D = 40.0, NREM-I = 40.5, $P=0.94$), the distributions were not equivalent. This is true for the distribution within-night and across nights. In the REM-R condition, both total awakenings and awakenings before 3AM demonstrate an increasing trend over the course of the experiment, while in the control condition they decrease. Awakenings before 3AM in the REM-R condition only reached 20 alarms on one night, whereas in the control condition they attained this level seven out of ten nights. In other words, at the beginning of the experiment controls were aroused nearly two fold more often, while in the later half of the study, participants in the REM-R group were aroused more often. Notably, controls improved more in the first half of treatment on the HAM-D-21 (decrease from baseline: REM-R = 5.3, NREM-I = 5.7), while the experimental group improved more in the second half (decrease from mid-experiment: REM-R = 3.4, NREM-I = 2.7). Furthermore, the distribution of awakenings through the night was different between groups, with the REM-R group having a clear left skew, while controls had a uniform level of awakenings.

The differential timing of awakenings is theoretically significant because REM sleep follows an ultradian rhythm (i.e. the NREM-REM cycle) with distinct characteristics across the night. REM periods become progressively longer and more intense

in the second half of the night, when REM pressure accumulates most strongly. Vogel et al. (1980) found that depressed patients had an unstable temporal distribution of REM sleep. From night to night, when REM sleep occurred early (long first period), it was reduced later, and vice versa, a pattern not seen in controls. Therapeutic improvement from REM sleep deprivation correlated specifically with increases in late REM % and recovery night late REM %, suggesting that modulating the NREM-REM cycle, particularly the capacity to generate REM sleep in the latter portion of the night, may be therapeutically relevant (Vogel et al., 1980). Returning to the Grözinger study, the observation that the REM-R group’s improvement corresponded to the period when they experienced more awakenings and late REM sleep was preferentially disrupted aligns with the hypothesis that therapeutic effects arise from disrupting REM sleep mechanisms during the period when REM pressure naturally accumulates (Grözinger et al., 2002; Vogel et al., 1980). This raises the possibility that the control condition’s concentration of early awakenings may have been a fundamentally different intervention than the REM-R group’s later awakenings, potentially explaining their different trajectories of improvement despite similar total numbers of awakenings.

Specificity of ANN. Detection of REM sleep was automated through an ANN developed through an iterative process (Grözinger et al., 1997). While the tested accuracy of this algorithm was 90% by epoch, the algorithm struggled with distinguishing REM from Stage I sleep. The algorithm misclassified 26.9% of true REM epochs as NREM (poor sensitivity for REM detection), but more importantly, misclassified 26.8% of Stage I as REM sleep (poor specificity). In other words, with a quarter of Stage I epochs misclassified as REM, the experimental REM-R group was likely awoken during most Stage I sleep periods, as each period consists of multiple epochs. This is particularly problematic because it transforms the experimental manipulation from selective REM-R into a confounded intervention combining REM-R with Stage I sleep fragmentation. No control logic is described in the methods, so it is unclear what protective measures may have been implemented to reduce false positives, if any (e.g., requiring consecutive epochs of REM before triggering).

Advantage of Control Group. Lastly, the author’s claim in the study that controls appeared to have a small but non-significant advantage regarding improvement by the study’s end. This is consistent when comparing the group improvement with baseline among the “intent-to-treat” participants on the HAMD-21 (difference from baseline: REM-D = 7.8, NREM-I = 8.6), but not for the completers where the REM suppressed group improved more from baseline than controls (difference from baseline: REM-R = 8.7, NREM-I = 8.3), contradicting the claim.

Summary. In 2002, Grözing et al. published the first and only replication of Vogel et al. (1975), although there were important differences between the studies (Grözing et al., 2002). This 10-day randomized controlled trial compared REM-R against NREM-I in 21 depressed patients. Unlike Vogel’s findings for endogenous depression, this study which used more modern DSM classifications found that both REM-R and NREM-I groups showed substantial and equivalent improvements in depression scores, with no advantage for REM suppression. Several methodological considerations complicate interpretation of these results, including the use of antidepressants, lower REM sleep reduction, differences in awakening distribution patterns, and specificity issues with the automated REM detection algorithm.

2.2.4 Cartwright et al., 2003

Cartwright et al. (2003) conducted an analogous study to examine the effects of REM-R on morning mood and the long-term effects of low-frequency REM-R administered over several months. Unlike earlier experiments, this study excluded hospitalized patients and applied a gentler REM sleep reduction protocol through reduced treatment intensity, frequency and fewer intervention nights. To the author’s knowledge, this is the last study to directly investigate REM-R and DS (Cartwright et al., 2003).

Study Design and Participants. ($N = 20$) participants were recent, first marriage divorcees who were described as “feeling blue.” The sample was evenly divided by gender (10 men, 10 women). Participants were screened in the context of unipolar depression related to marital discord. To qualify, participants needed a minimum score of 18 on the HAM-D-21, a minimum score of 14 on the Beck Depression Index (BDI), and had to meet the requirements on the Structured Clinical Interview for Non-patients (SCID-NP). Antidepressant usage or psychotherapy was exclusionary. Controls ($N = 10$) were divorcees who did not meet any of the above depression criteria. A preliminary sleep study was conducted to rule out any sleep disorders. All participants within the experiment underwent the same REM-R procedure. Twenty depressed participants completed the study with three additional participants withdrawing during the project. One withdrew because standard treatment intervention was deemed necessary by a psychiatrist, and two others who elected to take medication.

Procedure. After the screening process, treatment and evaluation occurred over a five month period. Participants underwent just three REM-R nights overall. These occurred once in month 1, 2, and 4. Participants were awoken during REM sleep and were asked to recall, “What was going through your mind just before I called you?” This was accomplished by awakening participants five minutes after onset of

the first REM period, and then an additional five minute delay was added for each subsequent REM period. Once the fourth REM period occurred, participants were woken 20 minutes after REM sleep onset, and this 20 minute delay was maintained for future REM periods. Each night of REM reduction was preceded by a “sleep through” night which served as a respective baseline, and did not involve any interruptions. However, Month 1 included an additional preliminary night, serving as an adaptation and diagnostic night.

Outcome Measures. Outcome measures included sleep measures, and both depression and mood scores. Each month participants completed the BDI, and on months 3 and 5 they completed the HAMD-21. A third scale, Profile of Mood States (POMS) was completed before and after each laboratory night. Participants in the experimental group were considered remitted if they no longer met the entry requirements on the HAMD-21, BDI, and SCID-NP.

Results. When all 30 participants are considered, a significant reduction in REM % is observed for all treatment nights compared with respective baselines. But this is not true when considering group conditions. There was no significant difference in REM % for the experimental group on the last treatment night compared to the respective baseline, and the controls never experienced a significant reduction on any of the treatment nights, compared to respective baselines. Although the protocol aimed to reduce REM sleep by 50%, actual deprivation averaged only 16.1% in the experimental group compared with baseline nights, and controls a 12.1% decrease. Similarly, an increase in REM periods was significant when considering all participants, but this was not true of the control group alone, or the experimental group during the second treatment night. The authors separated the experimental group into remitters and non-remitters, and compared depression score differences, but such analyses were not extended to associations with sleep outcomes.

A steady reduction in depression scores was observed in the experimental group. Scores on both HAMD-21 and BDI nearly halved from baseline to outcome (BDI: 28.9 to 15.3, HAMD-21: 25.2 to 14.6). However, no statistical tests were noted comparing baseline and outcome. Controls also experienced a reduction of BDI scores, on a lower scale (7.1 to 3.6), but this was not replicated on the HAMD-21 (5.9 to 4.2). Experimental participants were classified as non-remitters if they failed to improve to below study entry scores or status on any one of the BDI, HAMD-21, or SCID-NP. Remitters ended with a mean BDI score of 6.66, which is considered clinical remission. Non-remitters mildly improved, but remained well above clinical levels of depression at a mean score of 28.12 ([Beck et al., 1996](#)).

The authors found that individuals who experienced diurnal variation (DV), that is,

changes in mood or severity across the day, and felt the worst at night (evening worsening), were the most likely to remit. Significant correlations were found between baseline DV scores based on questions from the HAMD-21 with both remission ($r = 0.47, p < 0.03$) and reduction in POMS scores ($r = 0.46, p < 0.04$).

Methodological Considerations. While there are many considerations to note in the following study, this section focuses on methodological considerations related to REM-R and depressive scores. Additional considerations are provided in Appendix A.

Baseline Differences Between Remitters and Non-Remitters. Baseline differences between remitters and non-remitters reveal a substantial confound. Baseline analysis between remitters and non-remitters found that while there was no significant difference between scores for the HAMD-21, non-remitters had significantly higher BDI scores ($p \leq 0.05$, t -test). More critically, remitters and non-remitters differed substantially in months of separation from their ex-spouse. Remitters had a mean separation of 13.0 months, while non-remitters just 4.18 months. For comparison, controls had a mean separation of 13.2 months. The authors neither report statistical tests for this difference nor adjust for it in outcome models. Given their own citation that 60–70% of depressed individuals will remit within one year after a major stressor, this raises the possibility that time since separation, rather than REM-R, accounts for observed remission effects.

Regression Analysis and Interpretation. A regression analysis was conducted to distinguish remitters from non-remitters with four predictors: BDI baseline score, DV baseline score, Dream-like fantasy score at Month 1, and the sum of total decrease of POMS-D scores on all three REM intervention nights (night–morning). The authors claim that greater overnight improvement (a decrease) in POMS-D on intervention nights predicted remission ($\beta = 0.319, t = 2.28, p = 0.038$). However, Figure 2.3 (originally Figure 1 in Cartwright et al. 2003) shows larger overnight reductions in POMS-D score across the REM interruption nights for non-remitters. If the plotted data are accurate, this appears to contradict the stated result. Either the variable was coded in the opposite direction, or the relationship between REM reduction and remission is the reverse to what is proposed.

Summary. Cartwright et al. (2003) investigated whether minimal, longitudinal REM-R could alleviate depression in a cohort of recent divorcees. Twenty participants underwent just three REM-R nights over four months, averaging a 16.1% of REM sleep reduction per intervention night. Despite this minimal suppression, 60% of participants achieved remission (mean BDI: 28.9 to 6.66).

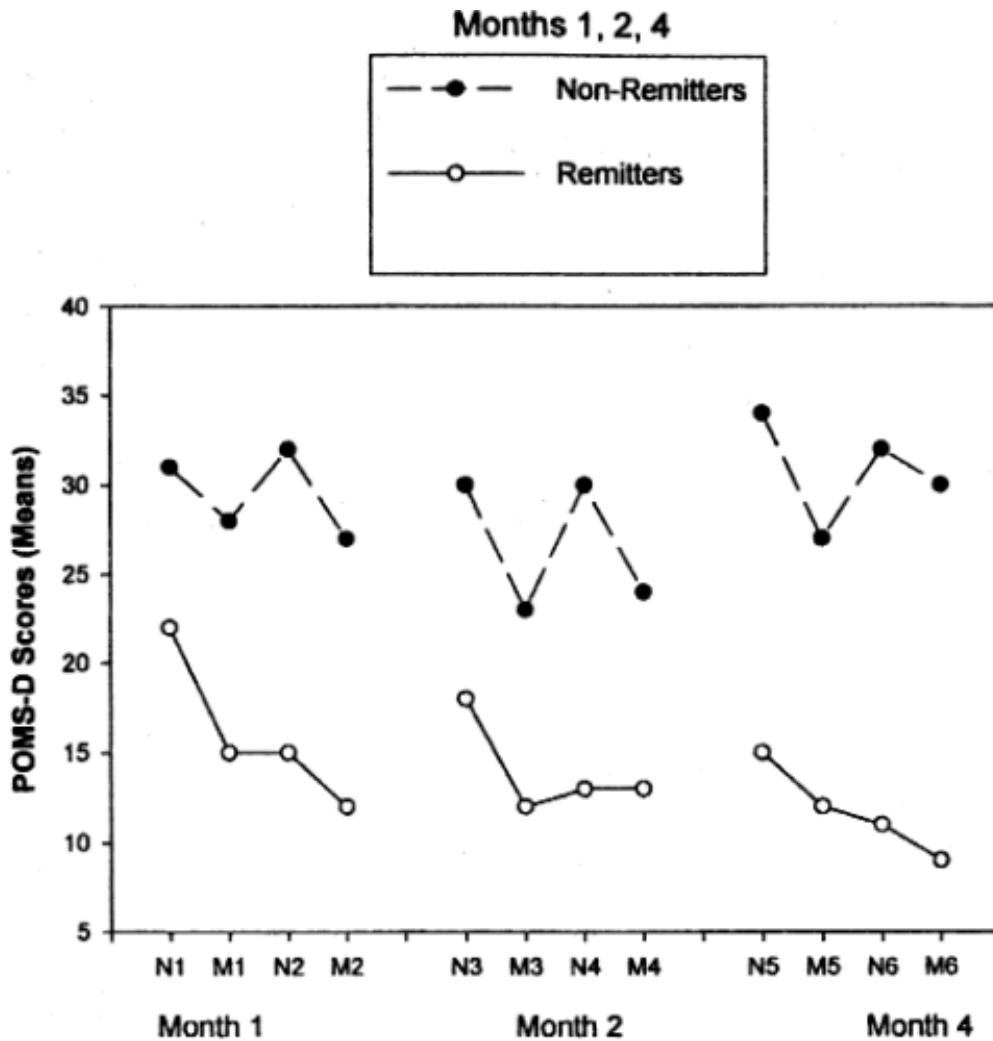


Figure 2.3: Figure 1 in Cartwright et al., 2003. Mean POMS-D score across the study, separated for Non-Remitters and Remitters. Each sleep intervention night was preceded by a baseline night.

Source: Cartwright R., et al., “REM sleep reduction, mood regulation and remission in untreated depression,” *Psychiatry Research* **121**(2), 159–167 (2003), Fig. 1.

The study demonstrates that minimal, spaced REM-R (3 nights over 4 months) is feasible and well-tolerated in outpatient populations with situational depression. However, the duration of separation between remitters and non-remitters is a notable confound that prevents any conclusion to what extent REM-R contributed to the observed remissions. Nevertheless, the study contributes to a broader pattern across all three REM-R trials; that despite variable methodologies and depression subtypes, consistent reports of clinical improvement emerge during REM-R protocols, even when causal attribution remains ambiguous.

2.2.5 Cross Study Synthesis

This synthesis integrates the three human REM-R trials (Vogel, 1975; Grözing, 2002; Cartwright, 2003) to highlight findings not apparent from any one single study. Table 2.3 summarizes the design and measurement details. This section considers areas of convergence and divergence, contextual and methodological differences, the emergent signal, and gaps from this synthesis. The resulting examination reveals consistent physiological responses, heterogeneous response outcomes, and critical methodological gaps that prevent definitive conclusions.

Areas of Convergence. All three studies reported improvement in DS during the study window, though methodological limitations prevent confident attribution of these improvements to REM-R specifically. Improvements were detectable on standardized clinical scales including the HAMD-17, HAMD-21, and CGI.

Safety and Tolerability. Selective REM-R through awakenings proved feasible and generally well-tolerated across all three studies. While some participants experienced adverse effects of anxiety or insomnia, these were transient and treatable, even in vulnerable psychiatric populations. However, dropout rates in the more severe REM reduction methodologies were considerable. Vogel et al. saw a 58.4% completion rate, and Grözing et al. a 56.8% completion rate, whereas Cartwright et al. experienced a 90.9% completion rate.

Physiological Response. All studies documented physiological markers of REM system engagement: shortened REM latency, altered ultradian timing, elevated REM pressure, or indicators of REM pressure. These objective changes confirm genuine sleep architecture perturbation occurred rather than simple placebo effects from awakenings.

Responder Heterogeneity. None of the studies showed uniform effects across all participants, consistently indicating a responder subset within each participant cohort rather than universal treatment effects. Given that diagnostic criteria for depression permit 945 distinct symptom profiles, heterogeneity in treatment response should be expected rather than surprising (Fried and Nesse, 2015a).

Areas of Divergence. The studies diverged substantially across methodological, clinical, and population dimensions.

Population, Diagnosis, and Medication. Each of the studies examined different participant group labels, in part due to changes in diagnostic categories. Vogel et al. classified participants as having endogenous or reactive depression based on clinical criteria by two independent clinicians. Grözing et al. used DSM-IV unipolar

Table 2.3: Summary of Differences Between REM Reduction Studies for Treatment of Depression

Study	Vogel et al. (1975)		Vogel et al. (1975)		Grözinger et al. (2002)		Cartwright et al. (2003)
Diagnosis ¹	Endogenous		Reactive		Unipolar		Unipolar
Differences	REM	NREM	REM	NREM	REM	NREM	REM
Participants (N)	17	17	11	7	11	10	20
Mean Age $\pm$ STD ²	58.6	55.8	40.4	44.3	43.6 $\pm$ 11.6	44.6 $\pm$ 10.2	36.2 $\pm$ 6.5
Est. Change in REM Sleep ³	-63.2%	+14.1%	-64.3%	+11.8%	-44.5%	+27.2%	-16.1%
Antidepressant Users ⁴	0	0	0	0	4	6	0
Anxiolytic Users ⁵	?	?	?	?	4	2	?
Arousal Verification	Wake for 3 min		Wake for 3 min		Button Press		Verbal Report
Arousal Dist. across Nights	matched		matched		unmatched		N/A
Arousal Dist. within Nights	matched		matched		unmatched		N/A

¹In Vogel et al. (1975) diagnosis of *endogenous* and *reactive* depression were made by independent, qualified clinicians based on a ranking of symptoms which loaded onto either diagnosis. The final label was made so that patients were conservatively placed in the *reactive*. In Grözinger et al. (2002), two independent psychiatrists made the diagnosis of unipolar depression according to the DSM-IV. In Cartwright et al. (2003), participants were screened for unipolar depression related to marital discord and had to qualify with scores to at least 18 and 14 on the HAMD-21 and BDI.

²Mean Age ($\pm$ Standard Deviation if shown in respective study).

³For Vogel et al. (1975) all estimations are calculated from Baseline Measures and REM % for Deprivation Nights during the first three weeks. For Grözinger et al. (2002) the REM group's reduction is calculated as mean REM % of the experimental condition over the mean REM % of the controls. The true reduction is likely lower as controls displayed increasing REM %. The NREM groups increase is calculated as the change in REM % from Nights 1–2 and Nights 9–10. Cartwright et al. (2003) reduction is the mean reduction across each baseline and arousal night. If measured over the course of the treatment period (3 months), this reduction would be far lower.

⁴No participants in Vogel et al. (1975) were on antidepressants during the study but the wash-out period is not noted. Participants in Grözinger et al. (2002) using antidepressants were either on them 2-weeks prior, without improvement, or started with the experiment.

⁵Vogel et al. (1975) notes medication for anxiety, but not frequency, distribution, or number of receivers. Cartwright et al. (2003) had no mention of anxiety medication.

depression criteria as determined by two independent psychiatrists. Cartwright et al. focused on divorce-related depression, requiring participants to meet threshold scores ($\text{HAMD-21} \leq 18$, $\text{BDI} \geq 14$) and qualify on the SCID-NP, but without formal diagnosis.

These groups were also permitted differences in medication. Vogel et al. allowed treatments for agitation (chlorpromazine, phenobarbital) and insomnia (chloral hydrate), but no antidepressants. Grözinger et al. permitted antidepressants (trimipramine) for participants already on medication for more than two weeks without improvement, and also allowed treatment for anxiety/agitation (lorazepam) and insomnia (chloral hydrate). Cartwright et al. did not permit any medication.

Participant age differed substantially across studies. Vogel et al.’s endogenous cohort averaged 58.6 years, the reactive cohort 40.4 years. Grözinger et al.’s participants averaged 41.9 years. Cartwright et al.’s averaged 36.2 years. This 20+ year span may be relevant given age-related changes in REM sleep architecture (Floyd et al., 2007).

Magnitude and Scheduling of REM-R. The magnitude of REM-R and scheduling differed significantly between studies. Vogel et al. effected REM-D ($\geq 60\%$ reduction of REM %) with periodic recovery nights over a period of 3–9 weeks, dependent on participant response. Grözinger et al. effected REM-R ($\approx 45\%$ reduction) for 10 consecutive nights, independent of participant response. Cartwright et al. applied modest REM-R ($\approx 16\%$ reduction) on just 3 nights over a 4 month period.

Methodological Fidelity and Awakening Procedures. Vogel et al. conducted REM-R in a clinical setting using PSG, verified wakefulness for 3 minutes after each awakening, and targeted awakenings for REM onset. In contrast, Grözinger et al. also used a clinical environment but employed an artificial neural network (ANN) to trigger an alarm requiring a button press from participants. However, the ANN’s poor specificity (producing false positives in 26.8% of Stage I sleep) potentially resulted in participants being awakened during most Stage I periods in addition to REM sleep. Like Vogel et al., Grözinger et al. targeted REM onset, though the ANN’s limitations may have also triggered awakenings during N1 sleep. Cartwright et al. used manual PSG with verbal confirmation from participants upon awakening. Unlike the other researchers, Cartwright permitted increasing levels of REM sleep throughout the night, with delays escalating from 5 minutes after the first REM period onset to 20 minutes by the fourth period.

NREM Control Procedures. Concerning controlled NREM-I, Vogel yoked interruptions to occur ten minutes after a REM period in controls, matching both the number

of awakenings and their temporal distribution. In comparison, Grözinger equally distributed yoked interruptions across NREM stages II, III, and IV. Notably, the distribution of awakenings both across and within nights was unmatched between groups. Controls experienced nearly twofold more total awakenings early in the experiment, while REM-R participants experienced more total awakenings in the later half than controls. Additionally, controls consistently had more early-night awakenings (before 3AM) throughout the protocol, reaching 20+ pre-3AM awakenings on seven out of ten nights compared to just once in the REM-R group.

Response Patterns. Vogel’s data signaled an advantage of REM-D over NREM-I in endogenous depression (parametric $p=0.029$; non-parametric $p=0.09$), but equal improvement for reactive depression. The reactive cohort showed a higher recovery rate (72%) and shorter recovery time (mean 5.6 weeks) compared with the endogenous group (50% recovery, mean 7.3 weeks). Grözinger showed equivalent improvement for both REM-R and NREM-I during the 10-day treatment period. Cartwright had no NREM-I control but found 60% remission despite minimal and infrequent REM-R.

Synthesis and Interpretation. Despite substantial methodological limitations that prevent confident attribution of causal mechanisms, the consistent pattern of mood improvements across three methodologically distinct REM-R studies, spanning different populations, intervention intensities (16-64% REM reduction), and treatment durations (3 nights to 9 weeks), indicates that dismissal of this research line was premature. The specificity question, whether improvements result from REM-specific mechanisms versus non-specific sleep disruption, remains unresolved and requires systematic investigation with rigorous designs that address the limitations identified above.

2.3 Hidden REM effects in NREM Interventions

The effectiveness of TSD as a rapid antidepressant treatment has long been recognized, but uncertainty remains about which element of sleep is responsible for this effect. Selective REM-R as a treatment was dismissed after studies of NREM sleep reduction (NREM-R) and partial sleep deprivation also demonstrated symptom improvement (Giedke and Schwärzler, 2002; Grözinger et al., 2002). It was proposed, perhaps prematurely, that the antidepressant effects observed in REM-R studies stemmed from general sleep loss, not suppression of any specific state (Giedke and Schwärzler, 2002).

Combined with the substantial costs and methodological difficulties of selective sleep reduction experiments, this reasoning likely contributed to REM-R falling out of fa-

vor as a potential treatment from active investigation. However, one of the principal challenges in selective sleep modification studies is the homeostatic and interdependent nature of sleep architecture. In practice, sleep interventions, whether REM-R, NREM-I, or otherwise, produce secondary changes in sleep architecture, altering the NREM-REM cycle, state latencies and proportional sleep state composition. Furthermore, few studies account for these changes on the individual participant level, instead focusing on averaged group data. Consequently, observed improvements for DS could reflect indirect effects on REM sleep rather than direct NREM mechanisms, or vice versa.

Nevertheless, REM sleep abnormalities remain among the most consistent biological features of depression: shortened REM latency, elevated REM density, and increased REM % appear in unmedicated depressives. These characteristics can even continue in remission ([Palagini et al., 2013](#)). The persistence of these features suggests that REM dysregulation is not merely epiphenomenal, but may represent a state-sensitive marker or a causal factor of DS.

This area therefore faces a conceptual challenge: if REM, NREM, and total sleep manipulations can improve mood, do they function through a shared mechanism, or through distinct processes with similar treatment outcomes? TSD likely works through different processes than REM or NREM manipulations due to the differences in response time and persistence of effects ([Wirz-Justice and Van den Hoofdakker, 1999](#)). However, resolving whether REM and NREM manipulations share a common mechanism is critical to understanding how sleep architecture relates to mood regulation, whether selective REM intervention warrants re-examination, and if so, how practical interventions can be developed for treatment. Landsness et al. (2011), though designed as an NREM-focused experiment for DS, offers a useful case study for addressing this issue.

2.3.1 Landsness et al., 2011

Landsness et al. (2011) conducted a study on NREM-R, specifically targeting slow wave sleep (SWS), to determine the effects on DS. In addition to focusing on NREM sleep instead of REM sleep, this investigation differed from the previously described studies in three significant aspects. First, participants only experienced a single treatment night. Second, PSGs utilized a high density EEG array. And third, researchers used a method for modulating NREM sleep which did not require awakenings. This was possible by means of appropriately timed acoustic stimuli which disrupted Slow Wave Activity (SWA) without arousing participants ([Landsness et al., 2011](#)).

Study Design and Participants. Participants ($N = 17$) met criteria for Major Depressive Disorder (MDD) as determined by the Structured Clinical Interview for DSM-IV (SCID), and were also evaluated on the HAMD-17. Participants were not on medication for at least six months prior to the study, were required to maintain regular sleep-wake routine, not nap, and to avoid alcohol, caffeine, and nicotine throughout the study. Sleep habits were verified by wrist motor actigraphy. Sleep disorders such as apnea and restless limbs were exclusionary conditions. Two participants were excluded from the analysis due to intervention difficulties, so the results consider the remaining ($N = 15$).

Procedure. Participants participated in five laboratory nights, three of which pertained to the focus of this study: a baseline, treatment and recovery night. To address order-of-effects, ($N = 9$) participants underwent treatment and recovery on nights 2 and 3 respectively, and ($N = 8$) participants on days 4 and 5. The treatment was Slow Wave Deprivation (SWD) and this was achieved by active manual monitoring of each participant during sleep. On each treatment night, when a participant displayed SWA a pre-tailored auditory tone was played through a speaker, enough to disrupt SWS but not enough to rouse the participant.

Outcome Measures. Outcome measures for the study included sleep, depression, and vigilance scores. Participants were monitored by conventional PSG with high-density EEG. Depression scores were recorded before and after treatment through a modified self-rated Inventory of Depressive Symptomatology (mIDS-SR). The mIDS-SR was based on the IDS-SR16, with questions regarding weight and sleep removed, and questions from the larger IDS-SR30 added concerning irritability, anxiety, mood, future outlook, and hedonic capacity. ($N = 6$) participants were additionally measured by a shortened version of the HAMD-17 without questions concerning appetite/weight and insomnia (HAMD-13). Vigilance was measured before and after treatment by the Psychomotor Vigilance Task (PVT), Visual Analogue Mood Scale (VAS), and Stanford Sleepiness Scale (SSS).

Results: Sleep Outcomes. The treatment was successful in significantly reducing the amount of N3 % ($p < 0.001$, ANOVA) and SWA ($p < 0.001$, ANOVA). Compared with baseline, this represents a reduction of 54.3% and 37.3%. There was a non-significant increase in REM % of 18.8%.

Results: Depression Outcomes. There was a significant reduction of depression scores on both the mIDS-SR ($p < 0.01$, two-tailed paired t -test) and HAMD-13 ($p \leq 0.01$, two-tailed paired t -test). Three predictors for change in mood were identified: SWA dissipation on the baseline night, SWA rebound on the recovery night, and REM % on the treatment night, each having similar weights in a multiple regression

analysis for predicting mIDS-SR scores. A correlation analysis between REM % and a decrease in mIDS-SR score revealed significant effects ($r = -0.536, p = 0.04$). It is not noted in the study what relation SWS on treatment night, or changes in SWS between nights, had with depressive scores.

Methodological Considerations. Two methodological considerations should be noted.

REM Sleep and Treatment Response. Although REM % increased non-significantly during the SWD treatment compared with baseline measures, an expected phenomenon with NREM manipulation experiments, the analyses showed that the less REM % participants experienced during the treatment, the more their depressive symptoms improved. This observation is notable because as early as Vogel et al. (1975) it has been shown that NREM sleep manipulations can both improve depressive symptoms and increase REM % *at the group level*.

NREM Associations and Depression. Although this was a SWA deprivation study, the authors test associations between SWS and depression only for the treatment night; they do not analyze whether changes in SWS on baseline or recovery nights relate to changes in depressive symptoms.

2.3.2 Assessment

Even within an intervention explicitly designed to suppress SWA (an NREM phenomenon), Landsness et al. (2011) found that participants with the lowest REM sleep percentage during treatment experienced the greatest improvement in depressive symptoms. This association, observed incidentally rather than by design, suggests that REM modulation may contribute to antidepressant response even when NREM sleep is the primary experimental target.

The result does not overturn the role of NREM mechanisms but complicates the prevailing narrative that any sleep disruption is equivalently therapeutic. Instead, it suggests antidepressant efficacy may depend on the overall configuration and dynamic interaction of REM and NREM system. That is, altering one state perturbs the other. These findings further imply that the improvements observed in earlier NREM-I studies may be due, at least in part, by individual-level decreases in REM sleep that were masked by group averages. It is worth highlighting, the association between decreased DS scores and REM % occurs between participants during treatment, not changes in REM % from baseline. Nonetheless, this reopens a question the field effectively closed two decades ago regarding REM suppression's therapeutic potential. It also suggests that sleep interventions may require more precise, combined NREM and REM targeting.

Although preliminary and in need of replication, this finding suggests that the antidepressant effects previously attributed to NREM manipulations may have been driven, at least in part, by unexamined individual-level variation in REM sleep. As the following section will show, pharmacological evidence reveals a similar pattern: REM suppression as a consistent but overlooked feature of antidepressant action.

2.4 Pharmacology, REM Sleep, and Depression

The Arousal-Type REM-D Hypothesis. One of the principal catalysts for Vogel et al.'s initial pilot study in 1968 was previous evidence that existing antidepressants reduced or completely eliminated REM sleep (Vogel and Traub, 1968b). After the positive findings of the 1975 investigation and in response to criticisms in his later review, Vogel refined the hypothesis that REM-D was an effective treatment for endogenous depression (Vogel, 1975, 1983). Vogel et al.'s paradigm of REM-D by arousal was characterized by three conditions: (1) substantial REM reduction ($\geq 50\%$ from baseline); (2) persistence over weeks without evidence of habituation; and (3) a REM rebound after discontinuation of treatment. Vogel hypothesized that any pharmacological agent which caused these three "arousal-type REM-D" criteria would produce antidepressant effects (Vogel et al., 1990). Explicitly, arousal-type REM-D is a sufficient but not a necessary mechanism for antidepressant effects; a treatment which meets these criteria will have antidepressant properties, but other antidepressant mechanisms may exist.

To test this hypothesis, Vogel et al. (1990) conducted a systematic analysis of the scientific literature from 1962 through 1989, focusing on how various pharmacological agents influenced REM sleep and if they were associated with antidepressant effects. The scope of this review was extensive, encompassing "amine precursors, antidepressants, antihistamines, antipsychotics, barbiturates, benzodiazepines, other hypnotics, drugs affecting cholinergic and noradrenergic neurotransmission, ethanol, lithium and narcotics." Overall this included some 26 first and second generation antidepressants, and 44 other pharmacological agents. A total of 215 studies met their inclusion criteria (Vogel et al., 1990).

Supporting the hypothesis, the authors found that any pharmacological agent characterized as causing arousal-type REM-D demonstrated antidepressant effects. Only four antidepressants did not produce arousal-type REM-D: iprindole (atypical TCA), trimipramine (TCA), amineptine (atypical TCA), and trazodone (SARI). In summary, the authors found that arousal type REM-D "was produced by most, but not all drugs that improved endogenous depression" (Vogel et al., 1990).

Contemporary Evidence through 2010. Given the volume of new medications and studies published since 1989, a systematic review is beyond the scope of this work. The following sections draw primarily on comprehensive 2010 reviews by Beitinger & Fulda and by Staner et al., which examined antidepressants developed through 2010 (Beitinger and Fulda, 2010; Staner et al., 2010). Additionally, a targeted literature search was conducted for ketamine studies, as ketamine represents an important mechanistic contrast to traditional monoaminergic antidepressants not covered in the prior works.

As shown in Table 2.4, the pattern identified by Vogel (1990) has largely held. The vast majority of antidepressants suppress REM sleep and reverse characteristic abnormalities in depressed patients' sleep architecture, particularly by increasing REM latency. This is true for MAOIs, TCAs, SSRIs, Selective Norepinephrine Reuptake Inhibitors (SNRI), and Dual NA/5-HT Reuptake Blockers (Staner et al., 2010).

Reported Exceptions: Nefazodone and Bupropion. Two antidepressants were reported as "exceptions" because of early findings suggesting REM sleep augmentation. These drugs are highlighted here because they have been cited as contradicting Vogel's arousal-type REM-D hypothesis, but such criticisms rest on a misinterpretation (Thase et al., 2010). Vogel's hypothesis states that arousal-type REM-D is sufficient to produce antidepressant effects, not that it is necessary for all antidepressant mechanisms. The hypothesis makes no claim that REM augmentation would worsen depressive symptoms (inverse fallacy) (Thase et al., 2010). When examined in light of subsequent evidence, these apparent exceptions actually support Vogel's framework.

Nefazodone. The first is nefazodone, a Serotonin Antagonist and Reuptake Inhibitor (SARI). Two initial studies in the early 1990s found an increase in REM sleep, in both depressed and healthy participants (Beitinger and Fulda, 2010; Vogel et al., 1998). However, there are now more studies showing no change in REM sleep with nefazodone compared to those showing effect, and it is generally considered that nefazodone does not affect REM sleep (Beitinger and Fulda, 2010). Out of twelve antidepressants in a 2008 meta-analysis of published and non-published FDA clinical trials, nefazodone was tied for third lowest effect size ($g = 0.26$) versus placebo for treating depression (Turner et al., 2008).

Bupropion. The second is bupropion, a norepinephrine–dopamine reuptake inhibitor (NDRI), presenting with a similar history to nefazodone where an initial study demonstrated increased REM sleep, but numerous follow-up studies failed to replicate this finding (Beitinger and Fulda, 2010). In the same 2008 meta-analysis,

Table 2.4: Antidepressant effects on REM sleep latency and REM sleep. Based on Staner et al. (2010).

Drug	REM %	Latency	Drug	REM %	Latency
<i>TCA</i> s			Viloxazine	↓↓	—
Amitriptyline	↓↓↓	↑	<i>SNRI</i> s		
Imipramine	↓↓↓	↑	Reboxetine	↓↓	—
Trimipramine	↔	↔	<i>Dual NA/5-HT</i>		
Doxepin	↓↓	↑	Venlafaxine	↓↓↓	—
Clomipramine	↓↓↓	↑	Duloxetine	↓↓↓	—
Nortriptyline	↓↓↓	↑	Milnacipran	↓↓	—
Desipramine	↓↓	↑	<i>Dual NA/DA</i>		
<i>MAOI</i> s			Nomifensine	↔/↓	—
Phenelzine	↓↓↓	—	Bupropion	↑/↔	↓
Tranlycypromine	↓↓↓	—	<i>Others</i>		
Moclobemide	↓	—	Trazodone	↔/↓	—
<i>SSRI</i> s			Nefazodone	↑/↔	—
Fluvoxamine	↓↓	↑	Mianserin	↓	—
Fluoxetine	↓↓	↑	Mirtazapine	↔/↓	↑
Paroxetine	↓↓	↑	Agomelatine	↔	—
Citalopram	↓↓	↑	Tianeptine	↔	—
Sertraline	↓↓	↑			

Notes: ↑ = increased; ↓ = decreased; ↔ = no change; — = not reported in the cited review.

Multiple arrows indicate magnitude of effect. REM = rapid eye movement. SWS and Continuity columns have been removed from the original table. A REM latency column has been added by the author.

bupropion showed the lowest effect size versus placebo ($g = 0.17$) (Turner et al., 2008).

Thus, both drugs share a common pattern: neither consistently alters REM sleep (contrary to initial reports of REM sleep augmentation), and both demonstrate relatively poor antidepressant efficacy. These findings do not challenge Vogel’s hypothesis, which posits that arousal-type REM-D is a sufficient but not necessary mechanism for antidepressant effects. Instead, weak antidepressants that fail to suppress REM sleep are what the hypothesis would predict.

Long Term Effects of Antidepressants on REM Sleep. A critical test of Vogel’s arousal-type REM-D hypothesis concerns whether antidepressants maintain their REM-suppressing effects over extended treatment periods, as the hypothesis requires REM suppression to be persistent without significant habituation. A 2010 review by Beitinger & Fulda conducted a systematic investigation of the long-term effects (≥ 6 weeks) of antidepressants on sleep EEG, providing evidence on this question (Beitinger and Fulda, 2010).

TCAs. The authors found that TCAs, with the exception of trimipramine, maintained their REM-suppressing effects for months to years, with no systematic evidence of tolerance developing. The longest observation was recorded for imipramine at 4 years after commencement in 27 patients with a daily dose of 200 mg and still found significantly reduced REM sleep, fewer REM episodes, and increased REM latencies and REM density. During the last 3 years of this maintenance study with sleep recorded every 3 months, REM sleep parameters showed no systematic variation over time (Beitinger and Fulda, 2010). Nortriptyline maintained reduced REM sleep and increased REM density even after 15 months (Beitinger and Fulda, 2010).

MAOIs. Non-selective MAOIs demonstrated the most dramatic and sustained REM suppression of all antidepressant classes. Phenelzine completely eliminated REM sleep for 3–6 months in depressed patients, with very low amounts reappearing only after approximately 6 months. These reduced REM levels persisted for up to 3 years with continued treatment. Tranylcypromine, showed similar persistence, with reduced REM sleep and increased REM latencies maintained after 6 weeks. (Beitinger and Fulda, 2010).

SSRIs. Results for SSRIs were more nuanced. The SSRIs fluoxetine, citalopram, and paroxetine which have clear acute REM reducing effects, generally continued to reduce REM sleep in patients with major depression.

The longest study ($N = 12$) examined fluoxetine for 8 months, using a titration plan over the study duration, from 20mg 5 times a week down to just once a week. Despite

this frequency reduction, REM sleep remained suppressed below baseline throughout the study. No REM rebound was observed above baseline after discontinuation (Haro and Drucker-Colín, 2004). However, the second longest fluoxetine study followed 13 patients for 7 months on stable dosing (20–40mg per day) and found that while REM suppression occurred during acute treatment (10 weeks), REM sleep increased above baseline levels during continuation treatment (measured at week 30) (Trivedi et al., 1999). This suggests a tolerance development to fluoxetine’s REM-suppressing effects with daily exposure. Seven weeks after stopping fluoxetine, REM sleep remained significantly elevated compared to baseline (19.0% vs 16.2%), suggesting a persistent rebound effect. REM latency dropped significantly from 104.3 minutes at week 30 (while still on medication) to 66.7 minutes post-discontinuation. While not statistically lower than baseline, this represents a substantially shortened latency approaching levels typically associated with depression (Trivedi et al., 1999). This dramatic shortening of REM latency after discontinuation, combined with elevated REM %, suggests a persistent increase of REM pressure following chronic fluoxetine exposure.

Among other SSRIs with long-term data, clear differences emerge. Citalopram demonstrated sustained REM suppression over 6 weeks with no systematic changes over time. Paroxetine maintained reduced REM sleep amounts through 8 weeks but showed progressive shortening of REM latency. Sertraline (12 weeks) increased REM latency without reducing total REM sleep. Notably, sertraline was tied with fluoxetine (and nefazodone) for the third-lowest effect size (0.26) in the 2008 meta-analysis of FDA clinical trials (Turner et al., 2008). Fluvoxamine showed the fastest tolerance development, with initial REM suppression completely returning to baseline by 3 months, but REM latency remained elevated (Beitinger and Fulda, 2010).

Ketamine. In recent years, ketamine has shown growing promise as an atypical antidepressant (Nikolin et al., 2023). Although relatively few studies have examined ketamine’s effects on sleep architecture, two that did reported increases in REM % on the treatment night and total time in REM sleep, with no significant change in REM latency (Duncan Jr et al., 2017; Kheirkhah et al., 2025). Kheirkhah et al. (2025) found that ketamine significantly reduced REM density (number of eye movements over time) during the first REM period (RD1) without altering total night REM density, and that higher baseline RD1 predicted antidepressant response the next day (Kheirkhah et al., 2025). Both studies were inpatient, single-doses with measurements taken on the night of intervention.

Implications

The pharmacological evidence provides convergent support for a relationship between REM suppression and antidepressant efficacy. Vogel's arousal-type REM-D hypothesis has proven remarkably predictive: the vast majority of antidepressants produce sustained and significant REM reduction with subsequent REM rebound on discontinuation. The exceptions are revealing. Among traditional antidepressants with established efficacy, only four drugs did not produce arousal-type REM-D in Vogel's comprehensive 1990 review, and the apparent exceptions in newer antidepressants, nefazodone and bupropion, demonstrated the lowest effect sizes in meta-analyses of investigations comparing them to placebo.

This convergence between behavioral REM-R studies (Vogel et al. (1975), Grözinger et al. (2002), Cartwright et al. (2003)) and monoaminergic pharmacological mechanisms suggests REM suppression represents a major therapeutic pathway. However, this relationship does not establish causality. REM suppression could be either a direct therapeutic mechanism, an epiphenomenon of other neurological changes, or a reliable biomarker of monoaminergic drug engagement.

Ketamine serves as an important boundary condition for traditional antidepressants. Ketamine produces potent and rapid antidepressant effects, often within hours, even for treatment resistant depression (Nikolin et al., 2023). The acute response time of ketamine differs fundamentally from traditional antidepressants. Traditional antidepressants typically require 4–6 weeks to achieve therapeutic effects and the majority suppress REM sleep chronically throughout this period (Beitinger and Fulda, 2010). These distinct response times suggest different mechanistic pathways. While ketamine increases REM %, Kheirkhah et al. (2025) found it significantly reduced REM density during the first REM period, with higher baseline RD1 predicting next day antidepressant response. This may indicate a convergence with REM suppression effects, or may be an independent pathway that merits further exploration.

These findings indicate that suppression of REM is one antidepressant mechanism among multiple pathways, not a universal one. The strength of convergent evidence between REM-R interventions and multiple drug classes through decades of research suggests however it is a major mechanism, particularly for monoaminergic antidepressants. Different drug classes achieving antidepressant effects through distinct pathways (monoaminergic versus glutamatergic) does not diminish the therapeutic relevance of REM modulation; rather, it highlights that responder heterogeneity may partly reflect differential engagement of these pathways.

The neurobiological basis for a REM-R antidepressant relationship remains unclear.

While REM sleep involves monoaminergic systems, particularly serotonergic and noradrenergic transmission, simple chemical imbalance theories, in particular the serotonin hypothesis of depression, remain unsupported (Beitinger and Fulda, 2010; Moncrieff et al., 2022). Understanding the mechanistic pathway through which REM suppression relates to mood improvement, possibly through neurotransmitter systems, direct effects on emotional processing, changes in cortical excitability, or other processes, requires further investigation. The persistence of REM-suppressing effects over months and years of treatment also raises important questions about long-term consequences that remain incompletely understood.

2.5 Discussion

2.5.1 Summary of Findings

Two RCT studies of REM-R and two additional investigations with varying designs have demonstrated mood improvement associated with reduced REM sleep. Vogel et al. (1975) conducted the most extensive investigation, using a double-blind, crossover design with ($N = 34$) endogenous and ($N = 18$) reactive depressed patients. Endogenous depressed patients who underwent three weeks of REM-D (pre-crossover) improved significantly more than NREM-I control subjects, with half of REM-D patients improving compared to less than a quarter of controls. The observation that improvement correlated with REM pressure buildup came from Vogel’s earlier pilot studies with nine patients, and through follow-up analyses of 14 of the endogenous patients providing additional support for this relationship (Vogel et al., 1968, 1977, 1980).

Grözing et al. (2002) replicated these findings after a 27-year gap using an automated ANN-based REM detection to treat ($N = 11$) patients with REM awakenings and ($N = 10$) with NREM awakenings for ten consecutive nights. Both groups showed substantial improvement (33% and 41% reduction in Hamilton scores), with no significant difference between conditions. However, the automated detection system introduced timing uncertainties, and the distribution of NREM awakenings across sleep stages may have disrupted the ultradian rhythm differently than Vogel’s protocol (Grözing et al., 2002).

Cartwright et al. (2003) employed a gentler, longitudinal approach, studying ($N = 20$) individuals experiencing divorce-related depression over four months with intermittent REM awakenings for dream collection. This study differed substantially in using a non-clinical sample with less severe depression but lacked control of variables for time-based adaptation to stressors (Cartwright et al., 2003).

Landsness et al. (2011) investigated SWD rather than REM-R, yet observed that lower REM percentage during treatment correlated with greater symptom reduction. This finding emerged from a post-hoc analysis in a study of 15 non-medicated patients with major depression. While the primary intervention targeted SWA, the unexpected REM correlation suggests that REM suppression may contribute to antidepressant effects even when not the explicit therapeutic target (Landsness et al., 2011).

All REM-R studies demonstrated considerable methodological heterogeneity. Populations ranged from hospitalized depressed individuals, classified with pre-DSM conventions to outpatient divorce-related “blues.” Each treatment method was distinct in intensity, duration, frequency, and target. The degree of REM suppression differed substantially, from approximately 16% reduction in Cartwright’s protocol (compared to NREM-I controls) to more severe deprivation in Vogel’s study. Control conditions included NREM awakenings and non-depressed comparison groups. Despite this variability, a consistent pattern emerged: reduced REM sleep, achieved through different methods and in different populations, was associated with mood improvement.

However, the specificity question, whether REM sleep plays a specific role in depression treatment or whether therapeutic effects arise from general sleep disruption regardless of stage, remains unresolved. Three distinct problems contribute to this uncertainty. First, the small number of controlled trials, only three to date, limits statistical power and precludes a meta-analytic synthesis. Second, inherent methodological weaknesses in each study reduce confidence in each study’s effect estimates. Third, and most consequential, specific confounders within two of the three studies are capable of either masking a real REM-specific effect or generating a spurious one.

Grözing et al. (2002) illustrates the masking risk. The concurrent antidepressant use of trimipramine in a subset of participants would tend to compress any difference between the REM-R and NREM-I arms and thus obscure a real REM-specific effect. Although trimipramine itself does not reduce REM sleep percentage, it does affect REM pressure, which means the same REM regulation system the intervention targets was already perturbed in those participants (Feuillade et al., 1992). Unmatched awakening distributions across and within nights, together with poor specificity of the automated ANN detection system (26.8% of Stage I sleep misclassified as REM), further erode the cleanness of the contrast between conditions.

Cartwright et al. (2003) illustrates the opposite risk, of attributing to REM what could be better attributed to a confound. Remitters had been separated from their

spouses substantially longer than non-remitters (13.0 vs 4.18 months), nearly identical to the separation duration of non-depressed controls (13.2 months), providing a plausible alternative explanation for the observed remissions through natural recovery rather than the intervention itself. A separate concern is the strength of the manipulation: the actual REM reduction achieved was minimal (16.1% on average) and inconsistent, weakening the basis for any inferred dose–response relation.

These constraints, singly and in combination, prevent confident conclusions about whether the observed improvements reflect REM-specific mechanisms, non-specific effects of sleep disruption, or other factors. What remains noteworthy, despite the limitations, is the consistent direction of mood improvement across methodologically distinct studies.

The Landsness study reveals an important analytical consideration. Their observation that individual REM sleep percentage correlated with treatment response, despite targeting SWA, demonstrates how individual response patterns can be obscured by group mean comparisons. This suggests that apparent non-specificity in group-level analyses may mask subpopulations of REM-responsive individuals. Future research must move beyond simple group mean differences to examine individual response profiles, potentially identifying which patients benefit from REM-specific versus general sleep disruption interventions.

Pharmacological evidence provides convergent support for a relationship between REM suppression and antidepressant efficacy. Traditional antidepressants across multiple drug classes (TCAs, MAOIs, most SSRIs) consistently suppress REM sleep, often dramatically and persistently. The exceptions are instructive: drugs producing minimal or inconsistent REM suppression (nefazodone, bupropion) demonstrated relatively limited effect sizes in meta-analyses of antidepressant efficacy against placebo. This pharmacological convergence, spanning decades of research and multiple mechanisms of action, suggests REM suppression represents a major pathway through which traditional monoaminergic antidepressants exert therapeutic effects.

However, this evidence remains correlational rather than causal. It remains unclear whether traditional antidepressants work, at least in part, through reduction of REM sleep, or whether REM suppression is a downstream consequence of other mechanisms. The existence of effective antidepressants with atypical REM profiles (particularly ketamine, which increases REM sleep) indicates that REM modulation is sufficient but not necessary for antidepressant action, consistent with Vogel’s original formulation of the hypothesis.

2.5.2 Mechanistic Clues from Adjacent Research

While the depression studies reviewed above provide behavioral evidence for REM-R's antidepressive effects, the mechanisms that might mediate these benefits remain unclear. Selective REM disruption in healthy controls offers an avenue to observe acute effects for possible mechanisms such as emotional processing and cortical excitability. Four modern REM-A experiments provide evidence which supports a testable hypothesis, that REM sleep applies a "brake" on emotional responsivity and cortical excitability. When REM sleep is suppressed, the system is more reactive.

Rosales-Lagarde et al., 2012. ($N = 20$) healthy male participants completed the same fMRI task on two days, separated by a single experimental night. Participants were randomly assigned to either REM-D or matched awakenings during NREM sleep, equating the number and timing of arousals across groups. During the fMRI task, participants were instructed to respond to presented images as soon as possible, choosing to either defend themselves or not, with responses recorded through button presses. Defense responses were indexed as "high emotional response" (HER). Compared against baseline measures, only HER reported trials increased after REM-D. The fMRI results showed that regions which normally down-shift with repeated exposure (visual association cortex; ventrolateral prefrontal cortex) failed to decrease after REM-D, whereas the NREM-I group showed the expected exposure reductions. Reaction times sped up in both groups, consistent with retest effects, so the between-group difference resides in change from baseline, not in initial levels. PSG confirmed significant reduction of REM sleep (just 4.03% of TST), and NREM-I controls experienced an increase ([Rosales-Lagarde et al., 2012](#)).

Spoormaker et al., 2012. ($N = 18$) healthy male participants underwent a three-day fear learning and extinction protocol. Participants were conditioned to receive a shock on Day 1, underwent extinction training (i.e. no shock with the stimulus) on Day 2, and repeated the extinction recall experience on Day 3 with skin conductance recording. On the night of Day 2, subjects were randomized to REM-D or NREM-I, with the same number and distribution of awakenings. During block 1 of the recall experiment, both groups showed an elevated skin conductance response. Only the REM-D group continued to show significantly elevated skin conductance responses on trial block 2 (out of 5 blocks) of recall. Neither group demonstrated elevated responses to the later blocks. Imaging analyses focused on prediction-error signals and revealed a group-by-stimulus interaction centered in lateral temporal/occipital cortex, consistent with altered updating of the extinction memory after REM suppression. PSG confirmed a significant reduction in REM % (5.1% of TST) but the control condition had a notably lower REM sleep percentage

(15.1%) than is typically expected (20–22%) (Spoormaker et al., 2012).

Placidi et al., 2013. ($N = 10$) healthy adults completed a within-subject protocol across three mornings: baseline, after selective REM-D via awakenings, and after a control night with matched NREM awakenings. Cortical excitability was measured with transcranial magnetic stimulation (TMS). Two inhibition-sensitive measures, the cortical silent period (CSP) and short-interval intracortical inhibition (SICI), both indicated reduced inhibition after REM-D. The CSP is the brief relaxation period in voluntary muscle activity after a single TMS pulse over the primary motor cortex. The SICI reveals neural inhibition by applying a small, priming pulse followed by a larger pulse which can cause a muscular response, and this is compared with the larger pulse alone. The CSP shortened and SICI decreased, while the resting motor threshold and unconditioned motor-evoked potential amplitude were unchanged. PSG confirmed robust REM-D (3.7% of TST) with similar total sleep loss across REM-D and NREM-awakening nights. The pattern indicates increased cortical excitability from REM-D (Placidi et al., 2013).

Glosemeyer et al., 2020. ($N = 42$) participants were assigned to REM-R, slow-wave sleep (SWS) suppression, or undisturbed sleep, next-day fMRI during a social exclusion task (Cyberball). Negative affect, measured by the Positive and Negative Affect Schedule (PANAS), increased most after REM suppression and scaled with the degree of REM reduction. Subjective feelings of exclusion and the efficacy of instructed cognitive reappraisal did not differ by group. Neurally, the REM-suppressed group showed stronger right amygdala responses to exclusion and more positive amygdala–anterior cingulate cortex coupling during passive viewing, pointing to heightened limbic responsivity rather than a deficit in deliberate regulatory control. (Glosemeyer et al., 2020)

Synthesis. Across these paradigms: threat-based decisions, extinction recall, social exclusion with affect ratings, and TMS measures of motor cortex, selective REM curtailment appears to produce acute, next-day increases in emotional response in combination with heightened neural excitability. fMRI findings support greater sensory and limbic responses or failure to habituate to previously exposed information, and TMS found reduced cortical inhibition. Because the effects appear when awakenings are matched in number and timing (REM-D vs NREM-I) or when an alternative stage (SWS) is suppressed, they are unlikely to reflect non-specific sleep loss alone. This supports a mechanistic account in which REM sleep imposes a following day damping function on emotional processing and neural excitability systems.

Three points constrain the interpretation. First, these are acute effects measured the next morning in small, healthy cohorts. They do not by themselves establish clinical benefit or significance. Second, translation to depression likely depends on repeated cycles of REM suppression and recovery, consistent with timelines seen in both previous research and antidepressant pharmacology that chronically suppresses REM sleep. Finally, affect measures an outward, momentary expression of emotional state, while mood is an internal, sustained one. Together with the clinical and pharmacological evidence reviewed above, these mechanistic studies motivate the hypothesis that systematically reducing REM could contribute to a specific antidepressant action by altering dysregulated systems in individuals with depression.

2.5.3 Responders and Depression

Evidence for Responder Subset. All three REM-R studies have revealed a consistent pattern, that treatment responses are heterogeneous, with distinct responder and non-responder subgroups rather than uniform effects (Cartwright et al., 2003; Grözinger et al., 2002; Vogel et al., 1975). Most notable is Vogel et al.’s (1975) observation concerning endogenous patients who failed to respond to REM-D. When these non-responders were subsequently treated with TCA antidepressant imipramine, only one out of nine completers responded ($N = 11$ intent to complete) (Vogel et al., 1975). Compared to typical remission rates for imipramine, this is unusually low (Stewart et al., 1993). This pattern suggests a common underlying mechanism. Explicitly, individuals who failed to respond to REM-R also failed to respond to an antidepressant that is an effective REM suppressant. Furthermore, Vogel et al. found in both a pilot study and follow-up analysis that REM rebound correlated with improvement (Vogel, 1983; Vogel et al., 1968, 1977). Those who did not show REM rebound did not improve. In other words, there is evidence for a subset of depressed individuals who respond to a particular biological mechanism related to REM sleep modification. This finding of treatment-responsive subsets should no longer be surprising given decades of accumulated research into what we call depression.

Heterogeneity Challenge of Depression. The existence of distinct responder profiles points to a more profound issue in depression research. That is, depression is likely not a singular disorder with a single pathophysiology, but rather a collection of distinct disorders grouped under a common diagnostic label. This heterogeneity is inherent to both the standards used by Vogel et al. in 1975 as it is in today’s DSM, now in the fifth edition (DSM-V) (American Psychiatric Association, 2013).

Consider the criteria for Major Depressive Disorder (MDD) as currently described in the DSM-5 (American Psychiatric Association, 2013). Nine different symptom

axes are considered for MDD diagnosis. To qualify, an individual must experience anhedonia or depressed mood, plus 4 other symptoms. Some symptom axes are non-specific, like sleep or appetite, where any increase or decrease, count towards diagnosis. If all possibilities from this scoring system are accounted for, there are 945 unique symptom profiles that qualify for an MDD diagnosis (Fried and Nesse, 2015a,b). Moreover, measurement scales such as the BDI, which assesses depression by symptom and magnitude, results in arbitrary cut-offs for diagnosis. Two patients who receive equal scores may have entirely distinct symptom expression and severity. One individual could score highly in few symptoms, while another weakly but in many (Beck et al., 1996; Fried and Nesse, 2015a). Most dramatically, it is possible for two patients to meet the criteria for MDD yet have no overlapping symptoms except for anhedonia or depressed mood (American Psychiatric Association, 2013). Therefore, it seems unreasonable to expect such clinically distinct individuals to share the same pathophysiology and respond to the same treatment, and is likely why antidepressant efficacy is weak compared to placebo (Turner et al., 2008).

Diagnostic Challenges in Psychiatry. This challenge of heterogeneity extends beyond depression to psychiatric disorders generally. The use of case-control designs with these symptom-based diagnostic categories is fundamentally problematic. First, between patients and healthy controls, diagnostic boundaries are determined by arbitrary thresholds on symptom count or severity scales rather than by any demonstrated alteration or impairment in underlying biology. An individual scoring just below the diagnostic cutoff may be more similar to someone just above it than either is to individuals at the extremes of their respective groups. Second, patients sharing the same diagnosis can present with radically different symptom profiles, creating substantial within-group heterogeneity that raises questions about whether they share a common underlying condition. Finally, the high rates of comorbidity between psychiatric categories further complicate classification. Depression and anxiety scores correlate at $R=0.81$ (Gillan et al., 2016), and more than 90% of Obsessive-Compulsive Disorder (OCD) patients suffer from a second disorder (Ruscio et al., 2010). This creates situations where two patients with different primary diagnoses may share more symptoms in common than they share with others in their own diagnostic group.

In short, the challenges of heterogeneity and comorbidity pose serious concerns for psychiatric research because it is ultimately unknown whether diagnostic labels appropriately classify individuals into meaningful biologically founded groups. This ambiguity in case-control definition muddles both treatment selection and the interpretation of experimental findings.

A Solution: Biomarkers for Population Subsets. In medicine, ambiguity in case-control design is dispelled through understanding of the pathophysiology and the use of relevant biomarkers. Both are largely lacking in mental health (Insel et al., 2010). However, unlike other areas of medicine, biomarkers for mental illness are not straightforward or in standard clinical use. Psychiatric diagnosis continues to rely on clinical symptomatology rather than the responsible underlying physiological cause. While these categories may have some practicality for clinical use, they lack predictive power for treatment selection. The finding by Vogel et al. (1975) illustrates this problem perfectly. The existence of a subset of patients who fail to respond to REM-suppressing antidepressants, and a failed response to REM-D by arousal, suggests a distinct biological subtype. Biomarkers are therefore needed to identify these individuals and determine proper class labels before or during early stages of treatment. Without identifiers, medical practitioners are left to trial-and-error approaches which delay effective treatment, and researchers must contend with muddy class labels.

Sleep as a Window for Psychiatric Biomarkers. Sleep offers a uniquely promising window into potential biomarkers for mental illness (Meyer et al., 2024). Sleep architecture can be measured objectively, non-invasively, and repeatedly. Sleep disturbances are nearly universal across psychiatric conditions, yet the specific patterns of disruption may differ meaningfully between biological subtypes. REM sleep parameters such as latency, percentage, and density show measurable, objective alterations in depression (Ahmadi et al., 2010; Moretto and Palagini, 2019; Palagini et al., 2013). These patterns may predict treatment response, as Vogel’s REM rebound findings suggest, and sleep measures can track longitudinal changes during treatment.

The convergence of behavioral (REM-R), pharmacological (antidepressants), and sleep architecture evidence suggests that sleep, particularly REM sleep, may provide an accessible window for identifying depression subtypes. Identification of which patients are likely to respond to REM sleep modulation would represent a significant advance toward a biologically based treatment in psychiatry.

However, fully exploiting sleep as a source of psychiatric biomarkers faces a critical barrier: the current tools for sleep monitoring and manipulation are expensive, labor-intensive, and largely confined to specialized research settings. Addressing the outstanding questions about responder profiles, dose-response relationships, and mechanisms requires accessible, validated tools for sleep assessment and selective stage manipulation. This methodological limitation has contributed to the decades-long gap between initial promising findings and a systematic follow-up investigation,

a gap that motivates the open-source platform development described in subsequent chapters.

2.5.4 Conclusion

This review examined whether the abandonment of REM sleep reduction (REM-R) research as a potential treatment for depression was justified. While all three controlled REM-R trials demonstrated consistent patterns of improvement for DS, significant methodological limitations prevent definitive conclusions about efficacy and specificity. However, the absence of rigorous evidence should not be mistaken for evidence of absence.

The methodological concerns including awakening distributions, selectivity, potential confound, and incomplete reporting establish that prior studies cannot definitively answer whether REM-R produces specific antidepressant effects. However, the consistent direction of mood improvements across diverse methodologies, populations, and intervention intensities, combined with convergent pharmacological evidence showing that most traditional antidepressants suppress REM sleep, suggests this therapeutic pathway warrants systematic reinvestigation.

The primary obstacle preventing the systematic reinvestigation called for above is neither conceptual nor clinical, but financial and technological. The tools required for multi-night sleep monitoring with selective stage manipulation remain prohibitively expensive outside well-resourced laboratories or non-existent. This bottleneck has shaped the literature reviewed here in concrete ways: only three controlled human trials of REM-R have been conducted in over fifty years; the most rigorous of these required hospitalized patients and weeks of in-laboratory polysomnography; and the open question of which depressed individuals respond to REM-specific interventions cannot be answered without studies large and varied enough to characterize responder subgroups. Each of these limitations is a direct consequence of the cost and complexity of the available equipment.

The convergent evidence gathered in this chapter, behavioral, pharmacological, and mechanistic, is sufficient to make the case that reinvestigation is warranted. What it cannot do on its own is make reinvestigation tractable. The shape of that reinvestigation is already visible in the chapter's findings. A credible follow-up must use REM-targeted awakenings, the only intervention method shared across the three controlled trials and the one whose effects the literature has actually characterized. It must run for multiple consecutive nights over weeks rather than a single laboratory session, since the clinically meaningful effects observed by Vogel and Grözinger emerged only over such timescales. And it must be deployable at a scale and across

a population diversity sufficient to separate responders from non-responders, a requirement that follows directly from the heterogeneity signal present in every REM-R study reviewed here. Each of these demands points toward the same class of solution: a device affordable enough to be deployed widely, comfortably wearable and durable enough to be tolerated for weeks of nightly home use, and capable of both detecting REM and delivering the intervention automatically.

Chapter 3 documents the design and iterative development of the Open-Source Sleep Monitor and Modulator (OSSMM), a platform built to meet exactly these requirements. OSSMM is wearable, self-buildable from commercially accessible parts, and engineered for durability over weeks of nightly use, with a target build cost well below that of any clinical-grade equivalent. It is intended to support automated REM detection and REM-targeted awakenings as a single integrated system. Without reducing the cost and complexity of the underlying tools, the questions left open by Vogel, Grözinger, and Cartwright are unlikely to be revisited at the scale required to answer them.

Design Evolution of the OSSMM System

“Le mieux est le mortel ennemi du bien. (The best is the mortal enemy of good.)”

—Montesquieu, *Pensées*, 1726–1727

Abstract

This chapter documents the iterative development of the Open-Source Sleep Monitor and Modulator (OSSMM) system through three major prototypes. This development was a technological response towards addressing the questions raised in Chapter 2 concerning associations between Rapid Eye Movement sleep and depressive symptoms, given the prohibitive costs and lack of automated solutions detailed in Chapter 1. Here, the reasoning behind the design decisions and implementations is explained.

3.1 Introduction

The Open-Source Sleep Monitor and Modulator (OSSMM) is a wearable platform developed in this thesis to democratize sleep research by making sleep monitoring and, in the future, sleep modulation affordable and accessible for widespread research use. Building such a device meant reconciling demands that pull against one another: optimal signal quality, all-night battery life, nightly comfort, low cost, and assembly without specialist equipment. Translating requirements like these into a working product is seldom a linear process, and the OSSMM was no exception. This chapter therefore follows the design as it worked through those tensions, tracing its evolution across an initial conceptualization, the foundational investigations that settled the component set, and the fabrication of three working prototypes, the last becoming the final release version. The account is largely chronological, except where steps are combined or reordered for clarity, illustrating how engineering constraints, product experience, and technical discoveries shaped the final system architecture.



Figure 3.1: Many of the 3D printed design iterations created during this thesis. On the far right, the three complete prototype versions can be seen.

The design evolution presented here demonstrates the transformation of theoretical constraints and aspirations into practical engineering solutions, addressing challenges in product miniaturization, physiological signal collection, cost reduction, user comfort, and manufacturing reproducibility. Each prototype iteration attempted to refine the balance between ideals and real-world practicality. Ultimately, it was possible to achieve a wearable unit build cost of under €40, while maintaining ergonomics for nightly wear, fidelity for multi-night usage, and a data collection quality suitable for machine learning sleep stage classification.

3.2 Design Requirements and Goals

3.2.1 Design Constraints

The OSSMM project commenced with several design aspirations. The purpose was to not only build a sleep monitoring and modulation system, but to design one that was affordable, accessible, professional, and safe. Therefore, the project was framed by several design constraints at the start. These are found in full in Table 3.1.

In short, the project aimed to build a wearable platform. This modality was chosen out of ethical concerns over the potential for abuse of non-contact sensing methods, since the deliberate act of donning equipment functions as a form of consent. Within this framing, the design intent followed a clear hierarchy of priorities. Safety came first. Comfort and ease of use came second. Since any system that resists use is likely to be abandoned, the platform was therefore designed to be light, comfortable, and sleek, avoiding external wiring that could tangle or protrusions that could catch on bedding. Beyond wearability, the system necessarily had to capture sleep-staging discriminants with adequate signal quality, providing a high performance-to-cost ratio relative to existing systems. Finally, to be useful it also had to be accessible, produced with readily available parts, and simple in-house assembly. Table 3.1 formalizes these constraints.

Table 3.1: System design requirements for the OSSMM system.

Req.	Description
<i>Physical Design Requirements</i>	
1	Any sensors must be worn or require some form of “contact.”
2	Any wearable portion must be light and small, in short, as unobtrusive to sleep as possible.
2a	– Must run off an internal power source (i.e., no power cables).
2b	– Battery life should safely exceed typical sleep duration (over 12 hours).
2c	– No exposed outer wiring, even if safely sheathed.
<i>Functional Requirements</i>	
3	The system must accurately detect REM sleep in near-real time (within 3 minutes of REM onset).
4	The system must incorporate an automated method of sleep modulation which can be enabled/disabled.
11	The system safely records all pertinent data and can carry out essential statistics (i.e., percentage of sleep states, time wake/sleep, number of awakenings, etc.).
<i>Cost and Accessibility</i>	
5	One wearable unit will not cost more than €130.
8	All parts must be commercially accessible without restriction.
13	The system should not require any consumables (e.g., disposable electrodes, gels, etc.).
<i>Assembly and Usability</i>	
6	Construction of the system must be such that individuals with a basic technology background can assemble it.
7	Construction of the system must be as straightforward as possible, not requiring advanced tools or techniques which cannot be found in a basic university laboratory.
9	Wearers must be able to easily learn how to use and maintain the system.
<i>Safety and Durability</i>	
10	The system should be durable such that usage over weeks to months does not diminish performance, or if there are susceptible parts, they are effortlessly replaced.
12	The system must not carry any overt risks, such as being constructed from materials which can cause irritation or allergy.

3.2.2 Design Goals

Beyond the primary design constraints the project aspired to incorporate or achieve further design goals. These are found in Table 3.2. These targets elevated the standards of the system, in terms of performance and affordability. While REM sleep was a priority within the design constraints for this thesis’ objectives, multi-stage sleep classification was also included for future, wider use of the system.

Table 3.2: Design goals for the OSSMM system.

Req.	Description
<i>Performance and Accuracy Requirements</i>	
1	Data collected by the system should strive for the highest possible quality approaching clinical grade standards.
2	The system must accurately detect Wake, NREM, and REM sleep in near-real time (within 3 minutes of REM onset).
4	Device should achieve 90% general accuracy for REM/NREM detection.
<i>System Features</i>	
3	The system should offer additional consensual methods for conducting tests and collecting information.
<i>Cost Requirements</i>	
5	One system unit should not cost more than €65.

3.3 System Conception and Architectural Decisions

The architectural decisions described in this section emerged in response to the market survey presented in Chapter 1, and the discussion of available sleep monitoring and modulation platforms in 2020–2022. That survey identified two related gaps: low-cost consumer devices (e.g. smartwatches, smart rings, etc.) lacked the staging accuracy required for research, while research-grade portable systems remained prohibitively expensive (i.e. greater than 3,500 per unit). No general-purpose, automated sleep modulation platforms were commercially available. The OSSMM system was conceived to address these gaps simultaneously, and the design constraints in Tables 3.1 and 3.2 are the formalization of what the survey indicated was missing: a low-cost, accessible, research-suitable platform with both monitoring and modulation capabilities.

The initial system conception established a partitioned architecture to resolve the competing demands of between data collection, processing, and wearable constraints. Given modern technological limits, it was believed that at the desired price point, no single device could simultaneously run for more than twelve hours, record high quality data, analyze and store that data in real-time, be worn comfortably while asleep, assembled without advanced equipment, and cost no more than €130. The solution was to partition the system such that collectively the requirements would be fulfilled while easing the demands on each part.

This partition resulted in two principal components, illustrated in Figure 3.2: a wearable unit called the Recorder responsible for physiological data collection, and a smartphone-based Base Station running a companion app for data analysis and storage. The Recorder, optimized for minimal size and low power consumption, continuously streams sensor data wirelessly to the Base Station. This separation meant the wearable portion could achieve the goal of having a small form factor with its limited energy requirements, while the Base Station could leverage existing computational resources for processing and storage.

Smartphones were a logical choice for the Base Station to reduce overall system cost, as they are widely available and provide a powerful computational platform. As of 2018, 84% of US households claimed to have at least one smartphone, and as of 2019, 84% of Irish adults claimed ownership ([Gibney and McCarthy, 2020](#); [Martin et al., 2021](#)). In Ireland, adoption rates reached 98% among 18–25 year-olds, suggesting higher future adoption ([Gibney and McCarthy, 2020](#)). The OSSMM system capitalizes on their widespread ownership through a dedicated companion app. This approach potentially eliminates additional hardware costs while providing sufficient processing power for real-time analysis, built-in wireless connectivity, and potential future cloud storage capabilities via existing cellular or WiFi networks. This cost reduction is contingent on whether researchers leverage participants' existing devices or supply dedicated Android devices. Nonetheless, the familiarity users have with their own devices (or smartphones in general) further facilitates the deployment of an intuitive OSSMM app for sleep data visualization, and the future potential of modulation control.

The Android operating system was selected for the Base Station application development due to its strong market share among smartphone operating systems (70% as of 2021), maximizing potential user accessibility ([Statista, 2024](#)). The Base Station handles data storage and, with future development, could provide real-time sleep stage analysis and cognitive tasks as additional measurement methods for researchers. Though never implemented during this thesis, previous research has

demonstrated the potential for assessing depression through cognitive tests, including video games (Gillan et al., 2016).

3.3.1 Hardware Architecture Conceptualization

The Recorder hardware architecture was conceptualized to meet Design Constraint #5 mandating unit costs not exceed €130, and Design Constraint #8, requiring all parts to be commercially accessible without restriction. As shown in Figure 3.2, the initial Recorder design comprised six principal components: Sensors/Electrodes, Amplifier and Filtering Circuits, the Microcontroller Unit (MCU), Stimulator, Battery, and a Bluetooth transceiver. All electronic components were specified as commercial-off-the-shelf (COTS) parts in keeping with Design Constraint #7, which mandated that construction be possible without advanced tools or techniques beyond those found in a basic university laboratory.

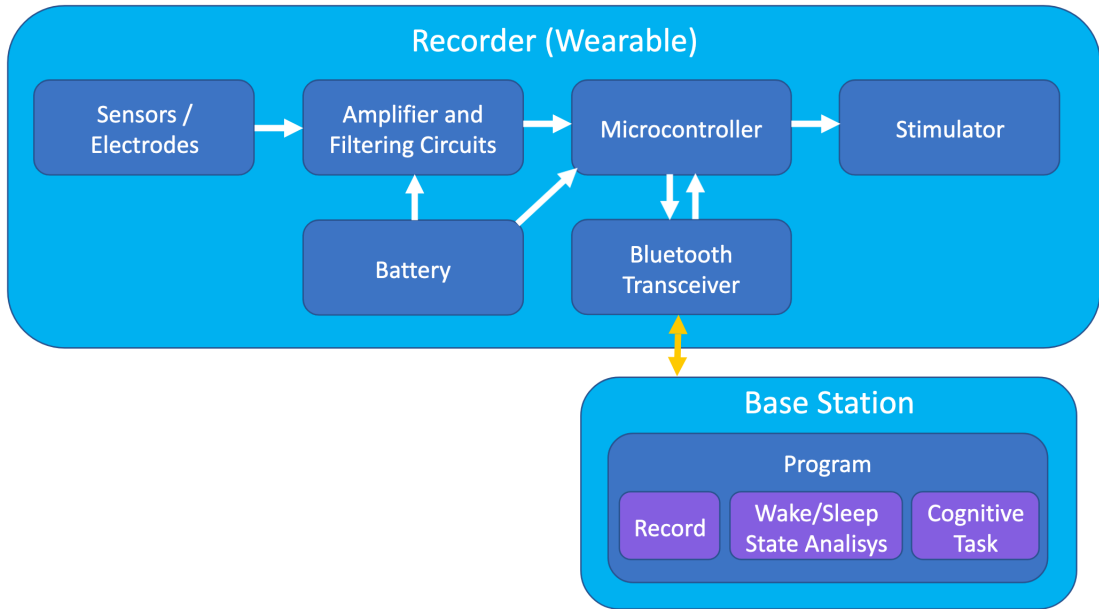


Figure 3.2: Early OSSMM System Overview

The conceptual action flow was as follows: The sensors/electrodes detect physiological signals such as eye movement, and these signals are amplified and filtered through electronic hardware before reaching the MCU. The MCU digitizes these analog signals, and with minimal onboard processing (in order to conserve power), transmits them wirelessly to the Base Station, with wireless transmission required by Design Constraint #2c, prohibiting exposed external wiring. The Base Station records this data in a unique session for each sleep recording and collects the data into a moving window of 30 second epochs. This collection of epochs is analyzed to classify the most central observation in the window. If a certain classification threshold is reached, for instance four consecutive classifications of slow-wave sleep,

the Base Station can communicate back to the Recorder to activate the stimulator for potential sleep modulation.

3.3.2 Hardware Safety

Due to the wearable aspect of this project, safety was a priority. Voltages were kept below 5 volts. All components were inspected prior to use. Data and safety sheets were consulted to confirm minimal risk, such as components which required extended dermal contact. Appendix F provides ethics approvals from Maynooth University for experiments applied for during this thesis. The final prototype, whose safety is discussed thoroughly in Chapter 4, underwent an independent expert review as part of ethics, and health and safety approvals.

3.4 Foundational Investigations

Before committing to a component set for Prototype #1, several investigations were carried out to address the principal unknowns left open by the design constraints (Table 3.1). This section is mostly written in chronological order, but for clarity, some information is presented or combined at different time points than the order may suggest. Four investigative threads are described. The first addressed the foundational engineering problem of this project; the development of a basic analog signal acquisition and wireless-transmission chain to remote storage, upon which all subsequent work was built. The second investigated the physiological acquisition front-end; the selection of target biosignals, the chosen amplification and filtering hardware, and the acquisition electrodes. The third assessed potential anatomical sensor placements for the primary biosignal. Finally, candidate MCUs were examined which in turn ultimately dictated the selection of the wireless communication protocol. Together these investigations established the empirical foundation on which future prototypes were built.

A note on naming: the project was originally called *WakeLift*, an echo of its initial focus on REM reduction for potential treatment of depression. This name appears in some of the early work, such as in user interfaces (UI); the project was renamed OSSMM and that name is principally used throughout this thesis.

3.4.1 Foundational Backbone - Analog Signal to Wireless Data Transmission Chain

The foundational problem of this project was the development of a system for capturing analog signals, digitizing them, and successfully transmitting them wirelessly to an Android platform. The thesis therefore began with the simplest possible model

project: the construction of a controllable analog signal that could be captured by an MCU and transmitted wirelessly through Bluetooth Classic (BC) to an Android application. A $10\text{ k}\Omega$ potentiometer connected to a 5 V rail served as a manually adjustable analog signal source. An Arduino Uno clone (Elegoo Uno R3) served as the MCU, with an input connection from one terminal of the potentiometer to one of its analog-to-digital (ADC) pins. The digitized value was sent from an output pin on the Uno to an HC-05 BC transceiver, which broadcast the data wirelessly. The HC-05 was selected because it is well known in the Arduino development community, works reliably, and performs serial data transmission, allowing easy interfacing with both the Uno and the Android application. This hardware setup is shown in Figure 3.3, with potentiometer voltage fluctuations shown through the Arduino Integrated Development Environment (IDE) in Figure 3.4.

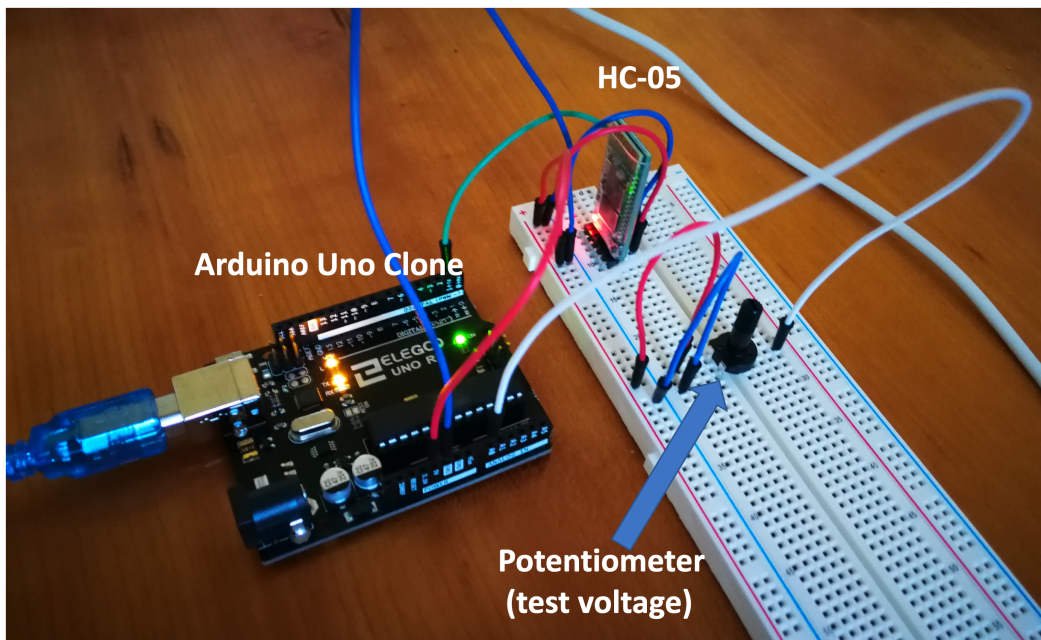


Figure 3.3: Foundational project setup consisting of an Elegoo Uno R3 measuring voltage across a potentiometer and transmitting the digitized value over Bluetooth through an HC-05 transceiver.

A simple companion Android smartphone application was built with MIT App Inventor, a web-based, block-based utility for building Android applications that requires limited programming experience (Patton et al., 2019). MIT App Inventor was selected initially as it permits rapid prototyping, even for those without Android programming experience. As shown in Figure 3.5, the completed setup successfully transmitted live voltage readings from the potentiometer to the Android application, confirming the basic signal-acquisition-and-wireless-transmission chain.

The companion application framework subsequently evolved through two further iterations before settling. MIT App Inventor’s block-based programming proved

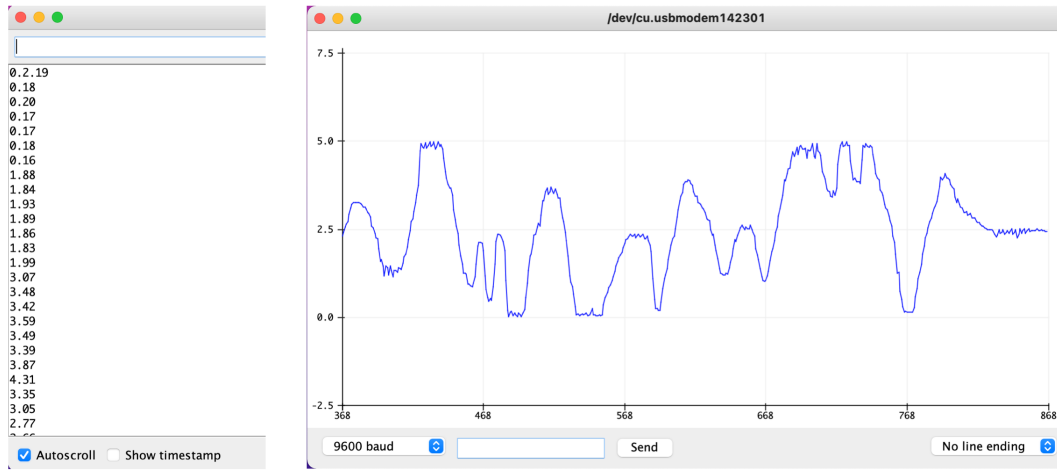


Figure 3.4: Read out of voltage over an adjustable potentiometer shown in 3.3 on Arduino Serial Monitor (left) and Arduino Serial Plotter (right).

cumbersome as functionality requirements grew and it had immediate limitations with BC connectivity and data storage. Kivy, an open-source Python framework, was investigated next, but was abandoned because no working Bluetooth API could be located for it in 2021–2022. Development finally settled on the Dart language and the Flutter open-source software development kit, which became the framework for all subsequent companion application work. Extended details of the MIT App Inventor and Kivy detours are provided in Appendix B, Section B.1.

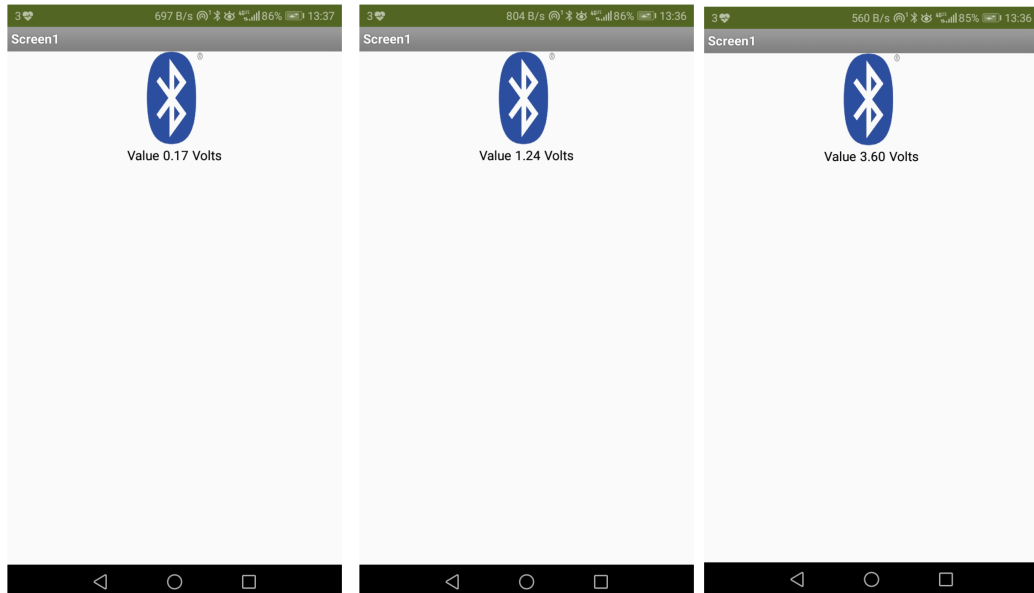


Figure 3.5: Display of momentary voltages on an Android app built with MIT App Inventor.

With the Flutter framework in place, the foundational experiment was repeated on the new application as a sanity check. The next step was to then extend the foundational experiment, and the app’s capabilities, permitting the sampling and

transmission of more than one analog signal. Figure 3.6 shows this extension in which five analog input channels were tested. Similar to the first experiment, a potentiometer was used to artificially simulate a changing analog signal. Five different input pins were used, but with only one of those pins sampling the potentiometer voltage at a given time. If the pins were not in usage, they were connected either to the power rail or to ground. In the right-hand screen of Figure 3.6, all five pins can be observed reaching 5 V at unique periods, underlining a success of the experiment. However, it was repeatedly observed that when the unused pins were connected to 5 V, a sinusoidal interference signal would appear in the data. It was later determined that one source of this interference was due to the HC-05 chip itself.

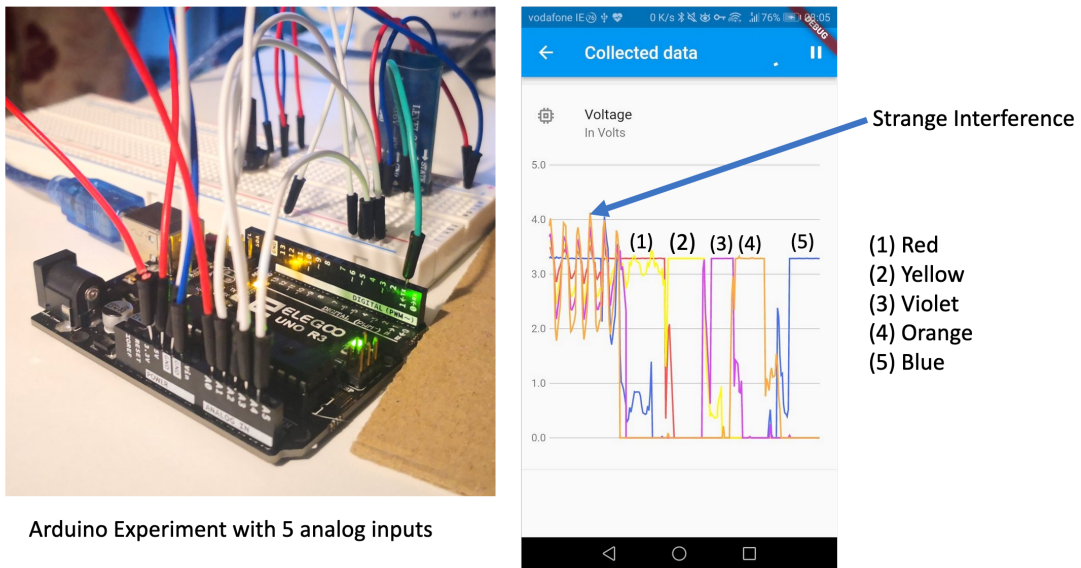
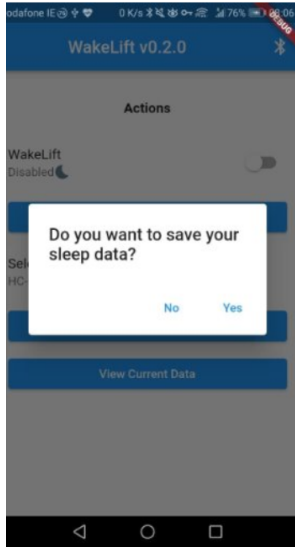


Figure 3.6: Multi-channel analog capture extension of the foundational setup. Five analog input pins on the Uno R3 were sampled, with a potentiometer connected to one pin at a time and the remaining pins held to 5 V or ground. The Flutter companion application reception is shown on the right, with each pin reaching its expected value at distinct periods.

With the system capable of handling multiple analog streams, the next progression was to enable the application to save the transmitted data. In the subsequent version, shown in Figure 3.7, when the user stopped the data transmission from the app, a prompt appeared asking, “Do you want to save your sleep data?” Replying “Yes” would save the data into a Comma-Separated Values (CSV) file. The initial saved data structure consisted of six columns: a single DateTime column marking the reception of each sampling round at the Android device, plus five columns containing the sensor values.



Save Dialog Pop-up
after recording

No.	Time	Sensor1	Sensor2	Sensor3
1	08:04:00.353	3.291	3.477	
2	08:04:00.795	3.315	3.482	
3	08:04:01.295	3.296	3.359	
4	08:04:01.797	3.306	3.291	
5	08:04:02.274	3.301	3.262	

Data saved as CSV

Figure 3.7: Subsequent version of Flutter app with CSV saving capability. In this version, a bug prevented sensor data for channels 3, 4 and 5 from recording.

3.4.2 Biosignal Selection and Component Foundations

With the analog signal acquisition and wireless transmission chain established, three component-level choices were made early in the project that directed subsequent investigations and prototypes: the main physiological signal targeted for sleep staging, the electrode material used to acquire that signal, and potential secondary signals for future investigation.

3.4.2.1 Biosignal Selection - Prioritizing EOG

Because sleep stages are primarily defined using EEG, purpose-built contact sleep monitors typically rely on EEG for staging. In this project, however, EEG acquisition was deliberately avoided due to the anticipated difficulty of obtaining stable, quality signals in an everyday wearable format, particularly without dedicated consumables like conductive gels or Ag/AgCl electrodes (Design Constraint #13). The decision to prioritize EOG over EEG stemmed from the project's original goal of achieving strong differentiation between REM and NREM sleep states for future REM modulation. Rapid eye movements are the hallmark physiological sign of REM sleep, and prior research had demonstrated that lateral EOG signals alone could permit REM sleep detection with relatively high accuracy, albeit in offline classification (Coskun et al., 2016). Focusing on EOG therefore reduced electrode type and configuration challenges while directly targeting the most distinctive physiological feature of REM.

Secondary consideration was given to actigraphy (movement) and heart rate signals. The initial Market Survey (Section 1.4.3) showed that standard consumer wearables, like smartwatches, relied almost exclusively on these two inputs for sleep staging. Though top devices achieved stronger performance by including EEG (like the Dreem headband) or full body movement (like the Somnify), standard Inertial Measurement Units (IMU, for measuring movement) and photoplethysmogram (PPG, for pulse) sensors offered easy, low-cost benefits. It was believed that by incorporating them, the platform should, at a bare minimum, match the baseline performance of standard consumer devices (roughly 60% per-epoch accuracy). However, during this initial investigative phase, the architecture was kept flexible given other sensor modalities could have warranted inclusion later in the project. Though these sensors were identified early, hardware selection for IMU and PPG was performed later, during MCU selection and later prototype development respectively.

3.4.2.2 EOG Amplifier: The Analog Devices AD8232

For EOG signal amplification and filtering, the Analog Devices AD8232 was selected as a suitable integrated circuit (IC). The AD8232 is widely available as part of a ready-to-use Printed Circuit Board (PCB) in the “heart rate monitor (HRM) configuration” (Analog Devices, 2012). Originally popularized by SparkFun and called the “SparkFun Single Lead Heart Rate Monitor,” this hobbyist PCB is typically used for detection of electrocardiogram (ECG) signals using Ag/AgCl electrodes placed on the left and right side of the chest, with a Right Leg Drive (RLD) electrode placed on the lower abdomen or leg (Analog Devices, 2012; SparkFun Electronics, 2014). Although marketed for cardiac monitoring, the AD8232 IC itself is designed for biopotential measurement applications, and the selected bandwidth in the HRM configuration is compatible with EOG signal detection. Typical EOG activity lies between 1–12 Hz, with a recommended band-pass of approximately 0.1–35 Hz (shown here is the maximum range between various sources) (Bhandari et al., 2006; Constable et al., 2017; Kryger et al., 2017). The AD8232 HRM configuration maintains a band-pass range of 0.5–40 Hz, making it largely suitable for EOG work (Analog Devices, 2012).

The AD8232 PCB functions as a differential amplifier, and for eye movement detection operates by amplifying the difference in voltage detected at the electrodes as the corneal-retinal dipole of the eyes changes position (Walczak and Chokroverty, 1994). Importantly, previous research supported the use of these PCB units, as adequate sleep stage detection could be performed when using them for EOG (Paglinawan et al., 2021). SparkFun’s AD8232 boards retail for approximately €20, while clones are available for as little as €2.50. Their low cost, broad availability, and straight-

forward interfacing with MCUs make the AD8232 PCB an appropriate choice for the platform.

3.4.2.3 Electrode Material: Conductive Thermoplastic Polyurethane (CTPU) from Heart Rate Monitor Chest Straps

Although the AD8232 had been successfully used to determine REM sleep in previous research, this had been carried out in conjunction with Ag/AgCl electrodes which are the standard for both EOG and EEG work (Paglinawan et al., 2021). However, using such electrodes would violate Design Constraint #13, which specified that the system should not require any consumables such as disposable electrodes or gels. Therefore, an alternate electrode solution was required.

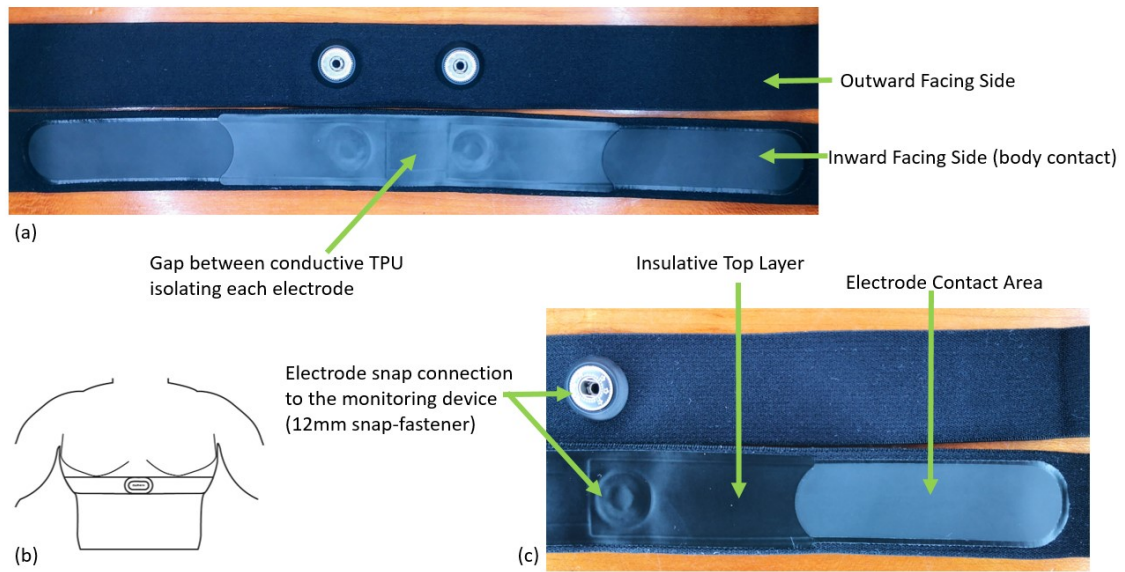


Figure 3.8: Heart Rate Chest Monitor Strap Breakdown

One readily available and affordable commercial option for reusable electrodes was found in heart rate monitor chest straps (HRMCS). As shown in Figure 3.8, these fitness bands contain two conductive Thermoplastic Polyurethane electrodes (CTPU) which sit just below or at the bottom of the pectoralis muscles when worn. They are commonly used by athletes or those interested in fitness monitoring, but have also been used in research for heart rate monitoring (Olstad and Zinner, 2020; Schaffarczyk et al., 2022). The strap consists of an elastic webbing base, typically a polyester and elastane combination, with layers of TPU impregnated onto this base. Figure 3.9 presents an exploded view of one of the electrodes and their layered construction. The TPU layers are attached to each other using ester-based polyurethane heat adhesive films. By sandwiching a layer of CTPU between non-conductive TPU, an inner conductive channel of material is formed and then later exposed in desired

sections to serve as an electrode. Dimensions of electrodes and exposed surfaces vary by brand. Typically, the surface of each of the two electrodes on each strap measures approximately 2.5 cm tall by 7.5 cm long, and they are separated by a distance of approximately 13 cm in the front.

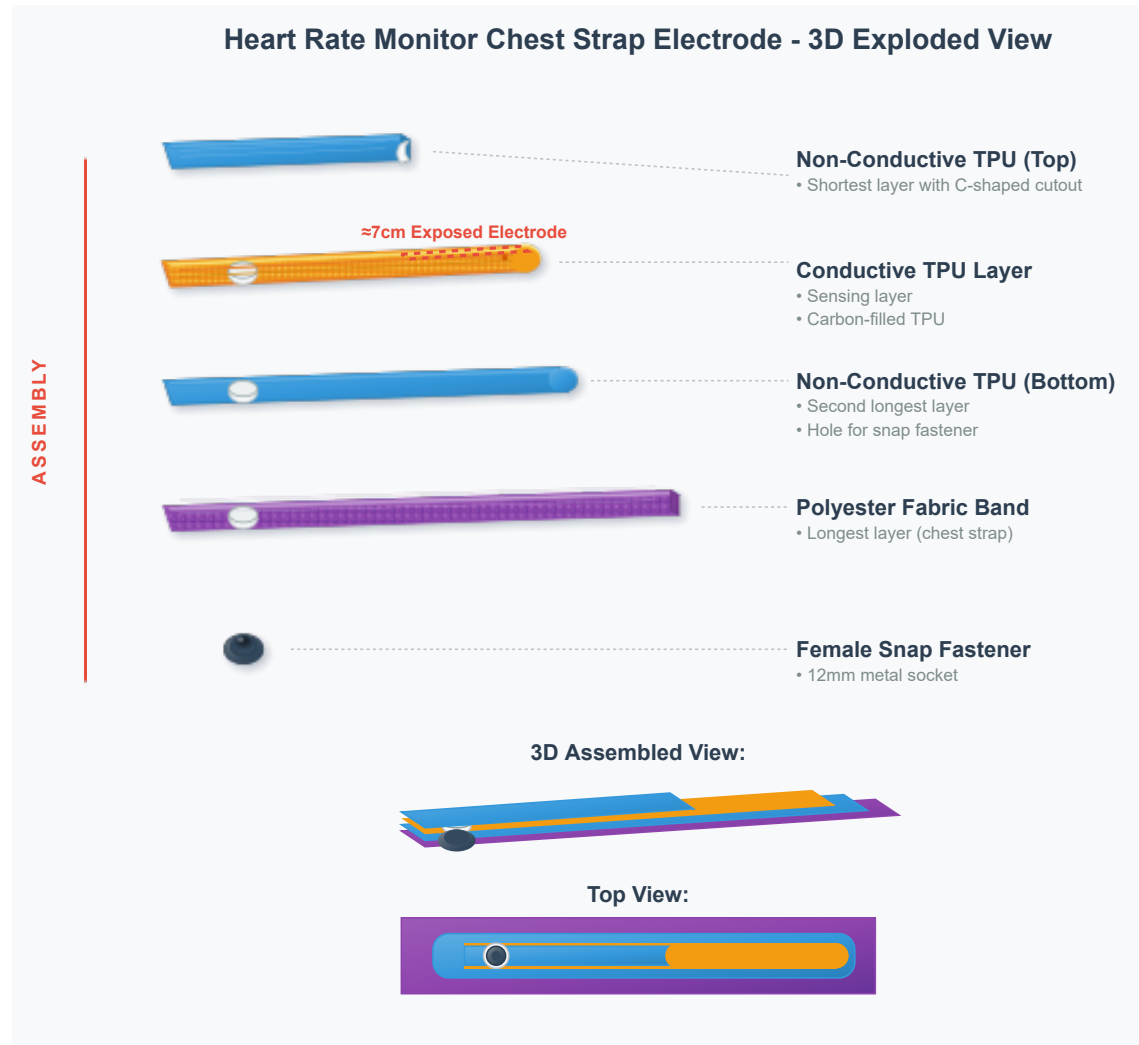


Figure 3.9: Exploded view of a typical Heart Rate Monitor Chest Strap CTPU electrode. Only one electrode is shown in the example. Dimensions and cut-out shapes vary by brand and model. Diagram by the author, based on physical teardown of multiple commercial straps.

The CTPU electrodes on the strap connect to their respective heart rate monitoring device through two conductive female snap terminals (sometimes called EEG, EKG, or electrode snaps) which are located on the mirror side. These electrode snaps are nearly identical in build to 12 mm female snap-fasteners, and are mutually compatible with male counterparts. There are several brand name HRMCS (e.g., Garmin, Polar, etc.) and many generic clones. These companies produce slightly varying versions differing in dimensions and materials, but experience during this thesis found that all tested HRMCS were functionally interchangeable, with

snap-electrodes consistently separated by approximately 44 mm centre-to-centre.

The CTPU electrodes in the HRMCS are known as wet-dry electrodes. They require no conductive gels, but application of conductive gels or water can help improve signal collection. Their conductivity is dependent on the amount of carbon-black doped into the TPU, and they are not purpose-built for EOG detection, rather ECG work, which features stronger biosignals. In short, CTPU electrodes sourced from fitness bands have weaker signal collection capability than dedicated Ag/AgCl electrodes. An investigation was therefore carried out to determine whether the CTPU electrodes in HRMCS were suitable for EOG signal acquisition in combination with the AD8232 PCB. A modified HRMCS was used to create electrode placements as shown in Figure 3.10, that is with an electrode placed above each eye, and the experiment showed that the CTPU electrodes with the AD8232 PCB were sufficient to detect lateral eye movement.

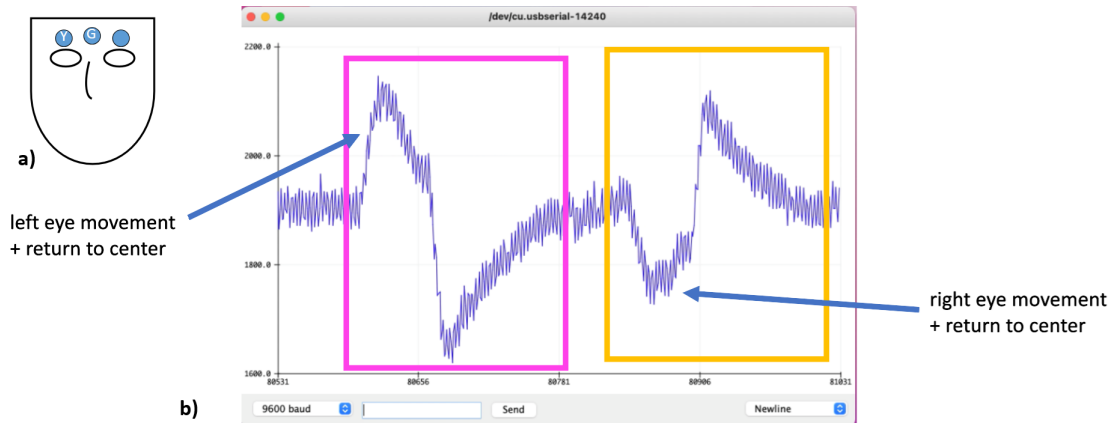


Figure 3.10: Investigation of CTPU electrodes for EOG detection using an AD8232 amplifier and filtering circuit. Subfigure *a* shows electrode placements while subfigure *b* shows Arduino Serial Plotter with highlighted eye movement detections.

Several further factors supported the decision to proceed with HRMCS. From a hygienic point of view, the electrode surfaces can be wiped clean and if deeper cleaning is required, such as for usage between participants, the straps can be washed. Overall, they are soft, flexible, and abrasion-resistant. Compared to fabricating custom metal electrodes, this design avoids the labor-intensive processes of cutting metal and soldering wired connections. Metal electrodes can also cause skin reactions in certain individuals, can be prone to corrosion and oxidation that changes conductivity, and may produce galvanic response due to sweat.

3.4.3 EOG Electrode Placement Investigation

Though previous research had indicated that lateral eye movement alone could be used for accurate classification of REM and NREM sleep states, that work relied on Ag/AgCl electrodes and offline classification with advanced neural networks (Coskun et al., 2016). Although standards for electrode placement in sleep-based eye movement measurement already existed, it was believed necessary to conduct optimization testing (AASM, 2023). Even though preliminary findings showed EOG signals could be detected with the HRMCS CTPU electrodes, it was clear that this set-up yielded weaker signals than standard Ag/AgCl electrodes. It was deemed best to determine the optimal electrode placements for the maximal signal strength. Furthermore, electrode placement in a wearable system is not always straightforward due to the engineering and design difficulties in accommodating the range of human face and head shapes.

Therefore, an investigation was conducted to determine the location or locations that yielded the optimal signal-to-noise ratio (SNR) for eye movement. Results are condensed here for brevity. A total of ten electrode placements were examined, as illustrated in Figure 3.11. The set included placements above the eyes, below the eyes, lateral to the eyes, and various combinations such as above-and-below or lateral-and-below for a single eye. Placement 7 was the closest analogue in the set to the AASM-recommended configuration, with one electrode above one eye and one electrode below the other eye (i.e., one electrode per eye), differing from the AASM standard in that the electrodes were placed directly above and below the eyes rather than lateral to them (AASM, 2023). Mirror placements (i.e., left versus right) were also examined to evaluate any differences in SNR between the eyes. An explicit limitation of this study is that the AASM-recommended configuration was not used as a controlled reference baseline against which the others were compared. However, one major downside to AASM-recommended placement is the need for adjustable electrode positions due to differing facial sizes between individuals. This is straightforward with disposable electrodes, but in terms of engineering design for a wearable device, nontrivial to achieve.

The testing setup consisted of an Arduino Nano 33 IoT MCU, connected to an AD8232 PCB in the heart-rate-monitor configuration. Data collection was performed via USB to the Arduino IDE. Adhesive Ag/AgCl electrodes were used in this study because the CTPU electrodes from a HRMCS are non-adhesive and rely on mechanical pressure (e.g. band tension, foam support, etc.) which is not possible in certain areas (e.g. below the eyes) without a dedicated apparatus. These experiments therefore could not determine how the specific shape, size, and conductive

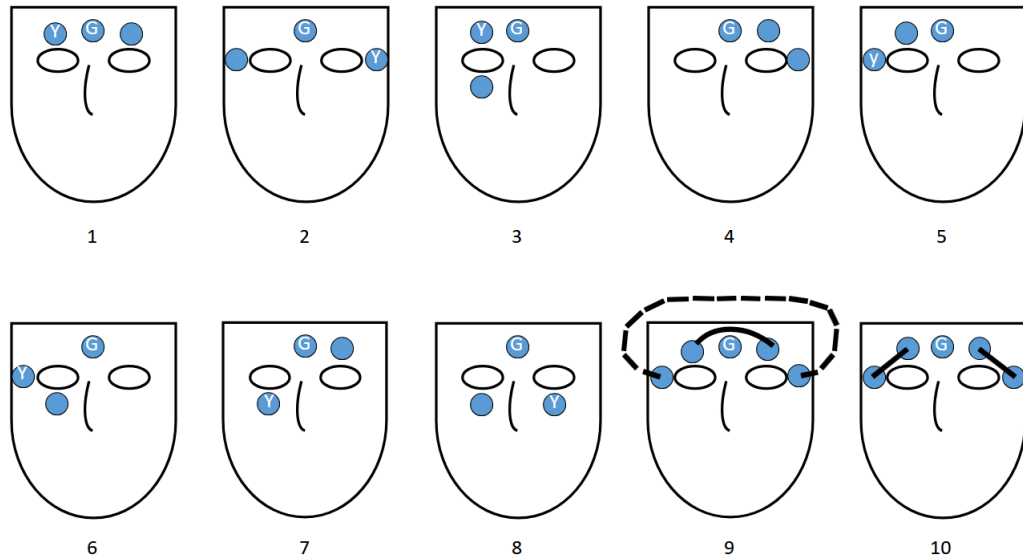


Figure 3.11: The ten electrode placements examined in the EOG placement study. Placement 7 was the closest analogue in this set to the AASM-recommended configuration, differing in that electrodes were placed directly above and below the eyes rather than lateral to them.

properties of the CTPU electrodes would themselves impact eye movement signals when placed in these positions; that question was deferred to prototype testing in which the required mechanical support or tension would be expected to be present. Each electrode location was examined for capture ability of up, down, left, and right eye movements, repeated movement sequences (e.g., left–right–left, up–down–up), and blinks.

The results indicated that optimal placements for vertical movement and blink detection involved positioning one electrode below the eye, with the second electrode either above the eye (placement 3) or laterally (placements 4 or 5). These configurations clearly detected blinks and vertical eye movements, and provided limited detection of horizontal eye movements. Figure 3.12 illustrates a trial using placement 3, demonstrating responses to a blink followed by up, down, left, and right eye movements, and concluding with continuous up–down repetition. Each described eye movement typically generates a complete sine-like waveform, due to two factors: first, the AD8232 functions as a differential amplifier, amplifying the voltage difference between the two electrodes; second, each eye movement involved a return to centre position after each directional movement (e.g., up–centre, left–centre), creating two distinct transitions.

While effective for blink and vertical movement detection, these configurations demonstrated limited sensitivity to horizontal eye movements. For optimal lateral eye movement detection, the strongest responses were found with electrodes



Figure 3.12: Recording of Placement 3 (above and below the right eye). Sequence: blink, followed by up, down, left, and right eye movements, then sustained up-down repetition. Green line shows maximum potential value, yellow line is measured AD8232 values, blue line is unused, second AD8232 input.

positioned either lateral to each eye (placement 2) or below both eyes (placement 8). Surprisingly, placement 8 produced stronger amplitude responses to horizontal eye movements than placement 2. Both of these placements produced virtually no responses to vertical eye movement. Forehead-mounted placements above the eye-brows (placement 1) produced the weakest signals overall; approximately half the signal amplitude compared to lateral placements.

Although signal response quality is important, practical considerations also influence implementation. Without adhesive, electrodes placed below the eyes are difficult to secure reliably. In comparison, electrode placement above the eyes is more practical, as a headband can apply even pressure across the electrodes. Below the eyes, the nose prevents this simple solution. Lateral electrode placements present additional challenges due to variations in head sizes and hairlines. Paradoxically, placement 1, which exhibited the poorest signal response during testing, proved the most practical to assemble and was ultimately adopted for the final release. This disconnect between optimal location for signal acquisition and wearability remained a tension throughout development.



Figure 3.13: Recording of Placement 2 (lateral to each eye). Green line shows maximum potential value, yellow line is measured AD8232 values, blue line is unused, second AD8232 input.



Figure 3.14: Recording of Placement 1. Eyebrow placement produced signal amplitudes approximately half those obtained from lateral placements. Green line shows maximum potential value, yellow line is measured AD8232 values, blue line is unused, second AD8232 input.

3.4.4 Microcontroller and Wireless Protocol Evaluation

MCU selection was constrained by four simultaneous requirements: a small physical footprint suitable for a head-mounted wearable, an on-board IMU to capture actigraphy, sufficient wireless throughput to transmit the planned sensor channels at clinical-grade sampling rates (>250 Hz), and an average current consumption low enough to support over twelve hours of operation from a battery small and light enough to fit inside the wearable (Constraint #2).

The foundational experiment used reliable and cheap components with success, an Arduino Uno clone and an HC-05 transceiver. However, these components were not suitable for the final design. While easy to prototype with, their physical bulk and lack of features, were a limiting factor. The next major design decision was therefore the MCU selection, which is the principal component of the Recorder. Several MCUs were evaluated before wearable prototyping began: the Arduino Nano 33 IoT, two ESP32-based modules from the M5Stack ecosystem, and the Adafruit Feather nRF52840 Sense. Each platform was rejected or adopted on the basis of experimental evidence, and in the course of this evaluation the project’s wireless communication protocol was also selected.

All report current measurements were obtained using the Nordic Semiconductor Power Profiler Kit II (PPK2), which acts as a variable power supply and current sensor in the 200 nA–1 A range with sample rates up to 100 ksps ([Nordic Semiconductor, 2020](#)). The author notes here that throughout the thesis, minor variations in current draw were routinely observed across trials, even with identical set-up and code, likely due to warm-up effects, ambient temperature, and inherent variability in measuring small currents.

3.4.4.1 Initial Wireless Protocol Choice: Bluetooth Classic

The decision to partition the system introduced the question of how the Recorder and Base Station should communicate. WiFi was rejected early because neither of its two implementation methods were without drawbacks. One method required using a shared router, which would add a dependency on uncontrolled external hardware and create a set-up burden and security concerns for users. The second method using SoftAP for a direct connection would force most users off their home network during recording, as most phones only have a single WiFi radio. This left BC and Bluetooth Low Energy (BLE) as the candidate protocols. At this stage in the project it was not clear what either the required transmission speed or the power requirements would be. Given that BC is typically utilized for wireless streaming, and given the desire for the Recorder to capture and transfer high-quality data

sampled at clinical rates (>250 Hz), it was believed that a limitation on data rate transmissions would prove a more pertinent bottleneck on the system design than power usage ([Bluetooth Special Interest Group](#); [Kryger et al., 2010](#)). In other words, the “low energy” characterization of BLE did not appear needed. BC was therefore selected initially, with the added benefit of being a simpler technology to work with. Unfortunately, this decision would prove to be mistaken, leading to significant time costs later, as energy consumption, not transmission speed, was discovered to be the greater challenge to overcome. The narrative below describes how this discovery unfolded across the three MCU platforms.

3.4.4.2 Arduino Nano 33 IoT

The Arduino Nano 33 IoT was the first candidate MCU beyond the foundational Uno setup. It was selected for several promising qualities. Arduino is a recognizable brand with quality products and a large support community, an asset both for continued development and for future researchers using OSSMM. The Nano 33 IoT features an on-chip wireless module, the NINA-W102 (itself containing an ESP32 MCU and supporting Bluetooth), and an on-board LSM6DS3 IMU. Because the wireless module is integrated rather than connected to a separate transceiver, it significantly lowers the risk of disconnection while also reducing the hardware’s volume. The on-board IMU similarly enabled actigraphy capture without additional components. Importantly, the Nano maintains a small footprint ($45\text{ mm} \times 18\text{ mm}$, 5 g with header pins), runs at 3.3 V for energy efficiency, and, at approximately $\text{€}20$, is inexpensive.

Though Arduino advertises the Nano 33 IoT as BC compatible, in practice, the BC capabilities of the board were discovered to be severely limited. The Nano 33 IoT has two processing units: the main SAMD21G, and the NINA-W102. For reasons that still remain unclear, the serial communication between the SAMD21G and the NINA-W102 is bottlenecked. It was found during this thesis that any information first processed by the SAMD21G, then serially sent to the NINA-W102 for BC transmission, never exceeded 112 Bytes/second . Experiments performed at this time were targeting an intermediate goal of $3,125\text{ Bytes/second}$, a 28-fold increase. In comparison, this higher data rate had already been achieved with the more basic Elegoo Uno R3 and HC-05 setup described in [Section 3.4.1](#).

A workaround exists for the limited BC transmission rate, but it requires hacking the board’s firmware and comes at the cost of lost functionality. The fix reassigns the NINA-W102 (i.e. the ESP32) as the board’s main processor rather than only its communication module. While this relieves the bottleneck, the NINA-W102

only has access to a single external ADC input channel which is too limited for the project’s sensing requirements (Guadalupi, 2022). The Nano 33 IoT was therefore ultimately found unsuitable. A more detailed description of the workaround and a verification test are provided in Appendix B, Section B.2.

Although the Nano 33 IoT did not advance in the project, the verification test performed on an ESP32 Lolin Lite indicated that the ESP32 architecture itself showed promise. Rather than continuing with the basic Lolin Lite, which lacks an integrated IMU beneficial for the OSSMM application, the next step was to investigate commercially available ESP32 development boards offering multiple onboard features. Selected results from those investigations are described in the following subsection.

3.4.4.3 ESP32 Platforms: M5StickC Plus and M5 Atom Matrix

A search for suitable ESP32 development boards led to the M5Stack StickC Plus (hereafter the M5StickC Plus), which emerged as a particularly promising candidate. As shown in Figure 3.15, it houses an ESP32-PICO-D4 in a protective casing with a compact form factor of just $48 \times 24 \times 14$ mm. For the UI, it provides a 1.14" LCD, two control buttons, and a power switch. Two onboard sensors support data collection relevant to sleep monitoring: an MPU6886 6-axis IMU and an SPM1423 microphone. The device also includes a 120 mAh internal battery with charging protection, supports standard USB-C charging, and incorporates a buzzer that could serve as a ready-to-use sleep-modulation stimulus. A single unit costs approximately €20 (M5Stack, 2025). All together, the M5StickC Plus met many of the system requirements through its default hardware. The principal remaining concerns were the integration of further sensors and the energy budget.

Switching to the M5StickC Plus marked a shift in engineering focus, as it was the first time the project involved battery operation and current consumption. It quickly became clear that energy consumption would be the most pressing concern for the system design going forward. Running a simple Arduino sketch using BC showed that the M5StickC Plus could not run for more than two hours, far short of the twelve-hour design constraint. For brevity, only the most relevant of the investigations conducted to evaluate and reduce energy consumption are presented here.

Table 3.3: Baseline current consumption of the M5StickC Plus running the “DoNothing.ino” sketch. No wireless activity was performed in this test. Values are 60-second means measured with the PPK2.

CPU Frequency	Configuration	Current Draw (mA)
80 MHz	Screen off, no IMU	31.92
80 MHz	Screen off, IMU initialised	44.04
80 MHz	Screen off, IMU initialised, accel + gyro sampled	34.57
240 MHz	Screen off, no IMU	65.47

library, an adaptive-screen control strategy, and appropriate Bluetooth settings, BLE transmission of IMU and ADC data drew approximately 72 mA ([h2zero and contributors, 2020](#)). At this current, the device’s internal 120 mAh battery would deliver only about 1 hour and 39 minutes of operation. Battery-life projections for the add-on batteries available within the M5Stack ecosystem are summarized in Table 3.4. The M5 TailBat, an external 190 mAh battery that connects via the Grove port for an additional mass of 10 g and a price point of €10, was appealing because it eliminated external wiring and more than doubled the run-time. However, at 4 hours and 18 minutes, this was still roughly a third of the minimum target. The M5StickC 18650C, a 2,200 mAh battery in a HAT-mounted enclosure, was the only ecosystem option that exceeded the twelve-hour target, but its additional 90 g of mass made it too cumbersome for a head-mounted device. Working backwards from the 72 mA draw, reaching the twelve-hour target would require roughly an additional 745 mAh (the 865 mAh row in Table 3.4). The author was unable to source a battery at that capacity at an acceptable weight and volume.

Table 3.4: Predicted run-time for the M5StickC Plus assuming a steady operational current of 72 mA (adaptive-screen transmission state).

Battery Configuration	Predicted Run-Time
120 mAh (internal only)	1 h 39 min
310 mAh (internal + M5 TailBat 190 mAh)	4 h 18 min
865 mAh (internal + 745 mAh external LiPo)	12 h 1 min
2,320 mAh (internal + M5 18650C add-on)	32 h 13 min

The M5 Atom Matrix (Figure 3.16), another M5Stack ESP32 offering, was evaluated in parallel as a more compact alternative. Like the M5StickC Plus, it uses an ESP32-PICO-D4, but in a smaller $43 \times 43 \times 20$ mm form factor, with no internal battery, the

LCD replaced by a 5×5 matrix of coloured LEDs, and a single UI button. Given these reduced features, the M5 Atom Matrix was expected to offer some power savings relative to the M5StickC Plus. Counter-intuitively, it drew more current than the M5StickC Plus across comparable configurations, despite having fewer hardware resources to power, and showed a smaller reduction in current consumption when the IMU was initialized and sampled (45–67 mA across configurations). In the interest of time these results were not investigated further, as the ESP32 architecture was determined to be unsuitable for the project.



Figure 3.16: Front comparison of M5StickC Plus (left) and M5 Atom Matrix (right)

Two further M5StickC Plus issues are noted briefly. A high-frequency “mosquito” tone was audible during operation on both tested units. The intensity of the sound correlated with IMU sampling rate but persisted even with the IMU uninitialized. Given that one of the design objectives was to be as unobtrusive as possible, this would have run strictly against user-experience goals. The sound was confirmed on both M5StickC Plus units tested, making it unlikely to be a one-off problem. Extended investigations of ESP32 sleep modes (Modem Sleep, Light Sleep, Deep Sleep) were also attempted in an effort to extend battery life, but these did not yield sufficient savings to make the architecture viable.

3.4.4.4 Switch from Bluetooth Classic to Bluetooth Low Energy

By this point in the project, two findings from the MCU investigations had re-framed the wireless protocol question. First, the BC throughput on the Nano 33 IoT was severely limited in practice, and the workaround restricted analog input access. Second, the M5StickC Plus and M5 Atom Matrix investigations had revealed

that power consumption, not transmission speed, was the dominant concern: ESP32 platforms could comfortably sustain the required data rate but at currents an order of magnitude above what a small, head-mounted battery could support. With a refined transmission rate estimate of 8 kB/s (or 64 kbps), well within BLE’s capability, and with the energy savings of BLE now identified as essential, the project switched its wireless protocol from BC to BLE.

It is noted that the switch from BC to BLE proved a tedious process as the two protocols are not interchangeable on any level. BC and BLE transceivers, software libraries, and APIs are different, both on the microcontroller and in the smartphone application. The Flutter application accordingly switched to the `Flutter_Blue_Plus` package, and each subsequent MCU evaluated was considered with its respective BLE library (Weinberger and contributors, 2024). In-depth technical descriptions of BC and BLE, including the GATT protocol, Maximum Transmission Unit (MTU), and the throughput calculations that informed the 8 kB/s estimate, are provided in Appendix B in Section B.3.

3.4.4.5 Adafruit Feather nRF52840 Sense

With empirical evidence of the energy requirements of ESP32-based boards, it became evident that a different MCU architecture was needed to advance the project. The Nordic Semiconductor nRF52840 chipset promised both the energy savings and the wireless-communication support the project required. However, despite reaching market less than two years after the ESP32, the nRF52840 lagged well behind it in popularity and development board options. As of late 2022, only a handful of development boards were available, and just two included an on-board IMU: the Adafruit Feather nRF52840 Sense and the Seeed Xiao nRF52840 Sense.

A feature comparison of the two boards is shown in Table 3.5. Although the Seeed Xiao Sense was cheaper (€17–20 vs. €45–50), smaller, and lighter, it lacked the brand backing, technical support, and additional sensors of the Adafruit Feather Sense, all of which were considered important for the project at the time. One deciding factor was the Feather Sense’s female JST-PH connector, an on-board standardized port that allows LiPo batteries to be connected and disconnected via a simple snap-fit. The Xiao Sense, by contrast, requires battery leads to be surface-soldered directly to the board, a delicate task, given how small and closely spaced the pads are, and which if done improperly, can permanently damage the board. For these reasons, the OSSMM project proceeded with the Adafruit Feather nRF52840 Sense, shown in Figure 3.17.

Table 3.5: Feature comparison of the two Arduino-compatible nRF52840 boards available with on-board IMUs in late 2022. Features better for the OSSMM project are highlighted in green.

Feature	Adafruit Feather Sense	Seeed Xiao Sense
Microcontroller	nRF52840	nRF52840
Memory	1 MB Flash, 256 KB RAM	1 MB Flash, 256 KB RAM
Wireless Connectivity	Bluetooth 5	Bluetooth 5
Dimensions	51 × 23 × 8 mm	21 × 17.5 × 3.5 mm
Weight	6.8 g	2.9 g
GPIO Pins	21 pins	11 pins
Built-in Sensors	9-DOF IMU, microphone, humidity, temperature, pressure, proximity, light, gesture	6-DOF IMU, microphone
USB Connector	Micro-USB	USB-C
Battery Management	Built-in LiPo connector & charging circuit	Charging circuit but no connector
Operating Voltage	3.3 V	3.3 V / 5 V tolerant
Price (approx.)	€45–50	€17–20
Documentation	Comprehensive tutorials, guides, and sketches	Adequate but sometimes confusing documentation
Assembly Difficulty	Easy hand soldering	Small soldering pins, surface-mounted soldering for battery

Power consumption difficulties in the project were resolved by the nRF52840. Table 3.6 shows the current draw measured on the Adafruit Feather Sense across four sketches. The “Feather-Sense-Example” sketch is provided by Adafruit and samples data from several on-board sensors. “Beacon.ino” is included in the Bluefruit nRF52 library and broadcasts a BLE advertisement every 100 ms. “Throughput.ino,” also from the Bluefruit nRF52 library, is a BLE stress test that streams data continuously at the maximum rate the connection supports; it therefore represents a worst-case sustained-transmission load rather than a typical operating scenario. The “Arduino-Science-Journal” sketch performs BLE transmission of all sensors at 1,000 ms intervals. In all scenarios, the nRF52840 architecture consumed no more than 10 mA.

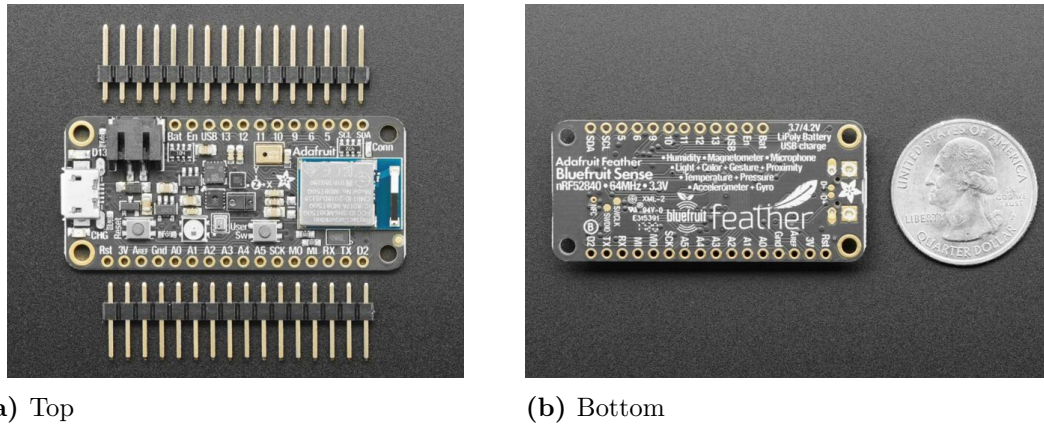


Figure 3.17: Top view of the Adafruit Feather Sense nRF52840 (left) and bottom view with American quarter for size comparison (right). These images display the older model (shown with green silicon), whereas the newer version features blue silicon. Images taken from ([Core Electronics, 2026](#))

Script	Other Information	Current Draw (mA)
Feather-Sense-Example (Adafruit)	Added 2 <code>analogRead()</code> calls, sampling every 300 ms	8.95
Beacon.ino (Bluefruit nRF52 library)	Increased TX power to 8 dBm (from 0)	1.01
Throughput.ino (Bluefruit nRF52 library)	Red LED enabled	8.63
Arduino-Science-Journal (Bluefruit nRF52 library)	BLE transmission of all sensors at 1,000 ms intervals	6.86

Table 3.6: Current consumption for different sketches running on the Adafruit Feather nRF52840 Sense at 64 MHz CPU frequency.

A direct comparison between the ESP32 and nRF52840 architectures is summarized in Table 3.7. Both platforms ran equivalent application firmware, BLE transmission of the IMU and two ADC channels under matched sampling and measurement conditions, differing only in the BLE stack each board supports (NimBLE on the ESP32; the Bluefruit nRF52 library on the nRF52840). At the default 0 dBm TX power the Feather Sense drew roughly one-thirteenth of the M5StickC Plus’s current, and varying TX power between -4 and +8 dBm shifted the draw by just 1 mA. Compared to the order-of-magnitude penalty incurred by the ESP32 architecture this difference was practically negligible.

Microcontroller	TX Power (dBm)	Current Draw (mA)
M5StickC Plus	0	≈ 72
Feather Sense	-4	5.44
Feather Sense	0	5.60
Feather Sense	2	5.93
Feather Sense	8	6.35

Table 3.7: Current consumption with BLE transmission of IMU and two ADC channels under matched conditions. The M5StickC Plus operated at 80 MHz CPU; the Feather Sense operated at 64 MHz. TX power values shown are board-level settings. Power recordings were made at different points in the project. Whole-milliampere resolution was sufficient for the M5StickC Plus, whereas 0.01 mA resolution was required to distinguish the differences in current draw between the Feather Sense settings.

The Feather Sense’s standalone consumption was promising, but the system would ultimately run the AD8232 EOG front-end alongside it, so the combined draw had to be verified before the architecture could be confirmed. As shown in Table 3.8, adding AD8232 EOG amplifier boards to the Feather Sense MCU produced only a modest, sub-linear increase in current draw. The non-linearity was unexpected. It is possible that this examination did not show typical power draw, as the boards’ electrodes were not connected to a user and their free-floating inputs.

Number of AD8232 Boards	Current Draw (mA)
0	5.70
1	6.04
2	6.14

Table 3.8: Current consumption of the Adafruit Feather nRF52840 Sense with zero, one, and two AD8232 PCBs connected. Boards were powered but their input electrodes were not connected to a body during this measurement.

The investigations described in this section established the empirical basis for the component selection for Prototype #1. The CTPU electrode choice, the use of the AD8232 as the EOG front-end, the migration from BC to BLE, and the selection of the nRF52840 architecture all flow directly from the investigations above.

3.5 Prototype #1: Minimum Viable Prototype

3.5.1 Design Objectives and Rationale

The first complete prototype, designated the Minimum Viable Prototype (MVP), explored a sleep mask form factor that departed from conventional headband designs prevalent in sleep monitoring devices. This approach was motivated by an informal survey within the laboratory indicating user concerns about “prone sleeping” or “pillow diving,” where traditional forehead-mounted electronics could create uncomfortable pressure points. The sleep mask concept aimed to serve as a “first pillow,” allowing face-down sleeping while maintaining stable electrode contact and a comfortable user experience. The primary goal of this iteration was to establish a working platform which could collect physiological signals relevant for sleep staging, and confirm the potential use of the selected hardware, specifically the electronic components and the CTPU (Conductive Thermoplastic Polyurethane) electrodes in the real world environment (i.e. off the laboratory bench).



Figure 3.18: Front view of Prototype #1 (MVP) showing the sleep mask form factor with integrated electronics housing (in silver PLA) and a Lyocell fabric covering. Visible are the ON/OFF switch and two windows in the electronic case for LED notifications.

3.5.2 Component Integration

The MCU selected for the MVP was the Adafruit Feather nRF52840 Sense, on the basis of the platform evaluation described in Section 3.4.4. The MVP exploited two of the Feather’s integrated features in particular: its on-board 9-DOF IMU, which captured actigraphy data, and its female JST-PH connector, which permitted direct attachment of a LiPo battery. Power profiling with the PPK2 confirmed the platform did not consume more than 7 mA on average during testing. This evaluation involved

sampling 8 ADC channels (6 from the IMU, and 2 from external pins). The Feather Sense was, however, notably expensive at €45–50.

The AD8232 PCB in the HRM configuration, introduced in Section 3.4.2, served as the EOG front end for the project. The MVP incorporated two such boards, one supporting each of the lateral and vertical electrode configurations identified as complementary in the placement study of Section 3.4.3. Each board’s amplified analog output was wired to a separate ADC pin on the Feather Sense, giving the MVP two simultaneous EOG channels. This dual-channel arrangement was adopted on the MVP on the principle that it is more feasible to eliminate unused channels in later iterations than to retrofit them.

3.5.3 Battery Selection

The initial battery selection followed a systematic protocol based on current consumption measurements from the Prototype #1 circuit, and accounting for potential long term use. Using the PPK2, the system was measured to consume 6.14mA with two AD8232 boards connected and running the standard collection and transmission sketch. The required battery capacity was first estimated using formula 3.1.

$$C_{\text{bat}} = I_{\text{mean}} \times SF \times t_{\text{max}} \times LF \quad (3.1)$$

where

C_{bat} = Battery capacity (mAh)

I_{mean} = Estimated mean operational current over total sleep period

SF = Safety factor

t_{max} = Maximum sleep duration (hours)

LF = Longevity factor

The *Safety Factor* provides a margin to cover uncertainties like load variation, temperatures, and battery aging, ensuring the platform runs throughout the night without failure. The *Longevity Factor* increases the lifespan of the battery, that is, the number of charge and discharge cycles. By creating an intended State-of-Charge (SoC) window, the depth of discharge (DoD) per cycle can be reduced. For example, keeping batteries within a SoC window between 20–80% extends their cycle life. The *Longevity Factor* is calculated as:

$$LF = \frac{1}{SoC_{\text{max}} - SoC_{\text{min}}} = \frac{1}{0.80 - 0.20} = 1.67 \quad (3.2)$$

In modern devices like smartphones, the SoC window is generally managed via software, with cut-offs enforced in hardware. For the hardware considered here, it was ultimately found that a top-of-charge buffer (i.e., $SoC_{\max}$) was not enforceable. To approximate the life cycle benefits without selecting a wastefully or prohibitively over-sized battery a $LF = 1.60$ was adopted. This choice still reduces the nightly DoD by maintaining the $SoC_{\min} > 20\%$.

With a safety factor of 1.30, a maximum sleep duration of 12 hours, and longevity factor of 1.6 to maintain battery health above 20% charge, the minimum required capacity was calculated at 153.25mAh:

$$C_{\text{bat}} = 6.14 \text{ mA} \times 1.30 \times 12 \text{ hours} \times 1.60$$

$$C_{\text{bat}} = 153.25 \text{ mAh}$$

Though the SF and LF calculations indicated a smaller battery would suffice, there was no cost difference between battery capacities which would fit within the permissible volume and shape of the platform's electronic case. Therefore, physical size was the deciding factor and a 3.7 V LiPo 352430 (220 mAh) was selected.

3.5.4 Additional Sensors and Stimulator

The primary goal of Prototype #1 was to establish a MVP. This was defined as a functioning wearable device which could be worn during sleep and collect data. Therefore, additional sensors like the on-board microphone of the Adafruit Feather Sense, a photoplethysmogram (PPG) for capturing pulse information, or a stimulator device (e.g., speaker, vibration motor, etc.) were considered but not pursued at this time.

3.5.5 Mechanical Design and Assembly

The MVP consisted of two main structures: an electronic case and a flexible frame, called the receiver or chassis. The electronic case design featured dimensions of $60.4\text{mm} \times 39\text{mm} \times 16.2\text{mm}$, yielding a total volume of 38.16 cm^3 . The case was produced via FDM (fused deposition modeling) 3D printing using Prusament PLA (Polylactic Acid). All 3D printing for working prototypes was completed with a Prusa i3 MK3S+, and all designs were created in OnShape ([PTC Inc., 2024](#)). As shown in Figure 3.19, the case incorporated support standoffs for the AD8232 boards, dividers for component organization, windows for LED notifications from the MCU, and a slot for the ON/OFF toggle switch. The microphone window was closed in this design due to battery placement blocking access.

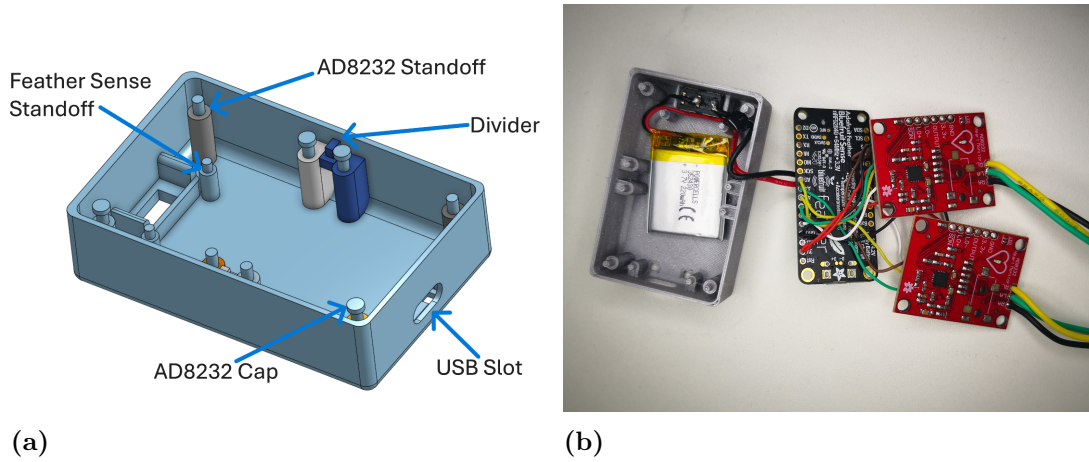
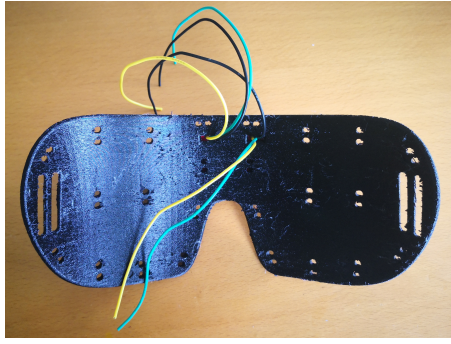


Figure 3.19: (a) CAD view of V0.1.0 of the Electronic Case for Prototype #1. Standoffs for the Feather Sense and AD8232 had different heights to allow the MCU to sit below. (b) V0.0.9 of the Electronic Case with Adafruit Feather nRF52840, 2 AD8232 PCBs, ON/OFF switching, and a 220 mAh LiPo battery. The LiPo battery sits below the MCU, with the two AD8232 PCBs placed on top.

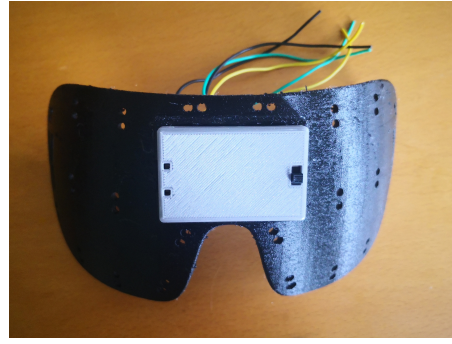
The chassis was also 3D printed but with flexible TPU-95 (Thermoplastic Polyurethane, Shore 95A). The chassis incorporated slots for headband strap attachment, anchor points for foam and fabric attachment, and an integrated slot for the electronic case. Figure 3.20 shows the TPU chassis and the progression of assembly stages.

Two layers of 12.7mm memory foam provided cushioning between the chassis and the user's face. The first layer covered the entire inner surface of the chassis, while the second layer created raised boundaries around the eye sockets. This served to block out light and permit the CTPU electrodes better contact. Lyocell fabric, a hypoallergenic wood-pulp derivative, served as the skin interface and was hand-stitched to the memory foam. This material was selected for its softness, hypoallergenic properties, and cost.

The CTPU electrodes sourced from commercial HRMCS required extensive manual modification. Eight cutting operations were needed to isolate the 5 CTPU electrodes used, followed by bespoke trimming to fit the electrodes to the memory foam. The electrodes were attached using a Velcro system backed with adhesive. This configuration permitted testing of multiple electrode placements including placement with electrodes positioned lateral to and below the left eye for optimal horizontal eye movement detection, and vertical placement with electrodes above and below the right eye for optimal vertical eye movement and blink detection.



(a) Rear of chassis with EOG leads coming through dedicated slots.



(b) Front of chassis with electronic case in position.



(c) Side view of chassis, with headband slots visible.



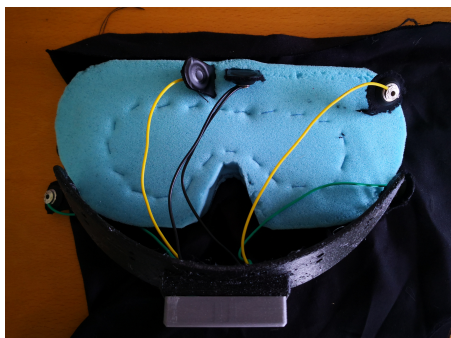
(d) Velcro patches and CTPU electrode cutout.



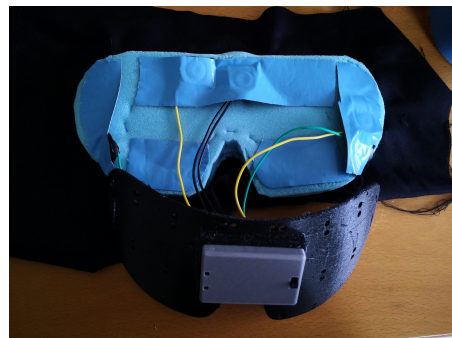
(e) Padded eye socket with a soft loop patch above the right eye position, and hook patch attached to rear of CTPU electrode.



(f) Rear view of memory foam for eye sockets with the snap-fastener connections coming through or around the foam.



(g) Connection between the EOG leads and the CTPU female snap-fastener connections.



(h) Placement of double-sided tape over the electrodes on the memory foam, in preparation for adhesion to the chassis.

Figure 3.20: Views of the construction process of the MVP.

3.5.6 Cost Analysis

The principal component costs and quantities for a single unit of Prototype #1 are shown in Table 3.9. This does not account for construction materials such as wiring, sewing thread, double-sided tape, and Velcro attachments. While the original MVP required over €100 to build, subsequent component sourcing experience identified more cost-effective suppliers. The revised pricing shown in the table reflects these optimized component costs; source notes are provided in Appendix B, Section B.6.

Component	Product	Quantity	Unit Cost	Total Cost
<i>Electronics</i>				
Microcontroller	Adafruit Feather nRF52840 Sense	1	€47.90	€47.90
EOG Amplifier and Filter	AD8232 PCB	2	€2.20	€4.40
Battery	3.7V 220mAh LiPo w/ JST Connector	1	€8.00	€8.00
On/Off Switch	Slide Switch	1	€0.40	€0.40
<i>Materials</i>				
Electrodes	CTPU from HRMCS	3	€2.70	€8.10
PLA Prints	Prusament PLA	9.7 g	€0.03745/g	€0.40
TPU Prints	multicomp Pro TPU-95	91 g	€0.05688/g	€5.20
Fabric	Lyocell	566 cm ²	€0.0016/cm ²	€0.90
Cushion	12.7mm Memory Foam	550 cm ²	€0.006/cm ²	€3.30
Total				€78.60

Table 3.9: Parts Cost for Prototype #1. 3D print masses were estimated at 15% infill with ‘snug’ support using PrusaSlicer. Each HRMCS provided 2 CTPU electrodes. Memory Foam and Fabric sizing was calculated as the minimum rectangular section required before trimming. Wiring, solder, sewing thread, double-sided tape, and Velcro costs are not included. Prices were raised to the nearest 10 cent, and sourced from 2022–2024; receipts and supplier details are provided in Appendix B, Section B.6.

3.5.7 Prototype #1 — Performance Observations

A preliminary trial was conducted on this first prototype to verify its fundamental signal acquisition capabilities. This initial test served as a proof-of-concept to guide the subsequent design process. The trial demonstrated successful capture of physiological signals. Figure 3.21 displays intentional circular eye movement at the start of the trial. Both EOG channels show similar signal amplitudes, and no head movement is observable.

It is important to note that voltages (V) provided within this chapter's plots are estimated based on the AD8232 operating range and should not be taken as ground truth. All observations are provided as 30-second epochs, which is standard in sleep research. Gyroscope channels are not presented here as they typically exhibit similar disturbances as the accelerometer axes in practice.

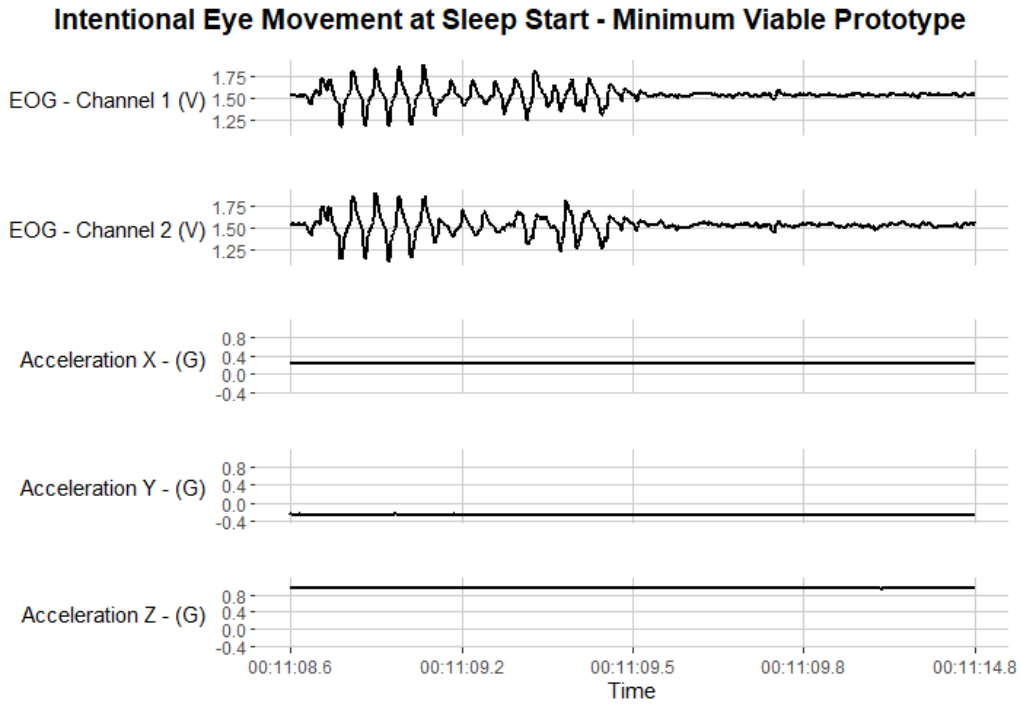


Figure 3.21: Intentional eye movement verification at trial onset showing clear detection of voluntary rotational eye movement, with no intentional head movement. After eye movement stop, the EOG signal stabilizes.

In comparison, Figure 3.22 shows an observation of REM-like sleep captured during the trial. Both EOG channels again show similar signals and amplitudes, while the accelerometer axes show no movement. It can not be determined if this was in fact a REM sleep state, only that the EOG channel detected what is most likely unconscious eye movement. Although there is a significant signal change at 00:54:25, a series of smaller eye movements follow after.

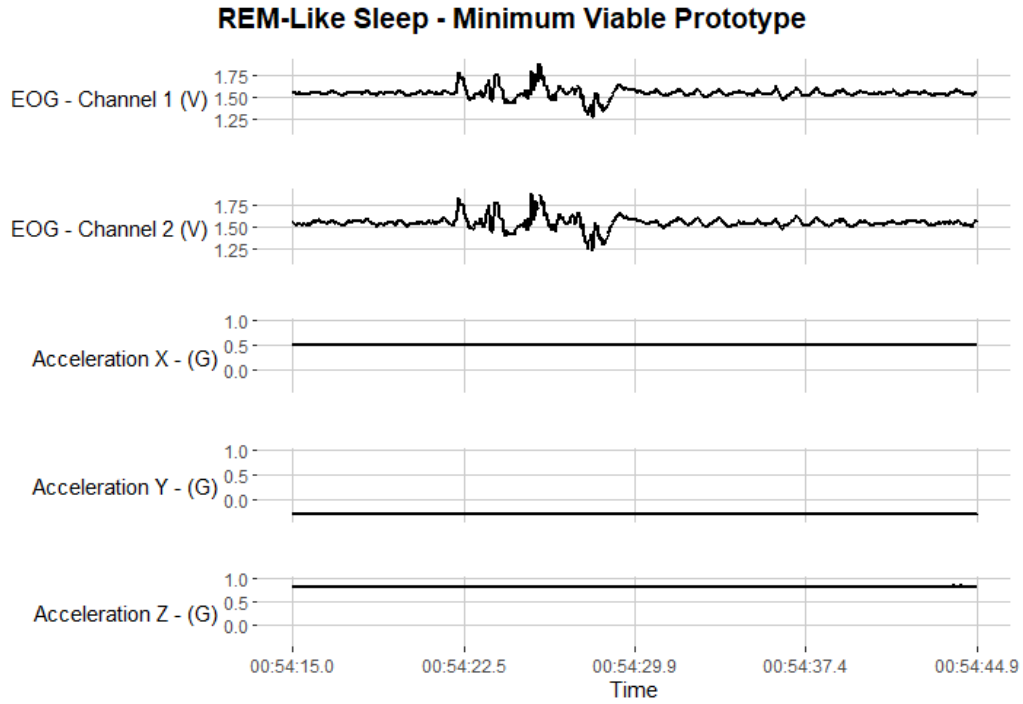


Figure 3.22: REM-like sleep period captured by Prototype #1, showing characteristic rapid eye movements centered around 00:54:25 with amplitudes matching intentional movements at the trial’s start. The signal then shows smaller amplitudes indicative of smaller eye movements for approximately 7 more seconds.

Figure 3.23 demonstrates the effect of movement on signal channels during the trial. All accelerometer axes show the strongest changes at two principal points during this 30-second observation, namely 01:09:35 and 01:09:47. During the first occurrence, the X-axis presents a significant dip, indicating left-right head rotation, and EOG Channel 2 experiences signal saturation. Note the change to the y-axis scale compared with Figures 3.21 and 3.22. This saturation is likely caused by electrodes on one side of the sleep mask being compressed against the skin during the unconscious movement, resulting in altered conductance. The second movement at 01:09:47 produces a similar but lesser effect on the signal channels. This saturation or signal distortion is common in biosignal acquisition, where electrode pressure can alter electrode-skin conductance.

Ultimately the trial found that both electrode placement configurations functioned effectively, with negligible differences in signal quality between the lateral and vertical arrangements. Intentional eye movements produced amplitudes inline with the strongest amplitudes observed during unconscious REM sleep eye movements. Head movements were clearly distinguishable through accelerometer data, though they created substantial EOG artifacts.

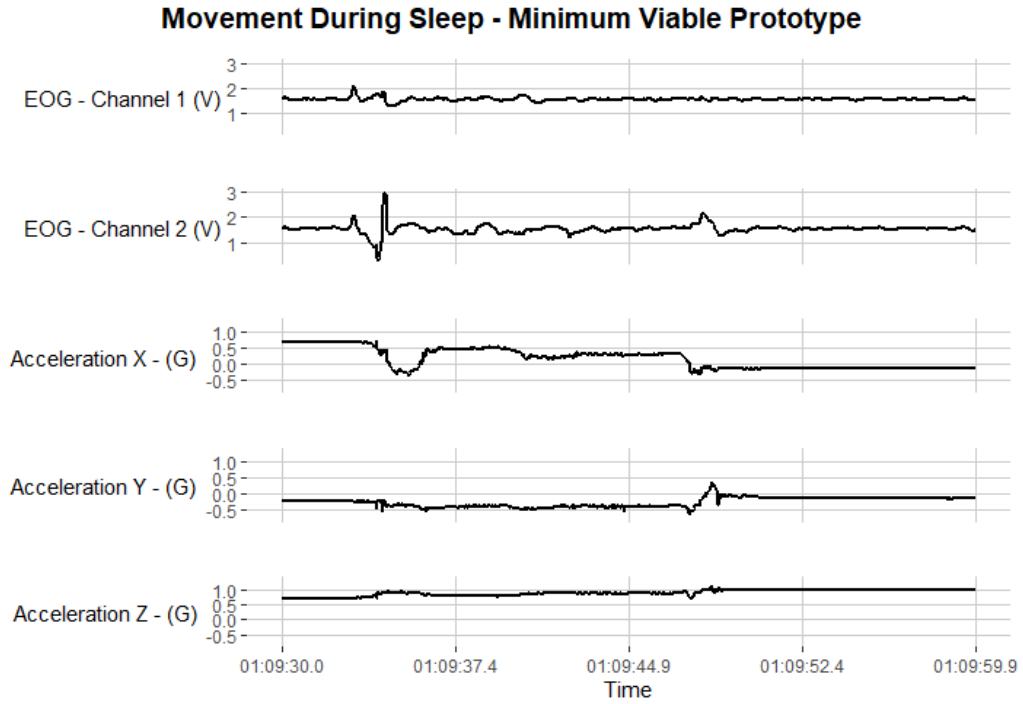


Figure 3.23: Demonstration of motion artifacts showing a synchronized response across the accelerometer and EOG due to unconscious head movement. Changes in the X-accelerometer plot at 01:09:35 are indicative of left-right head movement. This likely caused electrode compression from a pillow, resulting in saturation of the second EOG channel. A second movement at 01:09:47 produced similar effects but to a less extent. These are typical signal artifacts produced by movement.

3.5.8 Summary and Identified Limitations

The testing of the MVP served as an important milestone within the thesis. It was the first time the entire system (electronics, wearable mask, and companion Android app) was evaluated off the bench. The trial established that the selected electronic components, including the CTPU electrodes, in combination with the first working version of the Android application (built for Android Version 9), could successfully capture and record data throughout the night. An exploratory review of the data indicated it could be suitable for automated REM sleep detection, and possibly for stage 4 sleep classification. Finally, the MVP had been produced for under €130, suggesting that this project's price goal of below €65 was feasible.

However, the MVP also revealed several critical limitations that would drive the next iteration. First, the construction process highlighted the difficulties in assembling a usable prototype and the challenges ahead in developing a system which could be easily replicated. The manufacturing complexity presented a significant barrier to research deployment, requiring sewing skills and extensive manual labor that made large-scale production impractical. Assembly proved particularly

challenging, necessitating custom fitting for the memory foam, CTPU electrodes, and fabric attachments. Fabric preparation and sewing required hand stitching due to the complicated geometry of the mask. Manual trimming of CTPU electrodes consistently produced rough edges that were uncomfortable, particularly on the sensitive skin below the eyes. The flexibility of the TPU chassis, while necessary for comfort, allowed the electronics case to dislodge during use. The Lyocell fabric, selected for hygienic reasons, attracted dust, hair, and lint, raising concerns for short- and long-term usage.



Figure 3.24: Rear view of Prototype #1 showing memory foam padding and electrode positioning for comprehensive EOG coverage. This image was from 1 month after construction, and demonstrates the poor reliability of the Velcro adhesive to the Lyocell fabric.

Second, although the mask felt comfortable and balanced, it was noticeably heavy at 191 grams. The thickness also created pressure on the face while turning during sleep against a pillow, with the side sometimes catching on bedding. Finally, the durability proved inadequate. Adhesive failures occurred within one month of assembly as shown in Figure 3.24. Four of the five electrodes detached from their Velcro-adhesive mountings, leaving only the largest electrode with sufficient surface area still adhered.

3.6 Prototype #2

3.6.1 Design Evolution and Objectives

Following the lessons learned from the MVP, the second iteration pursued a fundamental shift in design philosophy. The sleep mask concept had proven too complex for assembly, and too bulky for practical use. Therefore, the primary objectives for Prototype #2 included reducing the system's weight and volume to improve comfort, simplifying the construction methods such as eliminating the need for hand sewing, and incorporating additional functionality, including a PPG sensor for pulse monitoring and a vibration motor for sleep modulation capabilities. While envisioned as a component from the start, it was believed that an additional sensor such as PPG would improve sleep stage classification. Therefore, Prototype #2's development began to consider incorporation for the final release.



Figure 3.25: Isometric view of Prototype #2 showing the headband design with integrated electronic case. The electronic case, in blue, is printed with PLA and the receiver, in black, in TPU-95. Visible on the front of the electronic case is the microphone window, and the new locking tabs to secure the case to the receiver.

3.6.2 Integrated Redesign: From Sleep Mask to Headband

The experience with the MVP revealed that its primary limitations, manufacturing complexity and user comfort, were fundamentally linked. The process of hand-cutting five CTPU electrodes from 2.5 HRMCS units was not only inefficient but also produced rough edges that compromised comfort. Integrating these bespoke

electrodes with multiple layers of memory foam and hand-sewn fabric resulted in a bulky sleep mask that would be difficult to replicate consistently.

To address these issues, the design strategy for Prototype #2 began with simplifying the CTPU electrode sourcing. The goal was to use the commercial HRMCS with minimal modification, thereby eliminating tedious cutting and improving replicability. This manufacturing-first approach, however, constrained viable electrode placements. Using the HRMCS as a continuous band made the electrode placements below the eye and/or at lateral positions exceedingly challenging. Respectively, the strap would either have to cross the bridge of the nose without undue pressure or become adjustable for varying facial widths. The only straightforward configuration was to place the electrodes on the forehead, directly above the eyes. This compromise was a necessary engineering trade-off, given this went against the optimal electrode placements found in earlier investigations (Section 3.4.3).

This decision rendered the entire lower portion of the sleep mask chassis obsolete from a sleep monitoring perspective as it would no longer support any electronics. Consequently, the design shifted from a sleep mask to a headband. The headband design involved a direct reduction of the existing sleep mask, maintaining the dimensions (except height) and curvatures, and was again printed in TPU-95. This new chassis housed the electronic case, guided the HRMCS straps and their folds, and provided a platform for the memory foam to ensure the CTPU electrodes made strong contact to the user.

However, this new configuration required adapting the torso-sized HRMCS to the smaller circumference of the head. As the approximately 13 cm fixed spacing between electrodes is too wide for the forehead, the excess strap length between them was folded inward within the receiver. Figure 3.26 shows an early version of this mechanism where the main HRMCS runs horizontally within the electronic slot of the chassis. In the figure, the strap appears rounded, but when secured by snap-fasteners under tension, it creates four vertical creases. While early versions housed these folds within the electronic case slot, the design was later refined to move them to a section behind the slot. The tertiary electrode, a RLD, was folded around the main strap and memory foam. This entire folded structure was ultimately secured using snap-fasteners integrated into the new electronic case lid. This solution simultaneously managed the strap geometry and provided a direct, solderless electrical connection.



Figure 3.26: Early Version of Prototype #2, featuring a PLA printed chassis, Lyocell wrapped memory foam support, and 1.5 HRMCS. This image shows the snap electrode connections of the HRMCS tucked within the electronic case slot, indicating how the tertiary electrode strap folds around the main, horizontal HRMCS.

3.6.3 Electronic Case Redesign and Integration

The folding of the HRMCS required a redesigned electronic case that incorporated snap fasteners directly into its structure. Three male snap fasteners were crimped through the new electronic case lid, creating rigid mounting points that served multiple functions. First, these fasteners maintained the folding geometry of the chest strap under the tension experienced when worn on the head. Second, they provided direct electrical connections to the AD8232 amplifier board without exposed wiring. Finally, they enabled a modular architecture for the system by permitting quick connection/disconnection for cleaning or part replacement. Figure 3.27 shows a view of the electronic case with aligned male snap fasteners, and an inside view of the electronic case slot on the receiver with corresponding female snap electrodes.

The electronic case dimensions expanded to $86.4\text{mm} \times 39\text{mm} \times 13.2\text{mm}$ to accommodate the planned sensor additions and the snap-fastener interface. While this represented an increase in length compared to the MVP's 60.4mm , it was necessary in order to permit usage of the Adafruit Feather Sense's onboard microphone. Annotated CAD views of Prototype #2 electronic case is shown in Figure 3.28.

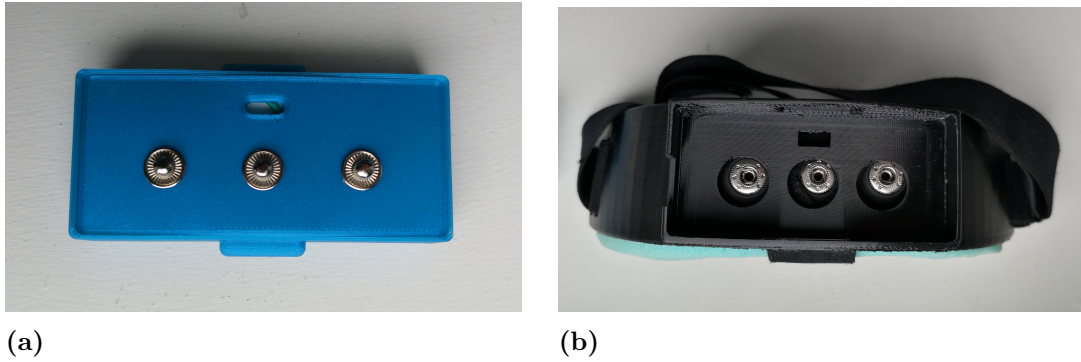


Figure 3.27: (a) The rear of the Prototype #2 electronic case, showing three male snap fasteners that maintain the folded chest strap configuration while providing electrical connections. (b) The inside of the electronic case slot and the respective female snap fasteners from the HRMCS straps.

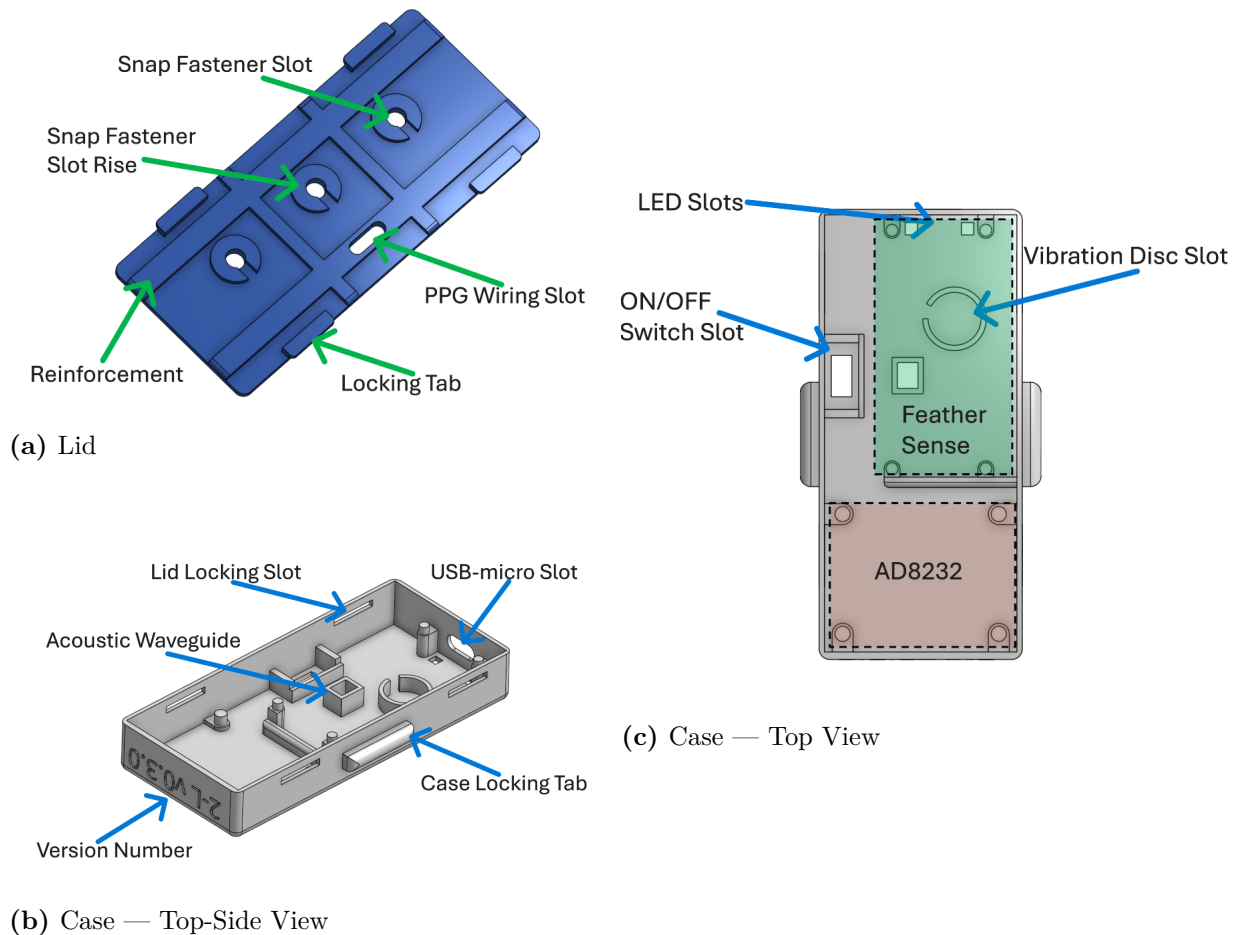


Figure 3.28: Prototype #2 Electronic Case and Lid CAD. Highlighted in (a) is the PPG wiring slot in (b) the Acoustic Waveguide for the on-board microphone and case locking tabs and in (c) the vibration disc slot.

The electronic case now featured its own lid, permitting it to become a modular component. The lid featured a series of locking tabs, asymmetrically offset to only permit insertion in a specific orientation with the case. Three snap-fastener slots were aligned in a row for the EOG connections, each with a horse-shoe shape rise which provided reinforcement to the lid while permitting cable access to the snap-fasteners serving as electrodes.

The electronic case incorporated locking tabs to secure it to the chassis, addressing the previous issue where chassis movement and flexing caused the case to work itself loose. An acoustic waveguide was incorporated into the case design, creating a channel that directed external sounds to the Adafruit Feather Sense's onboard PDM MEMS microphone. This feature anticipated future applications in sleep breathing analysis such as sleep apnea detection, though ultimately it was never utilized during this thesis.

The case also featured a slot to house the 10mm vibration disc which would serve as an example method of stimulation. A vibrator serves as a good stimulator choice for three principal reasons. First, from an engineering perspective, it consumes a significant amount of current. If a system can support this power draw, it can support most other types of stimulation (e.g., audio, visual, electrical). Therefore, it showcases the platform's existing potential for future modification. Second, it is targeted to the user and, depending on platform design, less likely to disturb others compared to sound or light stimulation. Finally, vibration carries fewer safety concerns than other methods like electrical stimulation (e.g., transcranial Direct Current Stimulation).

Using the experience from the MVP which indicated little difference in EOG signals from different CTPU electrode configurations, and the mechanical redesign into a headband, only a single AD8232 PCB was incorporated. This simplified the system and reduced cost.

3.6.4 Manufacturing Simplification and Modularity Improvements

The new configuration required just 1.5 heart rate chest straps and only four straight cuts. This was a significant reduction from the MVP, which needed 2.5 straps and more than eight cuts with further bespoke trimming to isolate the individual electrodes.

The snap-fastener solution eliminated the free-floating wires that had been present in the MVP. Wires soldered to the AD8232 pins on one end, with snap-fasteners

crimped at the other end, could now connect directly through the electronic case backing to the female snap-fasteners of the CTPU electrodes.

A single 12.7mm memory foam layer replaced the dual-layer, eye-cavity form of the MVP. Early prototypes included Lyocell fabric covering for hygiene and safety, but this was ultimately eliminated due to assembly difficulties. While direct contact between commercial-grade memory foam and skin raised health and safety concerns, this was accepted as a temporary compromise for the prototype stage.

The new modular architecture divided the system into five primary components that could be independently replaced: (1) the electronic case assembly, (2) the TPU chassis, (3) the memory foam padding, (4) one complete heart rate monitor strap serving as both headband and electrode carrier, and (5) half of a second HRMCS providing the central “reference” electrode (the RLD). This modularity meant that components could be cleaned between users, replaced if damaged, or upgraded without complete system reconstruction.

3.6.5 Implementation Challenges

Despite achieving the design goal of simplified assembly, several planned features were not successfully integrated into the final version of Prototype #2. The pulse sensor, intended for the central forehead position, was damaged during final assembly when its contact pads were accidentally removed. By the time a replacement arrived, testing had revealed fundamental design flaws that had prompted work on Prototype #3. Additionally, the central “reference” electrode (RLD) occupied the optimal position for PPG sensor placement and this placement dilemma had not yet been resolved.

The vibration motor presented an unexpected technical challenge. The nRF52840 microcontroller’s output pins are limited to 15mA, insufficient to drive even small vibration motors directly. To quantify the motor’s actual current demand, a 10 mm vibration disc was characterized using the PPK2, with the leads of the motor connected directly to the output of the power profiler ([Nordic Semiconductor, 2020](#)). Set voltages were swept from 1.0 V to 3.0 V in 0.5 V increments, and three current measures were recorded at each setting: (1) the 5-second average current under steady operation, (2) the maximum non-starting (i.e., non-transient) current observed during that period, (3) and the peak transient current detected during motor startup. Results are presented in Table 3.10.

Even at the lowest tested voltage of 1.0 V, the average steady-state current of 18.70 mA exceeded the 15 mA per-pin output limit of the nRF52840. At more useful operating voltages of 2.5–3.0 V, average currents climbed to 52–64 mA, with

Set Voltage	Avg Current (5 s, mA)	Max Non-Starting Current (mA)	Max Transient Current (mA)
1.0 V	18.70	41.83	48.19
1.5 V	29.88	66.01	279.72
2.0 V	40.96	93.79	372.92
2.5 V	51.99	114.00	459.02
3.0 V	63.89	137.31	540.00

Table 3.10: Current measurements at different set voltages applied to the 10mm vibration motor disc, measured with the Nordic Power Profiler Kit II. Average current values were recorded over 5-second periods. Max Non-Starting Current refers to the maximum non-transient current observed during that period. Max Transient Current refers to the peak current detected during motor startup.

peak transient currents at startup reaching as high as 540 mA, more than 35-fold the per-pin limit. The vibration motor disc therefore required some type of amplification circuit, with the leads of the motor connected to the power rails of the MCU board rather than directly to a GPIO pin. A transistor amplifier circuit consisting of two 2N3904 transistors was initially designed, but the discrete components and the perfboard required to implement them made the resulting circuit too large to fit within the Prototype #2 electronic case without a further redesign. At a stage in the project where work on Prototype #3 was already beginning, this redesign was not pursued, and the vibration motor was therefore not integrated into the final version of Prototype #2. This oversight in current limitation with the nRF52840 platform stemmed from earlier testing with ESP32-based platforms, which provide higher per-pin output currents and allow for direct pin powering of 10mm vibration motors, highlighting the importance of verifying specifications between platforms.

3.6.6 Prototype #2 — Weight and Cost Analysis

3.6.6.1 Weight

Prototype #2 achieved a total mass of 120.04 grams, representing a 37.1% reduction from the MVP's 191 grams. This improvement came primarily from reducing the amount of memory foam, nearly halving the chassis size, and eliminating fabric.

3.6.6.2 Prototype #2 — Cost Breakdown

Similar to the reduction in weight, Prototype #2 achieved an improved cost of €68.30, a 13.1% reduction from the MVP. This difference again came primarily from the elimination of fabric, reduction of memory foam, and removal of one of the AD8232 PCBs.

Component	Product	Quantity	Unit Cost	Total Cost
<i>Electronics</i>				
Microcontroller	Adafruit Feather nRF52840 Sense	1	€47.90	€47.90
EOG Amplifier and Filter	AD8232 PCB	1	€2.20	€2.20
Battery	3.7V 220mAh LiPo	1	€8.00	€8.00
On/Off Switch	Vertical Slide Switch	1	€0.40	€0.40
<i>Materials</i>				
Electrodes	HRMCS	2	€2.70	€5.40
PLA Prints	Prusament PLA	15.2 g	€0.0375/g	€0.60
TPU Prints	multicomp Pro TPU-95	52.6 g	€0.0569/g	€3.00
Cushion	12.7mm Memory Foam	125 cm ²	€0.006/cm ²	€0.80
Total				€68.30

Table 3.11: Parts Cost for Prototype #2. Actual price for components was higher than shown. Lower prices are shown here based on improved component sourcing experience. Note that 3D parts were evaluated at 15% infill with ‘snug’ support using PrusaSlicer. Each HRMCS provided 2 CTPU electrodes. Memory Foam sizing was calculated at the minimum rectangular section required before trimming. Wiring, solder, and snap-fastener costs are not included. Prices were raised to the nearest 10 cent and sourced in 2022–2024; receipts and supplier details are provided in Appendix B, Section B.6.

3.6.7 Prototype #2 — Performance Evaluation

An observational trial was conducted with Prototype #2 to determine if the new headband design and simplified single-channel EOG with above-eye electrode placements could maintain adequate signal quality throughout the night. The following plots are illustrative examples, selected to demonstrate the system’s signal acquisition ability.

The trial indicated that the simplified single-channel EOG maintained sufficient signal quality. The forehead electrode placement successfully detected lateral eye movements, however, this came at the cost of losing most sensitivity to vertical movements and blinks, which the MVP’s dual-channel configuration could detect.

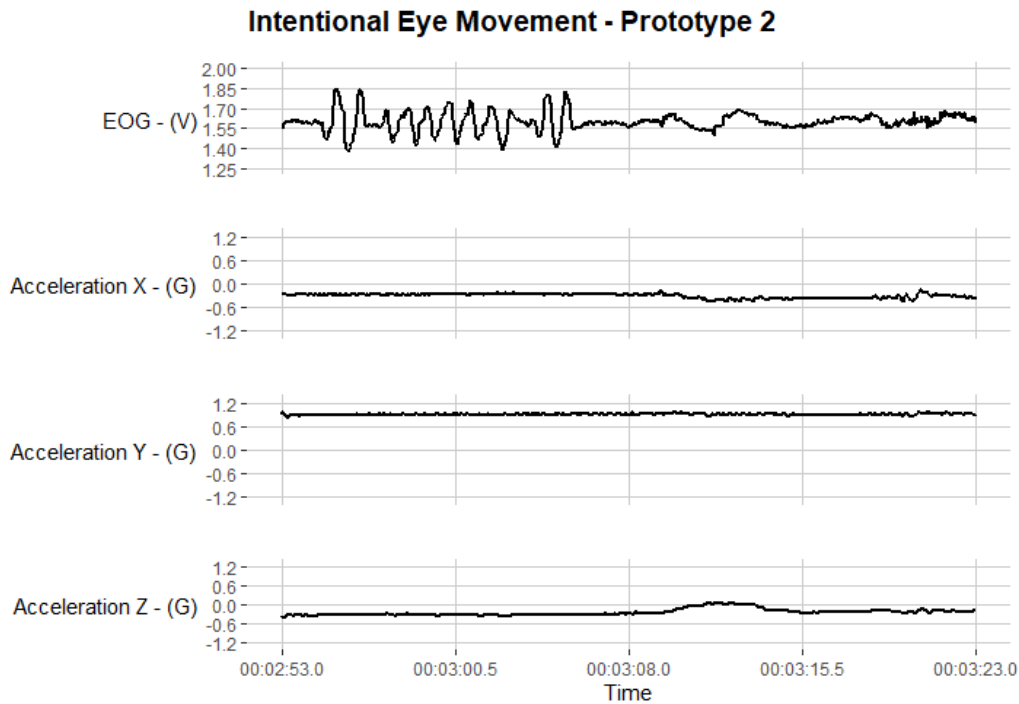


Figure 3.29: Intentional eye movement verification performed at the start of the trial. The EOG channels shows detection of the pattern of left-right saccades, followed by rolling eye movements, then left-right saccades.

Intentional lateral saccades are shown in Figure 3.29, and these movements can be observed within the EOG channel, with no effect on the accelerometer axes. In comparison, Figure 3.30 demonstrates a 30 second observational period which is REM-like in structure with signatures indicating multiple eye movements. The reader will note three “spike” artifacts in the first quarter of the observation, but it is unknown what causes such features. Interestingly, two of the eye movements captured, at 02:05:10 and 02:05:21, present stronger signals than those observed with conscious movements. Typically, conscious saccades show approximately the maximum signal limits. The improved signal strength could be due to compression

of an electrode against the pillow, which is suggested by the positioning show in the accelerometer axes, or sweat which improves skin-electrode conductance.

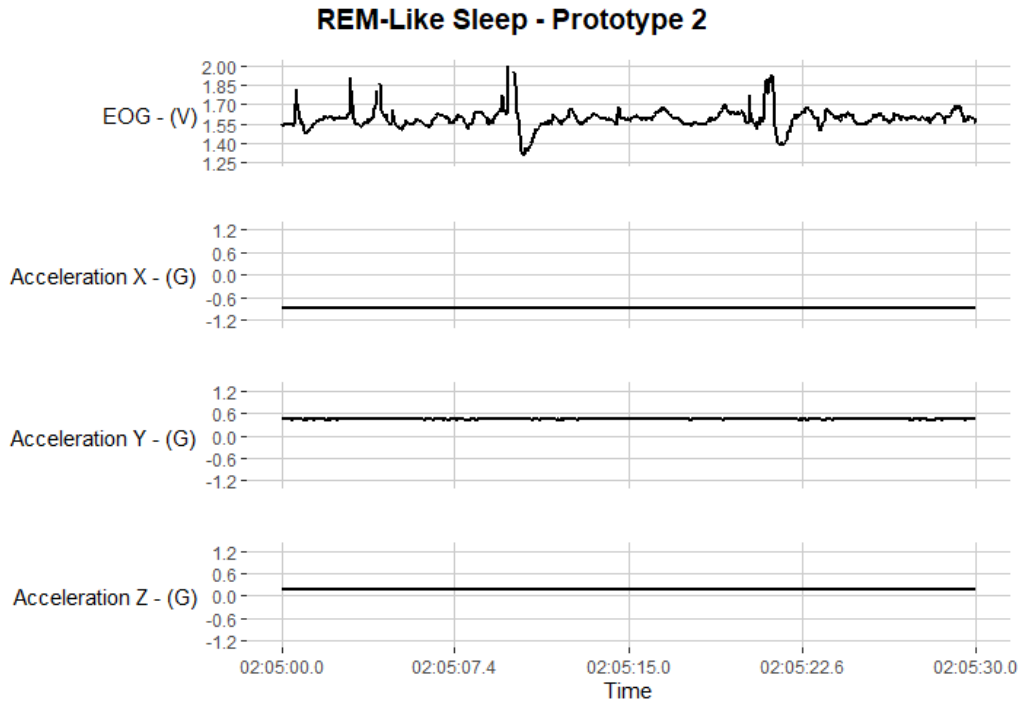


Figure 3.30: REM-like sleep period captured with Prototype #2 showing enhanced signal amplitude due to lateral head position improving electrode contact pressure. Three “spike” artifacts occur within the first quarter of the observation. Two major eye movements occur at 02:05:10 and 02:05:21, with signatures typical of rapid left-right movement evidenced by the peak and trough.

3.6.8 Critical Design Flaws

Despite improvements in manufacturability and weight reduction, Prototype #2 introduced several critical usability flaws. The method used to shorten the CTPU electrode distance, that is, folding the HRMCS, forced the electronic case to sit further from the forehead. Unlike the MVP, Prototype #2 lacked the broad support of a full chassis to counteract the resulting torque. This created a downward pressure on the lower forehead, producing a persistent sensation of the device tilting forward. Furthermore, the increased length of the electronic case proved problematic during side sleeping. The extended profile caught on bedding when the user turned their head, which not only resulted in signal artifacts but, more importantly, was a potential disturbance to sleep. While this risk had been anticipated, the practical impact of a 26 mm increase was underestimated. An unintended gap also formed between the apex of the forehead and the headband. This was caused by a combination of the reduced memory foam and the HRMCS folding mechanism where it connected with the chassis. Critically, this gap prevented consistent electrode contact

and would have rendered the planned central pulse sensor ineffective. Moreover, the complex chassis geometry, a remnant of the sleep mask design, made correct orientation difficult. Its subtle curvature was not immediately obvious, potentially leading to improper fitting and inconsistent electrode contact. This complexity also posed a barrier to future open-source modifications. Beyond the mechanical issues, a key UI flaw was identified. The ON/OFF switch lacked clear position markings, forcing users to guess the device's state or wait for LED notifications to confirm it was active. Finally, the removal of the protective Lyocell fabric introduced a design compromise. Using commercial-grade memory foam raised safety concerns related to direct skin contact, such as potential irritation from off-gassing or other chemicals. This approach was a temporary measure, as medical-grade foam was prohibitively expensive, highlighting an unresolved challenge for creating a safe, skin-contacting surface.

3.6.9 Lessons and Path Forward

Prototype #2 demonstrated that dramatic improvements in manufacturability were achievable through design simplification, validating the modular architecture concept and the elimination of fabric work. The reduction from over eight cutting operations to four, combined with the snap-together assembly marked the path for future designs.

The experience in designing and fabricating Prototype #2 revealed two typical design lessons. First, optimizing individual metrics like weight, sensor capabilities, or assembly time can compromise overall usability of the final product. Second, it is easy to be ignorant of previous design choices that worked because they did not call for attention. The forward weight bias of Prototype #2 and the pillow interference issues were direct consequences of design decisions which could have been informed by these lessons.

These findings informed three critical requirements for Prototype #3: reducing the electronic case profile to prevent bedding interference, reducing the forward positioning that caused weight imbalance, and maintaining the successful modular architecture while addressing the forehead-to-device gap dilemma.

3.7 Prototype #3: Release Version

3.7.1 Design Evolution and Objectives

Despite the successes accomplished in the design of Prototype #2, there was a particular disappointment in the change of design philosophy. Previous designs, such as the MVP, pursued an uncommon approach; integrating sleep-monitoring

(and modulation) into a typical consumer product, a sleep mask. The intent was that if it could be accepted as a sleep mask, it would be more than comfortable enough to function as a sleep monitor. However, this had proved too challenging, specifically in designing the mask in a fashion which lent to reliable and rapid replication. Instead, Prototype #2 focused on practical design choices, spurred by factors of cost, design constraints, and the thesis time line. In this direction it was more successful, but less innovative.

Virtually all purpose-built, contact sleep monitoring devices identified in Chapter 1 (Section 1.4.3) market survey consisted of a headband with the main electronics mounted over the forehead. There are obvious reasons for this common evolution in design choice. The forehead is the prime real-estate providing the most accessible window for capturing physiological markers of sleep (i.e., EEG and EOG). Due to the time pressures of the thesis deadline, it became necessary to create a simple, working system and not let perfection stop the progress of the good. This directed the project further into the headband evolutionary path. The result of this focus is shown in Figure 3.31, and would become the release version of the OSSMM project.



Figure 3.31: Isometric view of Prototype #3 in comparison with a €1 coin, showing the small size and minimalist design.

3.7.2 Mistakes and Progress

A significant design improvement in Prototype #3 stemmed from the discovery of a non-functional electrode connection in previous prototypes. This electrode, previously assumed to be a “reference”, was identified as an unconnected Right Leg Drive (RLD) pin on the AD8232 PCB. This error was due to two causes: first, a misidentification of the RLD pin’s function, and second, an unintentional open circuit created during hardware assembly process.

The issue was uncovered while investigating the effect of this “reference” electrode’s position on EOG signal quality in Prototype #2. Testing revealed that the electrode’s placement had no impact on the signal, and in fact, the signal remained unchanged even when the electrode was completely disconnected. A subsequent hardware inspection confirmed that the PCB trace for the RLD connection had been inadvertently severed during the removal of the 3.5 mm jack, as shown in Figure 3.32.

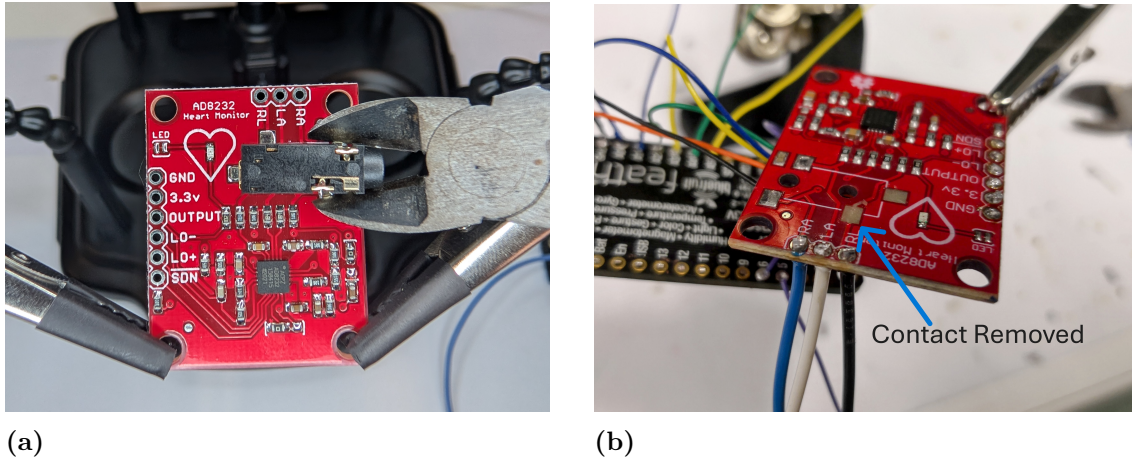


Figure 3.32: (a) The quick-removal method for 3.5mm jack using wire cutters, which occasionally severs the RLD trace but largely maintained functionality. (b) The AD8232 board showing missing surface contact for RLD connection, inadvertently severed during 3.5mm jack removal process.

Typically, the RLD circuit functions as an active noise-cancellation system. It detects common-mode noise present on both input electrodes and drives an inverted version of this signal back to the body, thereby improving the common-mode rejection ratio (CMRR). This is analogous to the technology in noise-canceling headphones. For instance, if both input electrodes detected 50 Hz mains interference (Europe), the RLD would apply an inverted 50 Hz signal to cancel it. Therefore, a functional RLD connection could have resulted in a higher Signal-to-Noise Ratio SNR in the captured data.

Despite the absence of the RLD, the earlier Prototype #2 trial indicated that the EOG signal quality was sufficient by visual standards. This finding yielded several design advantages for Prototype #3. It allowed for the removal of the third electrode, which eliminated half of a HRMCS from the material requirements. This in turn created space on the central forehead for the pulse sensor and helped reduce the device’s profile by removing folded strap layers that had contributed to a forward weight imbalance in Prototype #2.

In retrospect, an initial full appreciation of the RLD’s noise-canceling function might

have discouraged its intentional removal. Although REM-like signatures were visible in data from earlier prototypes, it was initially unclear whether they were adequate for machine learning classifiers. However, it was decided that the presence of these discernible signatures through visual inspection provided confidence that a functional RLD was not essential. Furthermore, removing the RLD connection was not only justifiable from a safety standpoint, but supported. The circuit actively injects a small current into the user. This would have been done at the forehead, a sensitive area, and therefore it was considered prudent to avoid its use due to potential safety risks (however limited) and complications with ethics approval. The current injection aspect of the RLD’s function does not appear to be recognized in published applications of the AD8232, where the electrode is incorrectly described as a “reference”; this is a safety consideration that warrants explicit acknowledgment ([Mendes Junior et al., 2023](#); [Paglinawan et al., 2021](#)).

3.7.3 Microcontroller Migration - Seeed Xiao Sense

Cost, availability, and size constraints motivated another change in MCU selection. Even into 2024, the 2021–2023 global chip shortage resulted in limited stock of the Adafruit Feather nRF52840 Sense in Europe. Beyond difficulty in sourcing units, the cost at €45–50 per board represented a significant price burden in the project which could be reduced without a board change. As described in Section 3.4.4.5, the Seeed Xiao nRF52840 Sense offered comparable capabilities at a €17–20 price point. Though the Feather Sense was originally preferred because of its JST-PH battery port, additional GPIO pins and sensors, and comprehensive support, the experience gained during Prototype #2 showed that board availability, lower pricing and smaller footprint were more important for advancing the project.

However, despite the Feather Sense and Xiao Sense featuring the same nRF52840 architecture, the migration required addressing several platform-specific challenges, beyond simple pin remapping for ADC channels. The main difficulty arose from Seeed’s support for two separate cores in Arduino: an mbed-enabled core and a non-mbed core. A core is a software API and toolchain that the Arduino IDE requires to communicate and compile code for different MCU architectures (e.g., ESP32 vs. nRF52840) or board layouts. In the case of Seeed Xiao Sense, having two cores resulted in differing library compatibilities that were not explicitly documented. For example, the non-mbed core supports Bluefruit, the BLE library previously used on the Adafruit Feather Sense, but it did not support the same library for IMU handling. This meant trial and error to determine which core worked with which libraries, committing to a final core, and sourcing replacements to restore functionality that was lost in the transition. Ultimately, the Seeed Xiao Sense

proved capable of achieving the OSSMM project goals in a smaller package for less than half the cost.

3.7.4 Mechanical Design Optimization

Prototype #3 achieved a successful miniaturization and improved ergonomics through a refinement of both the electronic case and chassis design. The complete redesign addressed all identified issues from previous iterations while maintaining the modular architecture.

3.7.4.1 Electronic Case Refinement

The electronic case achieved substantial miniaturization, measuring just 66mm × 38mm × 13.6mm, yielding a total volume of 34.11 cm³. This represented a length reduction of 20 mm compared with Prototype #2's electronic case, and a 23% reduction in volume. The reduced length acceptably addressed the bedding-catching issue in the previous version. Though it remained 6 mm longer than the MVP's electronic case, it achieved a 10.6% reduction in volume. The new electronic case achieved this smaller package while incorporating a vibration motor board and proven access for a pulse photoplethysmography (PPG) sensor.

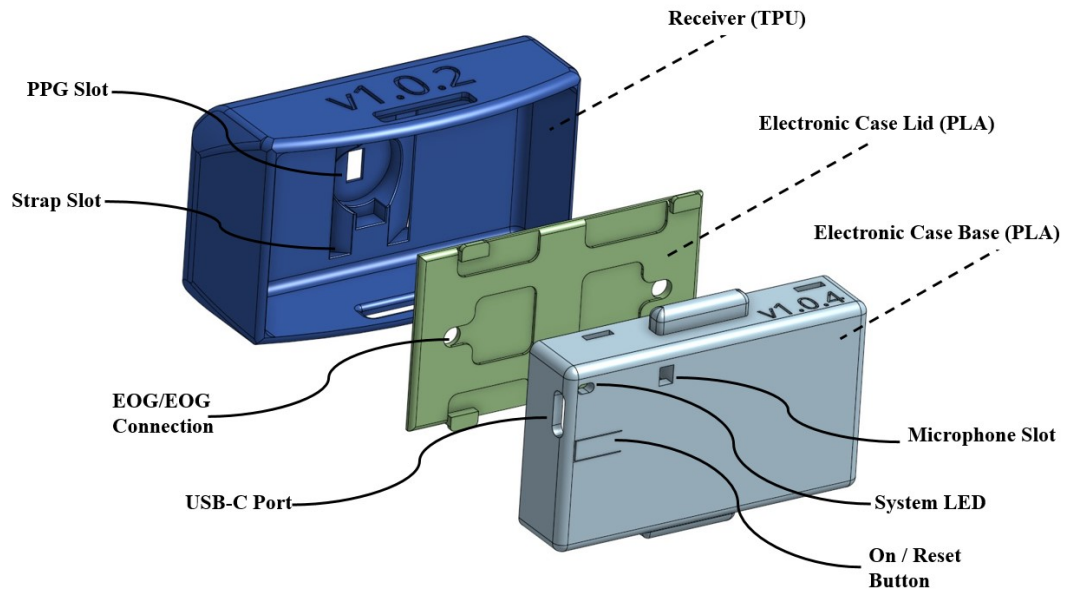


Figure 3.33: Exploded isometric CAD view of Prototype #3 components showing electronic case, lid, and redesigned chassis with annotations for key features.

Figure 3.33 shows an annotated CAD view of both the updated electronic case and the chassis. Another change in the new electronic case was the removal of the ON/OFF switch. This part, although inexpensive, added unnecessarily to the

complexity and assembly of the device. The simple optimization to eliminate this component was achieved by using the RESET button on the Seeed Xiao Sense as an ON button. This button was accessed through a compliant mechanism created in the electronic case PLA shell through a ‘U’ shaped relief cut. Turning the MCU OFF is accomplished through MCU programming, either through recognition of a non-serial USB connection for charging, or through a BLE command sent through the companion App.

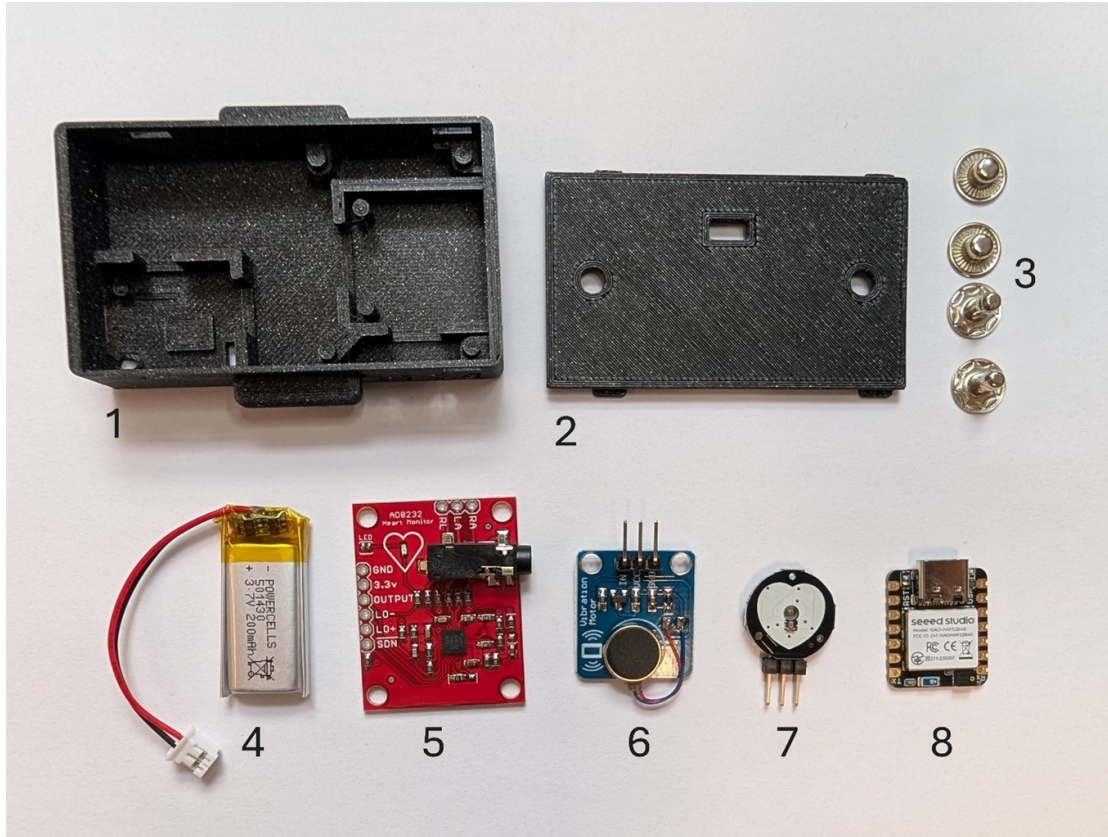


Figure 3.34: Complete component set for Prototype #3's electronic case: (1) Lower case, (2) Lid, (3) Snap-fasteners (x2), (4) 200mAh LiPo battery, (5) AD8232 board, (6) Vibration motor board, (7) Pulse sensor, (8) Seeed Xiao nRF52840 Sense.

A total of nine parts are required to assemble the electronics case (Figure 3.33). To address the 15 mA current limit of the nRF52840 microcontroller, a COTS vibration motor board was selected. These units are inexpensive and readily available. The chosen PPG sensor is based on the PulseSensor design ([World Famous Electronics LLC, 2024](#)). While the official module costs approximately €22, functionally equivalent clone versions can be sourced for as little as €2.20 representing a cost-effective solution for the prototype.

3.7.4.2 Battery Choice

The average current consumption of the final unit measured at 8.23 mA during operation with a sampling frequency above 250 Hz and a transmission power of 0 dBm. This increase was attributed to two main factors: hardware differences between the Seeed Xiao Sense and Adafruit Feather Sense boards, and software updates, principally the implementation of AES encryption for BLE transmission. A recalculation of C_{bat} using Equation 3.1 (Section 3.5.3) shows that a 205 mAh battery was recommended for this increased current consumption.

$$C_{bat} = 8.23 \text{ mA} \times 1.30 \times 12 \text{ hours} \times 1.60$$

$$C_{bat} = 205.42 \text{ mAh}$$

Ultimately, a 3.7V 200 mAh 501430 battery was selected. Although a battery with a larger capacity (up to 250 mAh) could be physically accommodated within the enclosure, the 200 mAh option was deemed sufficient. This decision is justified because the C_{bat} calculation is sufficiently conservative.

3.7.4.3 Receiver Redesign

A simplification of the headband was achieved by eliminating the memory foam padding. Inspired by 3D-printed flexible structures, a cushioning section was integrated directly into the chassis as a replacement. This integrated cushioning was created by printing the chassis with low-density (8%) ‘gyroid’ infill geometry. The gyroid structure provides near-isotropic force distribution, while the use of a softer TPU-85 filament further enhances comfort. This design change eliminated the need for additional materials and the associated manual trimming and fitting processes.

The chassis “wings”, which were included in Prototype #2 to apply tension over the memory foam, were also removed. Practical testing revealed they were not required to ensure adequate contact between the CTPU electrodes and the skin and were therefore deemed redundant.

The receiver module was redesigned to incorporate a dedicated slot and window for the PPG sensor. The external view of this window is shown in Figure 3.35. This slot and window design provided a stable position for the sensor, prevented any external wiring as per Design Constraint #2c, and protects the PPG circuitry while exposing only the essential LED and photodiode. Furthermore, the rear surface of the receiver

was designed with a subtle convex profile to ensure consistent sensor-skin contact and enhance user comfort.

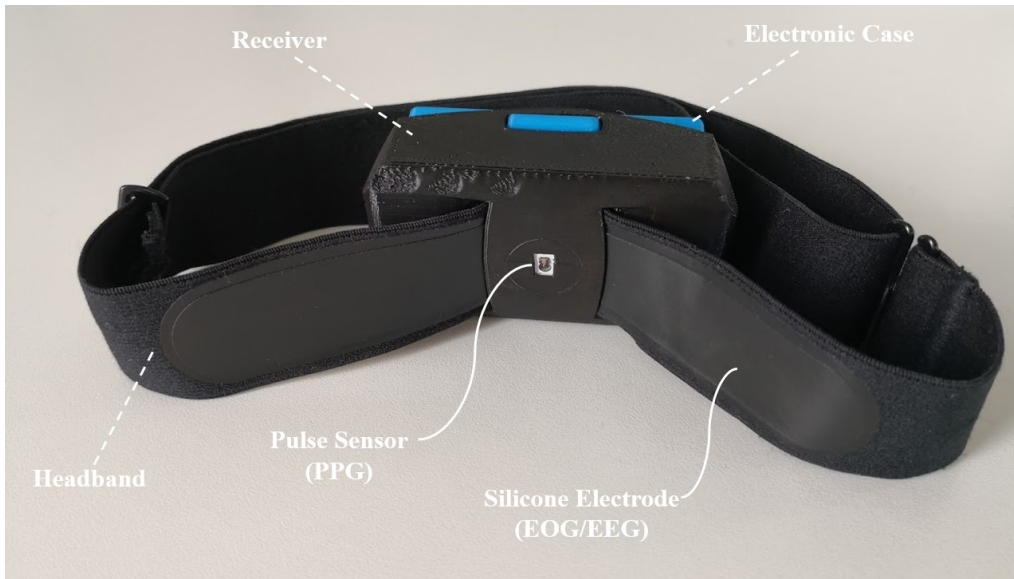


Figure 3.35: Annotated rear view highlighting the major components assembled. It is clear to see how the OSSMM captures sensor data from the EOG and the PPG along the forehead from this orientation. In the upper-left corner of the receiver it is possible to see 3 printing artifacts in the layer lines. Such errors do not affect the experience of the system.

3.7.5 Assembly, Manufacturing, and Accessibility

One of the central goals of the OSSMM project was that the system be straightforward for others to build and use. This subsection addresses both halves of that goal in turn: the assembly and manufacturing process for Prototype #3, and the technical competencies a user would need to construct the system from the released open-source files.

3.7.5.1 Assembly and Manufacturing

The final assembly process achieved remarkable simplification. The system is constructed from three primary modules: (1) the 3D-printed electronics case and lid, (2) the 3D-printed receiver, and (3) a single HRMCS, which is adjusted to size by 2 cuts. This revised approach eliminates the need for bespoke fabrication, although minor post-processing of the printed parts may be required depending on the 3D printer's quality.

As shown in Figure 3.36, the design is modular, comprising three primary components that can be easily interchanged: the electronic case, the receiver, and the HRMCS. Excluding 3D printing time, which varies by machine, a complete headband can be assembled from these parts in less than two hours.

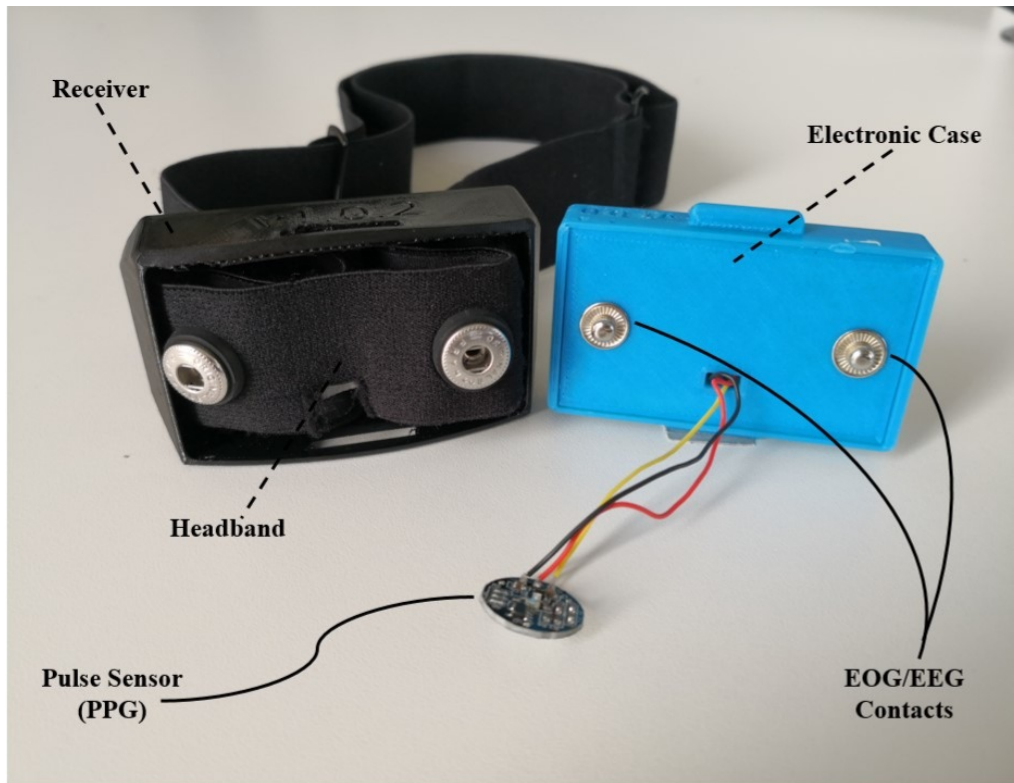


Figure 3.36: Annotated view of Prototype #3, separated in major components, highlighting the primary modules: the Electronic Case, the Receiver, and the headband (HRMCS). The notched cut-out in the strap, originally designed to accommodate sensor wiring, was later found to be unnecessary and is omitted from the final design.

3.7.5.2 Accessibility

Technical and knowledge accessibility for assembly is high. Assembly requires the following competencies:

- The ability to operate a 3D printer with a provided STL file.
- Basic soldering (through hole) for electronic components.
- Familiarity with the Arduino IDE to upload firmware to the MCU.
- The ability to install an Android application (APK file).

Assembly is intended to be accessible to undergraduate engineering or computer science students and hobbyists with basic maker experience. Free tutorials on 3D printing, soldering, firmware uploading, and Android app installation are widely available online for those new to any of these areas.

3.7.6 Prototype #3 — Weight and Cost Analysis

3.7.6.1 Weight

Due to the significant redesign of the chassis, particularly in the elimination of the side wings and the removal of cushioning, Prototype #3 achieved a mass of just 76.6 grams. This was a 36.2% reduction in weight from Prototype #2.

3.7.6.2 Cost

The final cost of Prototype #3 successfully met the project’s financial targets. The total unit cost was approximately half of the €65 target (Design Goal #5) and only one-third of the €130 maximum cost limit (Design Constraint #5). A detailed cost breakdown is presented in Table 3.12.

Table 3.12: Final bill of materials for Prototype #3. Note: solder, wiring, and snap fasteners are excluded. Prices have been rounded up to the nearest €0.10 and include shipping. Prices were sourced in 2024; receipts and supplier details are provided in Appendix B, Section B.6.

Component	Product	Quantity	Unit Cost	Total Cost
<i>Electronics</i>				
Microcontroller	Seeed Xiao nRF52840 Sense	1	€18.20	€18.20
EOG Amplifier	AD8232 PCB	1	€2.20	€2.20
PPG Sensor	PulseSensor Clone	1	€2.20	€2.20
Vibration Motor	Vibration Motor Board	1	€3.00	€3.00
Battery	3.7V LiPo 200 mAh	1	€7.80	€7.80
<i>Materials</i>				
CTPU Electrodes	Heart Rate Monitor Chest Strap	1	€2.70	€2.70
Electronic Casing	Prusament PLA	14.51 g	€0.03745/g	€0.60
Housing Attachment	Siraya Tech TPU-85A	29.01 g	€0.03755/g	€1.10
Total				€37.80

3.7.6.3 Brand names, clones, and safety

It should be noted that several components in the final prototype, including the AD8232 PCB, the PPG sensor, and the HRMCS, were not brand name parts. Counterintuitively, testing revealed that these non-brand ‘clones’ often performed equivalently or were more suitable for this project’s requirements than their brand name counterparts. For example, a generic AD8232 PCB functioned more reliably than the equivalent from SparkFun, although this thesis only examined small samples ($N = 2$) of these. Furthermore, the Polar Pro Strap, a widely used brand name HRMCS in cardiac research (Olstad and Zinner, 2020; Schaffarczyk et al., 2022), was found to have the highest surface resistivity of all CTPU electrodes tested (approximately 500Ω). Quantitatively, its surface resistivity was approximately 3.5 times higher than that of the most conductive generic straps evaluated.

The cost difference between brand name and generic components is significant. For instance, a SparkFun AD8232 PCB retails for approximately €20, whereas a generic equivalent can be sourced for €2.50 or less. Similarly, the €39.99 price of a single Polar Pro Strap exceeds the total material cost of an entire OSSMM unit assembled using non-brand components.

The use of non-brand components necessitates careful consideration of safety, particularly for parts with inherent risks (e.g., batteries) or those in direct dermal contact. A cautious approach was therefore adopted, involving either sourcing from reputable suppliers, despite higher costs, or conducting pre-emptive testing. For example, the TPU-85 filament used for the chassis was sourced from Siraya Tech, a manufacturer that provides biocompatibility and safety datasheets for that specific filament (Jiangsu Kerbio, 2024a,b,c). For other components like the HRMCS, it may be possible to source from vendors who can supply material requirement documentation like Polar. As an additional precaution, participants should be offered the opportunity to test materials for potential skin reactions before a study commences.

It is important to note that even established brands are not likely to provide data and safety certifications for components. Polar, for instance, does not offer dermatological safety certification for its Pro Strap as it is not classified as a medical device, but these are still used in human research (Olstad and Zinner, 2020; Polar Customer Care and Maria, 2025; Schaffarczyk et al., 2022).

3.7.7 Prototype #3 — Performance Observations

The final release version of the OSSMM system, Prototype #3, underwent a multi-night proof-of-concept experiment, the full quantitative results and systematic evaluation of which are detailed principally in Chapter 5 and Chapter 6. The purpose of this section is to provide an exploratory analysis of the multimodal data the system can capture. Therefore, the following plots are illustrative examples, showcasing the EOG, PPG, and actigraphy data in combination with sleep stage labels provided by a reference sleep monitoring device, the Somnify SM-100. The SM-100 has been validated against PSG with a Cohen’s Kappa of 0.62 for 4-stage sleep classification (Wake, Light Sleep (N1+N2), Deep Sleep, and REM) (Toften et al., 2020).

The signal quality of Prototype #3 exceeded expectations. In addition to the intended EOG signals for eye movement, the device also appears to capture frontal EEG signatures. Notably, visual inspection of the data revealed occasional waveform patterns consistent with slow waves (deep sleep) and sawtooth waves (REM sleep), although formal verification is not possible without a reference PSG (polysomnograph). Visual spectrogram analysis shown in Chapter 6 also provides evidence for capture of sleep spindles. The PPG sensor also provided reliable pulse information.

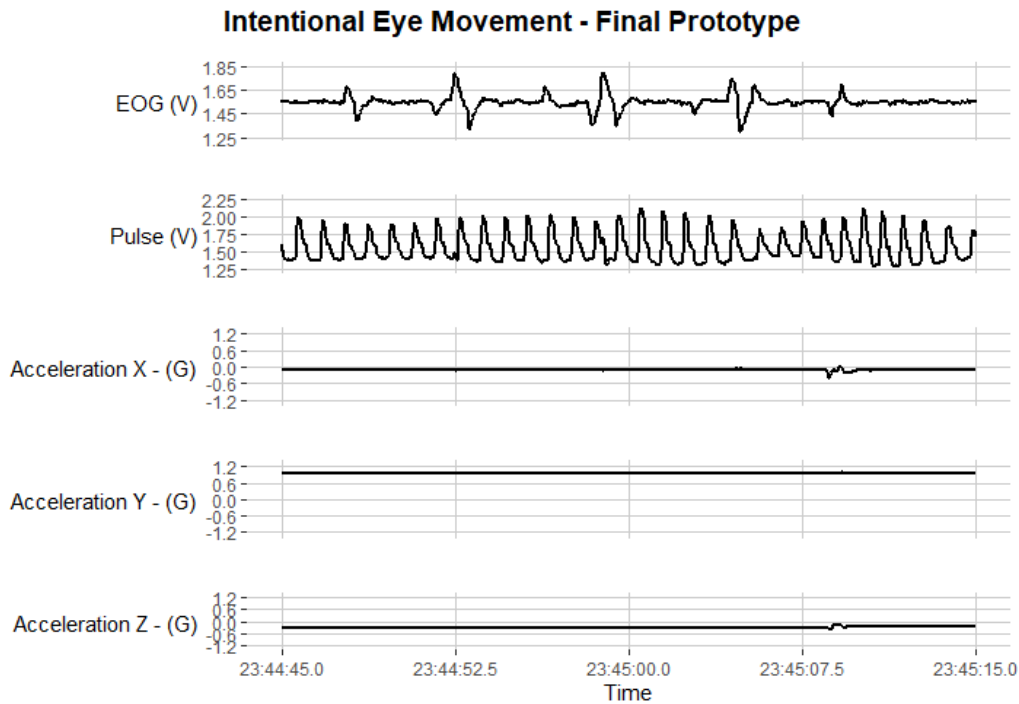


Figure 3.37: Intentional eye movement verification showing distinct saccades in EOG channel and clear pulse waveform detection in PPG sensor. This 30-second observation is from Night 1 of the experiment.

Figure 3.37 shows intentional eye movements performed at the start of Night 1. The EOG channel shows a mix of lateral saccade movements, while the PPG shows a steady heart rate of approximately 62 BPM (beats per minute).

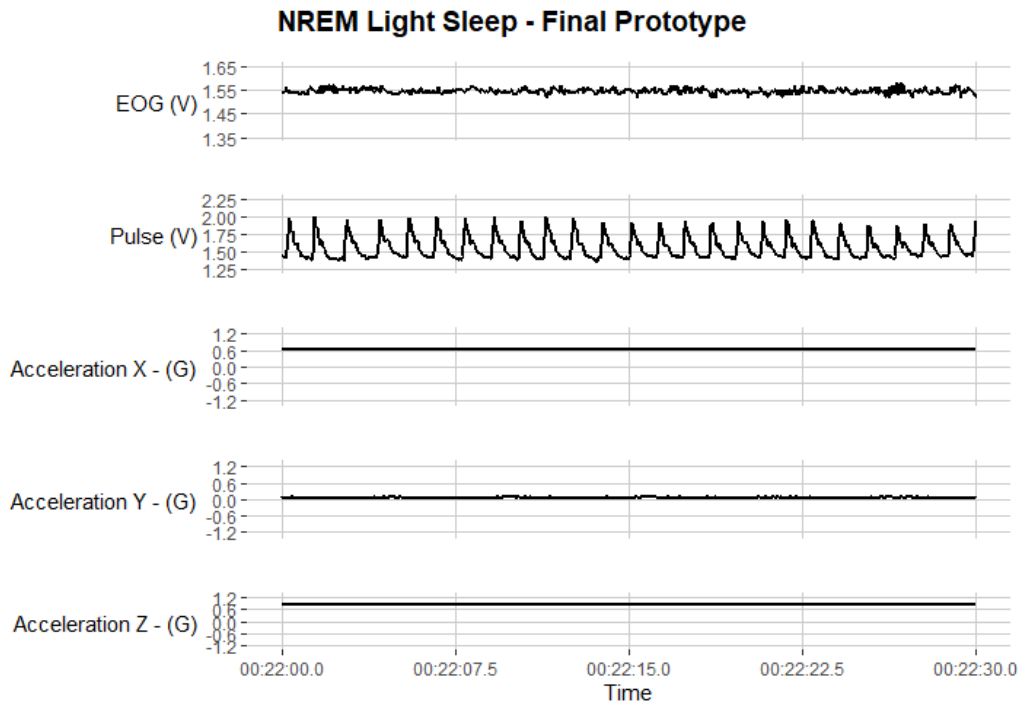


Figure 3.38: This is a 30-second observation collected from Night 1 of the experiment. The SM-100 labeled this period as “Light Sleep,” or N1/N2 NREM sleep. The EOG channel shows minimal activity, and the Pulse channel and stable heart rate of approximately 52 BPM.

Figure 3.38 shows a 30-second observation during a time period classified as NREM “Light Sleep” by the SM-100 unit. There is minimal activity recorded by the EOG and accelerometer channels, and a steady pulse of approximately 52 BPM. In contrast, Figure 3.39 presents a 30-second observation of NREM “Deep Sleep.” Here, the EOG channel shows increased waveform activity. The waveforms observed here in the EOG channel are consistent with EEG delta waves (0.5–4Hz) which occur during slow wave sleep (SWS), that is “deep sleep.”

Typically, delta waves occur in an anterior-posterior direction, but it is possible to observe them across hemispheres due to their global nature (Murphy et al., 2009; Uygun and Basheer, 2022).

Visual inspection of Prototype #3 data revealed equal success for REM labeled sleep. Figure 3.40 presents a 30-second observation with some minor head movement in the first quarter, three subsequent eye movements, and in the last quarter a notable change in heart rate. These identifiers are all characteristics of REM sleep. Minor head twitching is not uncommon, and heart rate variability is correlated with REM

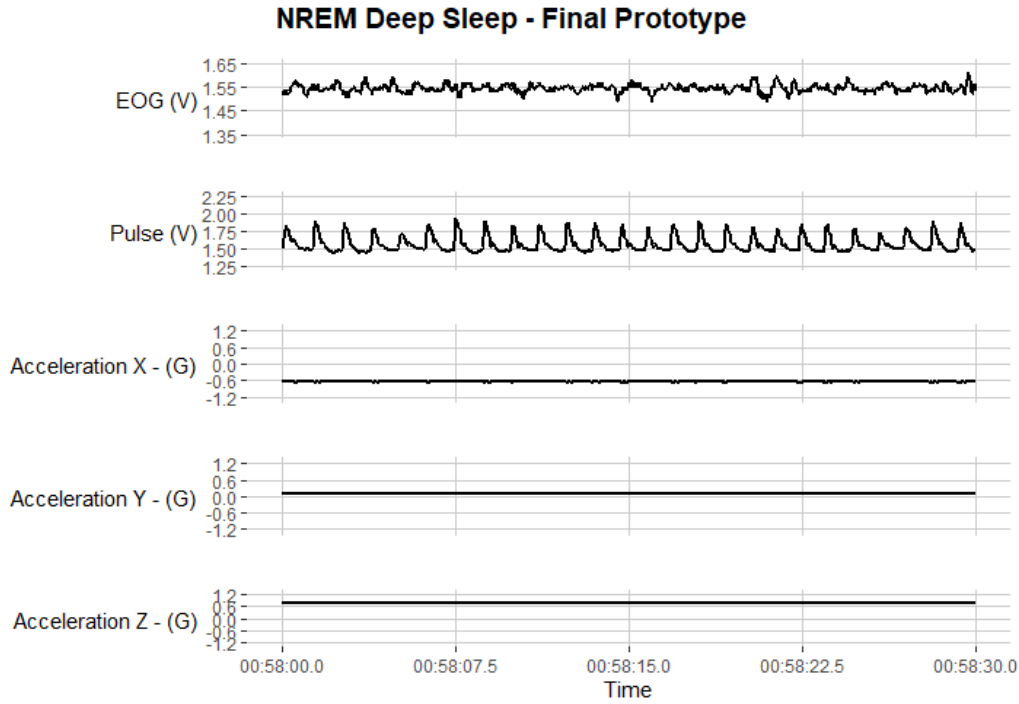


Figure 3.39: This is a 30-second observation collected from Night 1 of the experiment. The SM-100 labeled this period as “Deep Sleep,” or N3 NREM sleep. There is indication of slow wave EEG activity occurring through the EOG channel with a frequency of 0.8 – 0.92 Hz. The PPG sensor indicates a heart rate of 52 BPM.

sleep periods. This observation is notable in that during the first three quarters (22.5 seconds) the heart rate is approximately 76 BPM, then suddenly slows to $\approx$ 40 BPM.

During Night 1 of the experiment, the author noted waking at 5:45 from a dream in the Modified Consensus Sleep Diary (Carney et al., 2012). Exploring the data at this time point revealed the awakening took place at 5:43 AM, where it is then clear significant movement occurred due to a change in the accelerometer values and artifacts in the EOG and pulse channels (i.e. a waking event). Notably, during this observation is an approximately 2 Hz wave in the EOG channel. This is consistent with a phenomenon known as sawtooth waves which occur between frontal hemispheres during REM sleep (Frauscher et al., 2020). However, a PSG will be required in future work to confirm EEG capture of this phenomenon.

As a contrast to the presented results, Figure 3.42 displays unconscious movement during sleep and the effect on the sensors. Accelerometer axes show significant movement, and the PPG sensor shows minimum and maximum saturation.

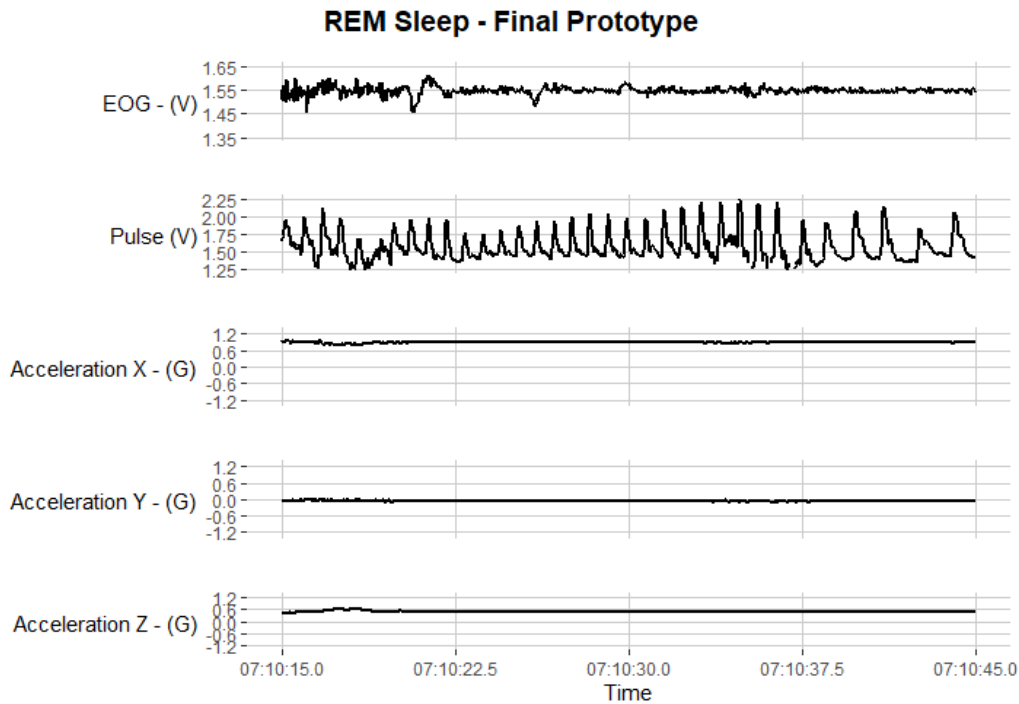


Figure 3.40: This is a 30-second observation collected from Night 1 of the experiment. The SM-100 labeled this period as REM sleep. The last quarter of the observation shows a marked change in heart rate variability, with heart rate dropping from 76 to ≈ 40 BPM.

It is important to highlight that the above observations are illustrative and must be interpreted with their selection method in mind. They are neither ideally representative nor truly random samples. When selecting an observation for presentation, a 5–10 minute window was first randomly chosen from within a larger classified sleep stage (e.g., REM, Light Sleep, etc.). This short window was then manually inspected, and a single 30-second epoch was selected. Consequently, while the initial search location was random, the final selection was subjective and limited to a very small portion of the total data. These examples serve to demonstrate the system’s capabilities but are not presented as systematically validated representations of each sleep state. The exception to this method is the observation corresponding to the recalled awakening from a dream which was noted in the sleep diary.

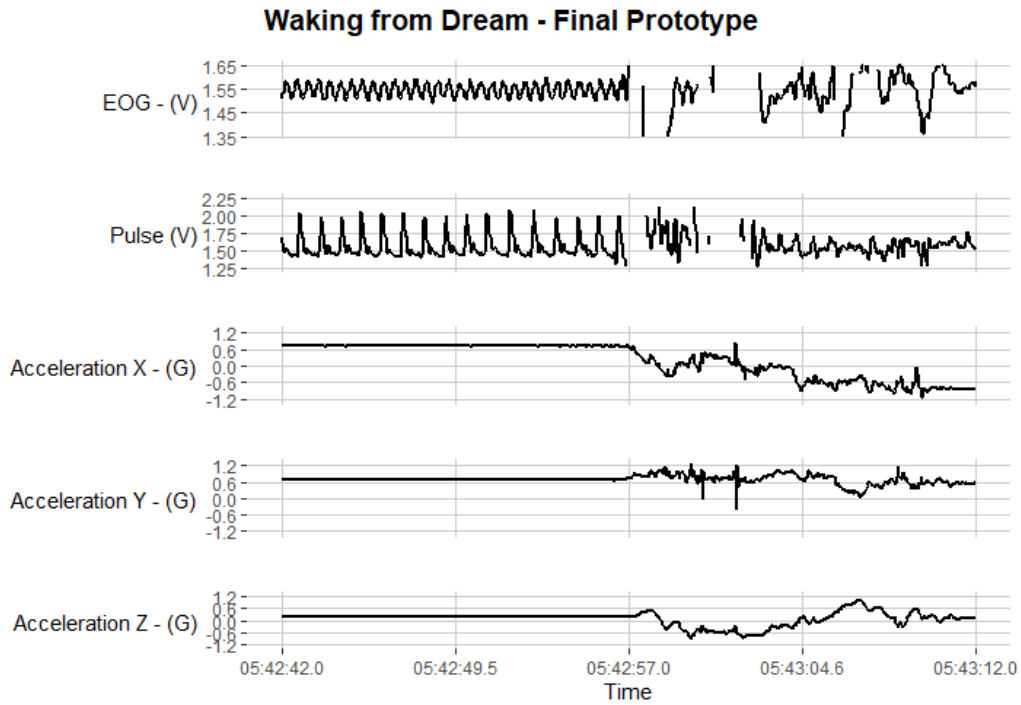


Figure 3.41: Dream period at 5:43 AM showing sawtooth waves in EOG channel, followed by awakening at 5:45 AM with characteristic signal disruption across all sensors.

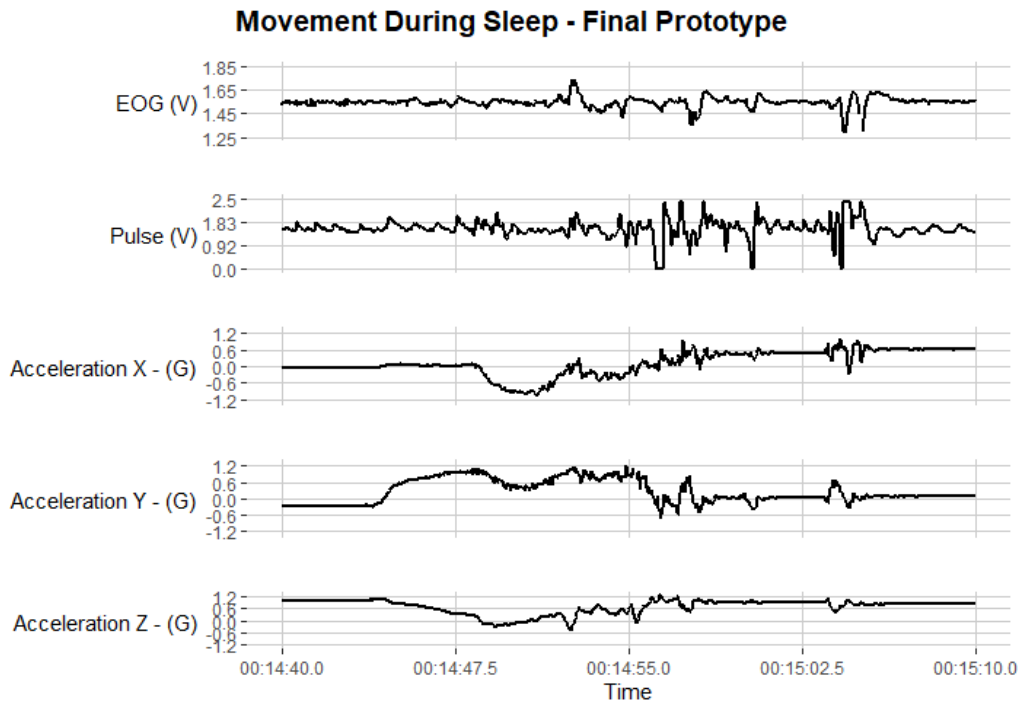


Figure 3.42: This is a 30-second observation collected from Night 1 of the experiment. Movement during sleep showing coordinated changes across all sensor channels with PPG saturation.

3.7.8 Identified Flaws and Future Work

While Prototype #3 represents a significant advancement, further hardware refinements are possible. As detailed in Chapter 5, a key research question was whether the utility of the PPG sensor justified its added cost and complexity. It was realized during Prototype #3 development that if the PPG sensor were omitted, the TPU chassis could potentially be eliminated entirely. This change would reduce the device’s overall bulk and weight and potentially improve user comfort. It would also make the OSSMM system more accessible, as printing TPU requires a 3D printer with a direct drive extruder and is more challenging to print compared to PLA. Ultimately, inclusion of the PPG presents a design trade-off, as pulse data is valuable for certain sleep studies. Future work could address this by developing two distinct open-source forks of the headband: a lighter, more streamlined version without the PPG sensor, and a more comprehensive version that includes it.

The performance of the vibration motor was also found to be suboptimal. Vibration was selected as a stimulus due to its targeted nature, but the haptic feedback was significantly dampened, likely by the folded HRMCS strap. As a result, the vibration was more audible than tactile. Further investigation is also required to determine if this stimulus can effectively influence sleep in its current form or if an alternative solution is needed.

3.8 Comparative Analysis and Design Evolution

The progression from the initial MVP to the final release version was an iterative process of refinement, driven by practical testing, and design discoveries. This evolution was fundamentally a journey from a complex, conceptual design to cost-effective, robust and accessible system. This section provides a comparative analysis of the three major prototypes, and evaluates the final prototype against the project’s original design constraints and goals. Figure 3.43 shows all final prototype versions side by side for comparison.

3.8.1 Evolution of Key Metrics

The design process yielded substantial improvements in cost, weight, and functionality, while dramatically simplifying material requirements. Table 3.13 demonstrates a comparison of the three working prototypes, charting the changes in cost, size, sensors and other metrics. The data therein makes clear a consistent trend towards miniaturization, cost-efficiency, and expanded capability.



Figure 3.43: View of all prototypes. In the left-rear is Prototype #1, the middle Prototype #2, and in the front-right Prototype #3.

Table 3.13: Final comparison of OSSMM prototypes showing evolution across key metrics. Current consumption values measured during active BLE transmission with connected AD8232 boards.

Metric	Prototype Version		
	#1 MVP	#2	#3 Release
<i>Design</i>			
Form Factor	Sleep Mask	Headband	Headband
<i>Cost</i>			
Total Cost (€)	€78.60	€68.30	€37.80
<i>Weight</i>			
Total Weight (g)	191	120.04	76.6
<i>Electronics</i>			
MCU Model	Adafruit Feather nRF52840 Sense	Adafruit Feather nRF52840 Sense	Seeed Xiao nRF52840 Sense
Battery Capacity (mAh)	220	220	200
Current Consumption (mA)	~6.14	~6.14	~8.23
Sampling Frequency (Hz)	200	200	250
<i>Electronic Case Dimensions</i>			
Length (mm)	60.4	86.4	66.0
Width (mm)	39.0	39.0	38.0
Height (mm)	16.2	13.2	13.6
Volume (cm ³)	38.16	44.50	34.11
<i>Capabilities</i>			
IMU	✓	✓	✓
Vertical EOG	✓	×	×
Lateral EOG	✓	✓	✓
Frontal EEG	×	?	✓
Pulse (PPG)	×	×	✓
Vibration Stimulation	×	×	✓
<i>Security Features</i>			
BLE Pairing	×	×	✓
Data Encryption	×	×	✓
<i>Material Requirements</i>			
Fabric	✓	×	×
Separate Cushioning	✓	✓	×

3.8.2 The Four Horsemen of Progress

The development of the OSSMM wearable can be framed by four key areas: ergonomics, manufacturability, cost, and sensing capability.

3.8.2.1 Form Factor and Ergonomics

The most significant architectural change was the pivot from the sleep mask design of Prototype #1 to the headband form factor of Prototypes #2 and #3. The MVP's mask design, while attempting to address comfort for prone sleepers, was ultimately impractical for lab-based fabrication. Its weight (191 g) and bulk created significant comfort issues. Prototype #2 adopted a more conventional headband form, reducing weight by 37% to 120 g, but its elongated electronic case and poor weight distribution created new ergonomic flaws. Prototype #3 focused on ameliorating the head band design and further reduced the weight to 76.6g (a 60% reduction from the MVP). This final version improved the chassis and electronic case profile to ensure comfort and stability.

3.8.2.2 Manufacturability and Assembly

An area with substantial improvement was in the manufacturability and assembly process of the prototypes. It is not possible to quantify this as no systematic measure was taken to account for the assembly times between the various prototypes. However, given the reduction in design complexity and material requirements, the evolution in this area demonstrates an important shift towards ease of assembly and reproducibility.

Prototype #1 was effectively a bespoke craft project, requiring specialized skills like sewing and tedious manual trimming of materials. This approach was fundamentally at odds with the open-source goal of easy and repeatable replication. Prototype #2 introduced key design improvements, specifically a modular architecture with swappable components and a snap-fastener system that eliminated loose wiring beneath fabric. Prototype #3 perfected this approach by eliminating the need for all upholstery materials like foam and fabric, instead using a purpose designed 3D-printed structure for cushioning. This reduced the component list to a minimum, making the final assembly process streamlined and repeatable. One of the most important factors is the reduction in need of “hands-on” knowledge.

3.8.2.3 Cost and Accessibility

A primary driver of this project was to greatly reduce the cost of an accurate sleep monitoring system to enhance accessibility to this technology for the research com-

munity. The final prototype's cost of €37.80 represents a 52% reduction from the MVP's €78.60, and under a third of the initial Design Constraint of €130.

This affordability was achieved through several key decisions: migrating to the economical Seeed Xiao MCU, reducing the number of AD8232 boards from two to one, and eliminating the cost of build materials like memory foam, fabric, Velcro, adhesive and shrinking the physical design. This price point makes the OSSMM system a promising tool for researchers who may need to deploy multiple units for parallel studies for long periods of time. Perhaps the most significant benefit achievement demonstrated was the elimination of consumables, like Ag/AgCl electrodes or conductive gels, which can cost several hundred euro even over a short experimental time span.

3.8.2.4 Sensing Capabilities and Electronics

During the design process, despite design simplification and cost reduction, technical capabilities expanded. The MVP featured dual channel EOG (but not EEG) and head motion tracking. Prototype #3 incorporates a PPG for pulse monitoring, a vibration motor for stimulation, and the EOG electrode positioning which permits the capture of frontal EEG signals.

A key enabler for this was an error in disconnecting AD8232's RLD pin from use. Proceeding without what was thought to be a necessary electrode proved that the system could acquire quality signals without it. This unexpected finding simplified the electronics, reduced material requirements, and critically, freed the central forehead location for the PPG sensor. This serves as a clear example of how challenging initial assumptions, even if mistakenly, during the prototyping process can lead to design breakthroughs.

3.8.3 Revisiting the Design Requirements and Goals

The success of an engineering process can be measured by comparing the release against the initial project objectives. Here a review considers the constraints and goals laid out in Section 3.2 with what was achieved in the release.

3.8.3.1 Assessment of Design Constraints

The final OSSMM prototype achieved the majority of the original design constraints, as shown in detail in Table 3.14 and 3.15. All physical design requirements were achieved or partially achieved, including the contact-based sensor approach, elimination of external wiring, and battery life exceeding 12 hours. The system substantially exceeded cost constraints. Material accessibility, assembly simplicity, and

safety requirements were fully satisfied, with all components commercially available and construction requiring only basic technical skills. The modular architecture demonstrated robust durability over one month of testing. However, several functional requirements remain incomplete: real-time REM detection and automated sleep modulation have not been attempted, though the hardware foundation is in place. Ease of use (Constraint 9) remains untested without formal user feedback. These unmet constraints represent software implementation challenges and required validation testing in a typical population, rather than discovered design limitations. This establishes a clear roadmap for future development.

Table 3.14: Achievement Status of Design Constraints (Part 1: Design and Performance). Status symbols: ✓ = Achieved, ○ = Partially Achieved, ✕ = Not Yet Achieved.

Constraint	Status	Description and Justification
<i>Physical Design</i>		
1	✓	Contact-based Design: The final system features a wearable headband using CTPU electrodes, a PPG sensor, and an IMU, all requiring physical contact.
2	○	Unobtrusive Wearable: The final weight of 76.6 g and the compact headband form factor are minimally obtrusive. Power is internal, battery life exceeds 12 hours, and there is no exposed wiring. Future iterations can improve on comfort and miniaturization.
<i>Functional Performance</i>		
3	✕	Real-time REM Detection: Proof-of-Concept testing indicates the OSSMM system provides useful data for sleep stage classification. However, this has only been tested with a silver standard device and offline analysis.
4	✕	Automated Sleep Modulation: The system incorporates a vibration motor that can be triggered remotely by the Base Station. While this serves as a sleep modulation example, it is not automated at this time.
11	○	Safe Data Recording: The system records all relevant sensor data, and the companion app includes functionality for data encryption. However, statistics on sleep measurements are not yet provided.

Table 3.15: Achievement Status of Design Constraints (Part 2: Implementation and Safety). Status symbols: ✓ = Achieved, ✓ = Exceeded, ? = Untested.

Constraint	Status	Description and Justification
<i>Cost and Accessibility</i>		
5	✓	Cost < €130: The final cost of €37.80 (approx. \$39 USD) is substantially lower than the initial constraint.
6 & 7	✓	Simple Construction: Assembly requires basic soldering and 3D printing skills. No advanced techniques or tools needed.
8	✓	Accessible Parts: All components are commercially available off-the-shelf or printable.
<i>Usability and Safety</i>		
9	?	Ease of Use: The modular design and simple operation via the companion app make the system easy to operate. However, without impartial user feedback, it is not possible to determine if this was achieved.
10	✓	Durability and Replaceable Parts: The OSSMM headband has been tested for over 2 weeks of consistent night time usage, with 2 different units (>1 month of total testing). The headband has never failed. The modular architecture also permits rapid and easy change of parts.
12 & 13	✓	No Overt Risks / No Consumables: The primary skin-contact materials are available on the commercial market, and 3D printer filament with extended dermal contact can be readily purchased with biocompatibility certifications. No consumables are required, and the CTPU electrodes work repeatedly even with daily use.

3.8.3.2 Assessment of Design Goals

The OSSMM release achieved mixed results against the project’s aspirational goals. Most notably, the system exceeded the ambitious cost target, achieving €37.80, nearly half of the €65 goal. However, more work remains to determine the capabilities of the hardware.

Table 3.16: Achievement Status of Design Goals. Status symbols: ✓ = Achieved, ○ = Partially Achieved, ✕ = Not Yet Achieved, ✓ = Exceeded, ? = Unknown/Untested.

Goal	Status	Description and Assessment
<i>Performance and Accuracy</i>		
1	?	Clinical Grade Data Quality: The system successfully captures quality multi-modal data, including PPG, EOG, and frontal EEG signals. Direct comparison to clinical-grade PSG to quantify its quality remains an area for future work.
2	?	Real-time Wake/NREM/REM Detection: Near-real time testing of the system has yet to be performed. This is an area for future work.
4	?	Accuracy >90%: Provided the results from the Proof-of-Concept experiment, this is an area for future work.
<i>System Features</i>		
3	✕	Additional Consensual Information: While the companion app allows for the future integration of cognitive tasks or questionnaires, this has not been implemented so no additional consensual methods have been employed.
<i>Cost Target</i>		
5	✓	Cost < €65: The final cost is approximately half of this aspirational goal.

3.8.4 Summary Design Evolution

In summary, the design journey of the OSSMM system was one of persistent simplification and optimization. The process began with an ambitious but impractical concept and concluded with a design that is simple, affordable, reliable. Each iteration served as a critical learning step, revealing that the most effective design solutions emerged from addressing practical manufacturing and usage constraints. The final prototype successfully balances these competing demands, delivering an

affordable and accessible open-source tool that not only meets but, in some areas, exceeds the initial design goals set forth for the project.

3.9 Software Ecosystem Evolution

Due to the split architecture, the OSSMM system comprises two software components: firmware for the Recorder and a companion application for the Android Base Station, developed in parallel with the hardware. All Recorder firmware was written in the Arduino IDE for its extensive support and accessibility, while the final companion app was built in Dart/Flutter. The application's evolution through the earlier frameworks that preceded Flutter is described in Section [3.4.1](#).

3.9.1 MCU Firmware

The firmware on the Recorder is optimized for low-power data acquisition. Its principal responsibilities are to sample the onboard accelerometer, gyroscope, and two analog pins (for PPG and EOG) at a default frequency of 250 Hz. To minimize payload size for wireless transmission, the firmware performs basic data compression by reducing sensor data from 4-byte floats to 2-byte unsigned integers. Data packets are then transmitted securely over BLE. The final version of the firmware also incorporates security protocols for BLE pairing and is capable of receiving and acting upon bidirectional commands from the Base Station, such as activating the integrated vibration motor or powering down the unit.

3.9.2 Companion Android App

The OSSMM companion application serves as the system's control center and data hub. The application architecture was completely reconstructed for Android 15 compatibility, a milestone that introduced critical security and data integrity features. The application manages the secure BLE connection, provides real-time plotting of incoming sensor data for immediate signal quality verification (i.e., actigraphy, pulse, and eye movement). It also saves the complete, timestamped data from each sleep session. To ensure data is not lost during application crashes or unexpected shutdowns, the software continuously writes to a temporary Comma-Separated Values (CSV) file. Upon the successful completion of a recording session, this raw data file is compressed and encrypted using a unique key accessible within the application's settings, after which the temporary file is deleted by default to protect user privacy.

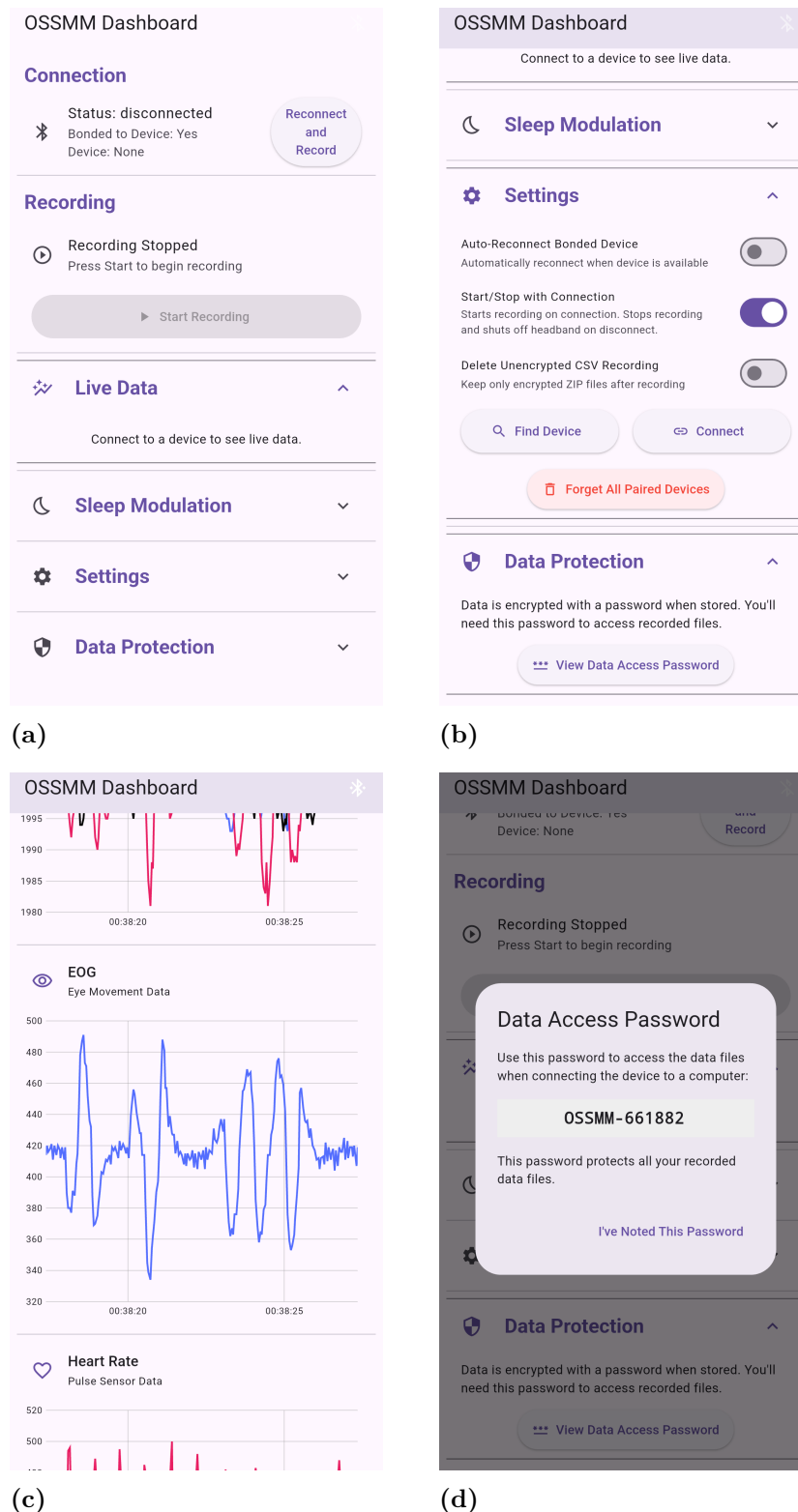


Figure 3.44: (a) A view of the OSSMM dashboard prior to connection with an OSSMM headband. (b) View of the settings section. In this case, the “Delete Unencrypted CSV Recording” is turned off, preserving the temporary CSV. (c) View of the EOG graph under “Live Data.” Users can clearly determine if the headband is functioning correctly with left-right saccades or rolling eye movement as is shown here. (d) Opened from the “Data Protection” area, this window shows the password needed to open the encrypted data files. This password is semi-randomly generated on the install of the app.

3.9.3 Future Software Work

While the current software ecosystem provides a platform for secure, long-term data collection, the primary area for future work lies in the implementation of on-device, near real-time sleep stage classification. It is envisioned that this can be accomplished using TensorFlow Lite within the Android app once an appropriate classification model and its corresponding weights have been developed. The existing architecture should support this functionality, which would enable the system to perform autonomous, closed-loop sleep modulation experiments, thereby fulfilling the project's scientific objective.

3.10 Conclusions

The iterative development of the OSSMM system indicates that steps towards democratizing sleep monitoring at consumer-device prices require not revolutionary technology, but careful engineering optimization and a willingness to abandon complex solutions in favor of simple ones that work. Each prototype iteration revealed that apparent limitations often masked simpler solutions waiting to be discovered through persistent investigation.

A key contribution of this project was finding strong supporting evidence that inexpensive (€2.50) commercial fitness electrodes capture EEG signals of sufficient quality for machine learning classification and, to some extent, for visual analysis of waveforms. This finding challenges common assumptions regarding the electrode quality required for effective EEG recording. This capability, combined with reliable PPG heart rate monitoring and precise head motion tracking, created a robust multi-modal dataset for investigating sleep stage classification algorithms. When combined with a final unit cost of €37.80 and a two-hour assembly time, this work makes multi-participant sleep studies financially and logistically feasible for a wider range of research groups.

The documented design evolution also provides valuable lessons for biomedical device development. Success emerged not from rigid adherence to initial concepts but from the willingness to fundamentally reimagine the system based on empirical findings. The transition from sleep mask to headband, from sewn fabric to 3D-printed cushioning, and from Ag/AgCl electrodes to CTPU, represent the abandonment of initial assumptions in pursuit of a simpler, more effective solution.

However, the system's effectiveness relative to gold-standard devices and across diverse populations remains to be formally established. The open-source nature of the OSSMM project ensures that this development can serve as a foundation for continued innovation. All hardware designs, software, and assembly instructions, are

openly available inviting researchers worldwide to build upon, adapt, and improve this work. The OSSMM system stands as proof that purpose-built sleep monitoring devices need not be expensive or complex to be effective, opening new avenues for large-scale sleep research and personalized health monitoring.

Technical Specification of the Open-Source Release

“An engineer can do for a dollar what any fool can do for two”

— *Attributed to Arthur M. Wellington, 1877*

Abstract

This chapter provides technical details and specifications of the release version of the Open-Source Sleep Monitor and Modulator (OSSMM) system, expanding on the engineering foundation for the subsequent technical chapters. Mechanical design, Android and Arduino software, power budgeting, wireless connectivity, and safety specifications are covered. A comprehensive open-source repository is provided as a key deliverable, containing all project assets, design files, and an in-depth assembly guide to enable replication and extension of the OSSMM system.

4.1 Introduction

The development of the OSSMM system, traced through its prototype iterations in the previous chapter, culminated in Prototype #3, the first release version. Before the platform’s capacity for automated sleep stage classification is evaluated, the exact technical reality of the hardware and software that produced the data must first be established. Establishing that foundation is the purpose of this chapter.

4.1.1 Chapter Scope and Objectives

This chapter documents the final release version of the OSSMM system, designated as V1.0.4, for the entire hardware and software project. The primary objective is to provide a technical description that is sufficiently detailed for another researcher to understand the system’s design choices, and replicate the findings of this thesis. The successful evaluation of the OSSMM system with machine learning methods, provided in Chapter 5 and Chapter 6, depends on the hardware and software implementation presented herein. The scope of this chapter therefore encompasses a complete breakdown of the system’s architecture, an examination of its hardware and firmware, an overview of the companion software, a review of safety and ethical considerations, and a presentation of the open-source documentation as a key research deliverable. The latter of which will permit other researchers to rebuild the system and record overnight datasets using only the materials and instructions provided.

4.1.2 Design Philosophy and Open-Source Principles

Therefore, this project committed to open-source principles, to ensure transparency, accessibility, and reproducibility. This approach provides researchers with complete access to the hardware designs and software code, enabling modifications and customizations for unique experimental needs. By making all hardware designs, software code, and documentation freely available, the OSSMM system will permit researchers to make design improvements and create standardizations for future sleep research.

4.2 System Architecture

The OSSMM system is built on a partitioned architecture designed to resolve the competing demands of continuous data acquisition, processing power, and comfort. The system is composed of two principal components: a Recorder, responsible for physiological signal capture, and a Base Station responsible for data storage, advanced processing, and user interaction. The Recorder takes the form of a wearable

headband, and the Base Station an Android device through a dedicated companion app. Leveraging the availability of Android devices, and separating tasks by hardware specialization, enabled the system to meet the design constraints for performance, accessibility, and affordability.

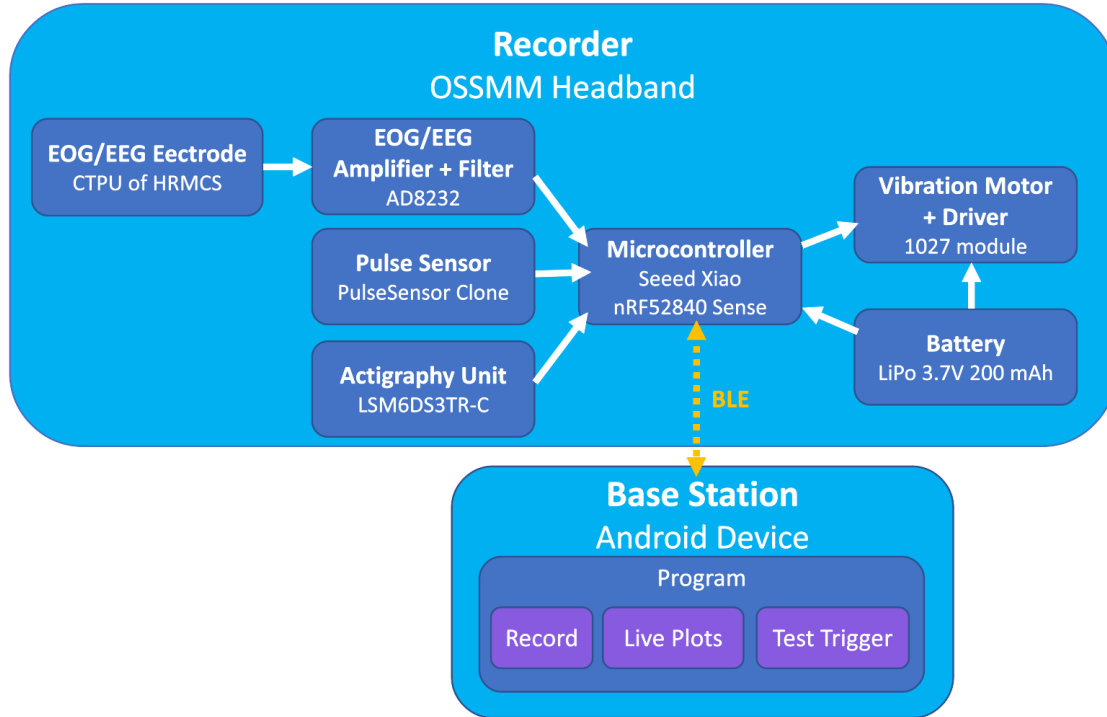


Figure 4.1: Block diagram of the OSSMM system architecture. Arrows denote the directionality of information or power pathways. Within the Base Station’s Program block: ‘Record’ refers to data collection, saving, and encryption from the MCU; ‘Live Plots’ to real-time visualizations of EOG, pulse, and actigraphy data; and ‘Test Trigger’ initiates a short pulsed buzz of the vibration motor.

The functional architecture of the OSSMM system is illustrated in the block diagram in Figure 4.1. This diagram presents a conceptual overview of the release version of the system, showing the major subsystems and the flow of data, power, or control signals between them, including the chosen hardware in the release version. The core of the wearable Recorder is the MCU, which interfaces with all other components. It acquires raw analog signals from the physiological sensors, either directly or through filtering and amplifier circuits. In total, four types of physiological data are captured: EOG and frontal EEG signatures, pulse, and actigraphy data (specifically head movement) from three different sensors. The MCU digitizes this data, packetizes it, and transmits it wirelessly via its Bluetooth transceiver to the Base Station. Power is supplied by a battery with integrated management circuitry. The system also allows for bidirectional communication, enabling the Base Station to send commands back

to the Recorder to trigger a stimulator, here a vibration motor, for potential sleep modulation.

4.3 Hardware Implementation: The Recorder

This section provides a detailed breakdown of the Recorder, detailing how the conceptual overview is accomplished in the final hardware of the OSSMM system. The components and design described herein correspond to V1.0.4. The mechanical construction, rationale behind the component selection, and summary of the final manufacturing cost are covered.

4.3.1 Mechanical Design and Materials

The mechanical design of the OSSMM Recorder prioritized user comfort, manufacturability, and performance. The platform consists of three major modules: (1) the electronic case, (2) a receiver, and (3) a commercial heart rate monitor chest strap (HRMCS). The HRMCS that serves as the headband strap and provides the EOG/EEG electrodes, while the receiver principally unites the electronic case and HRMCS. This modularity permits components to be readily swapped, for hygiene or repair.

Electronic Case. The electronics case and the receiver are both produced using fused deposition modeling (FDM) 3D printing. This allows for straightforward and virtually identical reproduction by other researchers. The electronics case is printed from Polylactic Acid (PLA), a standard material in hobbyist 3D printing, providing a rigid material suitable for protecting and isolating the internal electronics. In this thesis, units were printed with Prusament PLA. Its design dimensions are 66.0 mm x 38.0 x 13.6 mm.

Receiver. The receiver, which makes direct dermal contact, is printed from a flexible Thermoplastic Polyurethane (TPU). Both TPU-95A and TPU-85A are acceptable for the platform, but TPU-85A is recommended. For the units produced during the thesis, TPU-85A was selected for its increased softness. For safety reasons, we chose the manufacturer Siraya Tech as they produced multiple biocompatibility assessments for this specific filament, specifically analysis of cytotoxicity, irritation, and sensitization ([Jiangsu Kerbio, 2024a,b,c](#)). An innovation of note in the receiver's design is its independence of fabric or foam cushioning. Instead, the unit consists of integrated cushioning achieved by printing with a low infill density of 8% with a gyroid pattern. This internal geometry allows the structures to act

like an air cushion, distributing pressure while remaining easy to manufacture. Figure 4.2 demonstrates the receiver within PrusaSlicer, a program used to prepare CAD for 3D printing. The images highlight the gyroid infill within the design.

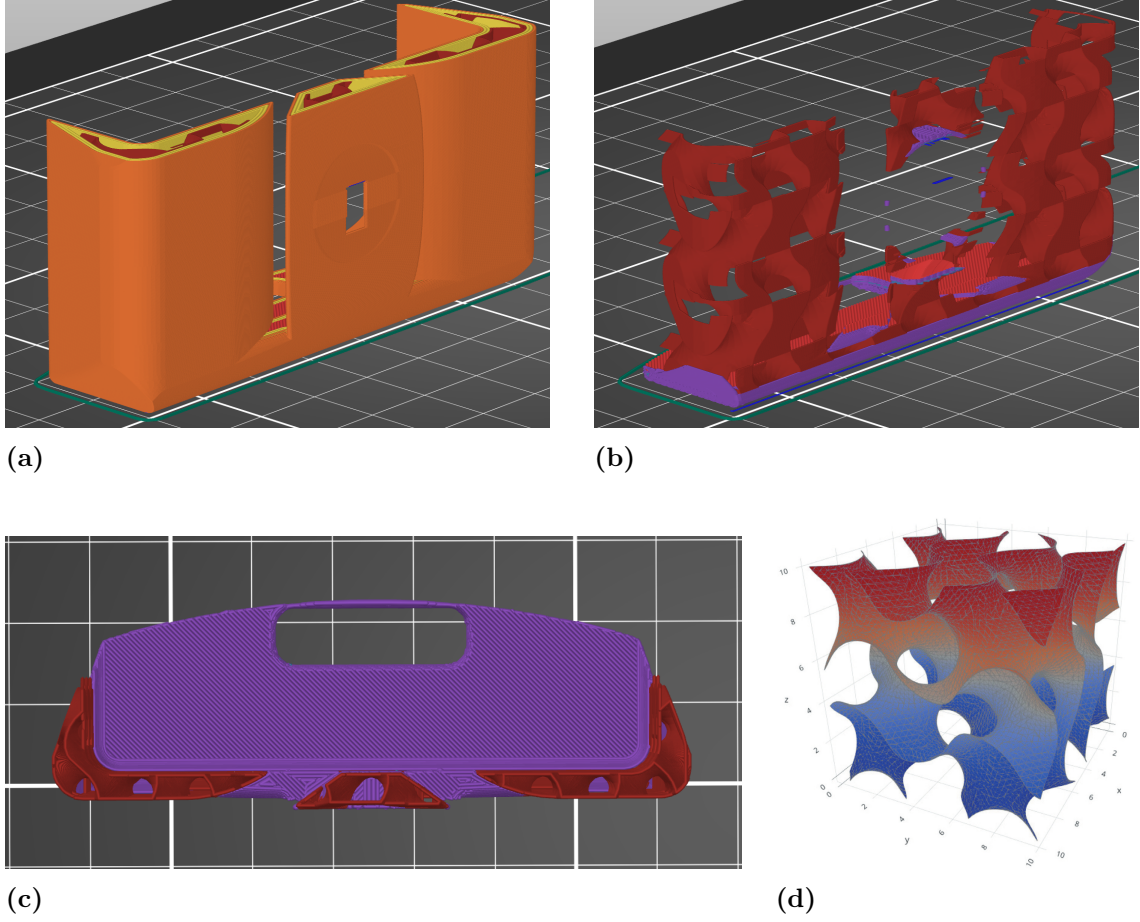


Figure 4.2: (a) the Receiver imaged at printline 374 (b) same image as *a*, but without the internal and external perimeters, highlight the gyroid infill shown in red. (c) Receiver at printline 350 from a top down view, highlight the depth of the cushioning. The slanted gaps in the back are slots for the HRMCS (d) Model of gyroid pattern taken from (Letov and Zhao, 2022). All screenshots are taken from PrusaSlicer 2.9.1.

Assembled Components and Measures When the electronic case and receiver are assembled, their nominal design dimensions are at $79.1 \times 47.2 \times 31.0$ mm. After several weeks of usage a tested unit measured at $78.6 \times 47.0 \times 32.8$ mm, as measured with a Kynup digital caliber with an accuracy of ± 0.2 mm. The minor deviation between physical and design dimensions is likely due to 3D printing tolerances. The increased depth of the unit is likely due to TPU material stretching from real-world usage over time, particularly as the folds of the HRMCS within the receiver maintain constant pressure on the electronic case, pushing it forward.

The final assembled device has a total mass of 76.6 grams with a 200 mAh battery.

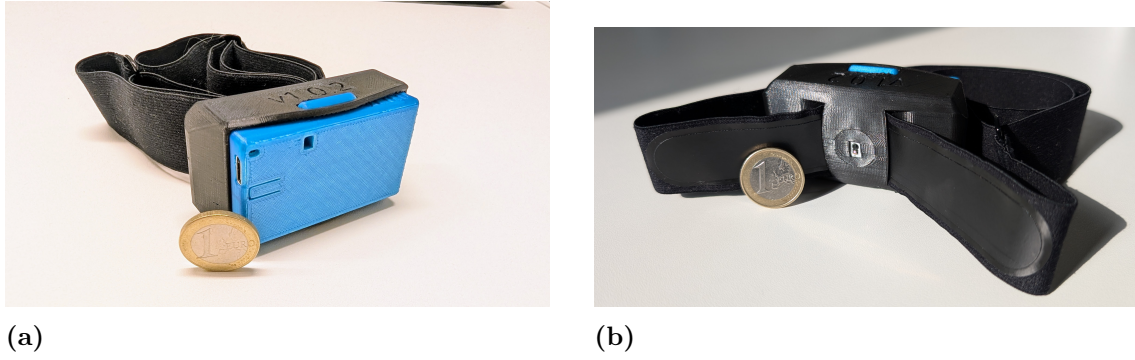


Figure 4.3: (a) Front view of the assembled headband. €1 coin for scale. Visible on the front of the electronic case are the USB-C port, the compliant mechanism serving as the ON/RESET switch, an LED notification window in the top-left corner, and the microphone window. (b) Rear view of the assembled headband, with €1 coin for scale. The CTPU electrodes are visible, and the window for the PPG sensor.

The weight will vary slightly depending on HRMCS brand, selected battery size, 3D printing variables (filament brand, filament print density, etc.).

4.3.2 Electronic Hardware

The selection of electronic components was guided by the project’s constraints, prioritizing low cost, low power consumption, small form factor, and widespread commercial availability.

Microcontroller Unit. The core of the Recorder is the Seeed Xiao nRF52840 Sense microcontroller (MCU). This board is built around the Nordic Semiconductor nRF52840 System-on-a-Chip (SoaC), which features a 64 MHz ARM Cortex-M4 processor, and integrated Bluetooth Low Energy (BLE) 5.0. This processor provides ample computational power for the data acquisition tasks while its architecture is optimized for low-power operation. The board’s small physical footprint (21×17.5 mm) was essential for the final compact design.

Physiological Signal Acquisition. The system was designed to capture three distinct physiological signals: EOG, PPG, and actigraphy data. Evaluation of the system strongly supports capture of EEG signature through the EOG channel. The analog signals from the EOG/EEG electrodes and the PPG sensor are digitized by the MCU’s analog-to-digital converter (ADC) with a 10-bit resolution, yielding integer values from 0 to 1023.

1. **EOG/EEG:** These signals are captured using reusable CTPU electrodes sourced from a commercial HRMCS. The signals are amplified and filtered

by an Analog Devices AD8232 integrated circuit, supplied on a COTS PCB. The AD8232's 0.5–40 Hz band-pass filter, designed for cardiac monitoring, is largely compatible with the frequency bands for both EOG and frontal EEG signals.

2. **PPG:** Pulse rate is monitored using a PPG sensor based on the PulseSensor design, an output from an open-source project [World Famous Electronics LLC \(2024\)](#). This non-invasive optical sensor uses a green LED and a photodiode to measure reflected light, which corresponds to volumetric changes in blood flow in the capillaries of the forehead. The sensor's onboard circuitry amplifies and filters this signal before it is passed to the MCU's ADC.
3. **Actigraphy:** Motion tracking is provided by the onboard LSM6DS3, a 6-axis IMU. For this application, the firmware configures the accelerometer to a range of ± 8 g and the gyroscope to a range of ± 2000 degrees per second (dps). These ranges are sufficient to capture subtle head movements during sleep without saturation during more significant postural shifts.

Stimulator — Vibration Motor. For haptic feedback, the system includes a simple vibration motor. Due to the low current output limit of the nRF52840's GPIO pins, the motor is driven by a small, dedicated driver board that amplifies the control signal from the MCU. The disc motor itself is a 1027 module (10mm in diameter, 2.7 mm in thickness).

Power and Safety. The Recorder is powered by a 3.7V 200 mAh Lithium Polymer (LiPo) battery. This capacity was selected to safely provide over 12 hours of continuous operation. The required capacity was determined based on the measured current consumption and the inclusion of safety and longevity factors, as detailed in Equation 3.1. Current measurements were estimated with the Nordic Power Profile Kit 2 (PPK2) during typical usage with sampling and BLE transmission ([Nordic Semiconductor, 2020](#)).

The system incorporates a two-tiered safety architecture for battery management. The LiPo battery itself includes an integrated SH-1738A protection circuit that prevents over-charge, over-discharge, and short-circuit conditions at the cell level, with a maximum short-circuit protection delay of 200 μ s. The Seeed Xiao MCU adds a second layer of safety with an onboard Texas Instruments bq2510x linear charger IC. This component manages the charging protocol and provides system-level overvoltage protection, short circuit protection to the battery, and thermal regulation, including a thermal shutdown function if a high temperature is detected.

4.4 Firmware and Software Ecosystem

The OSSMM system’s partitioned architecture is supported by a dual-component software ecosystem. This consists of dedicated firmware for the wearable Recorder and a companion application for the Android-based Base Station. The firmware, developed within the Arduino Integrated Development Environment (IDE) is designed for low-power data acquisition and secure wireless transmission. Likewise, the Android companion app, developed using the Dart and Flutter framework, provides a user interface (UI), data storage, and real-time visualization. This section provides a detailed technical description of the software responsible for the functionality of the Recorder and Base Station, and how they work in concert to form robust data collection platform.

4.4.1 Recorder Firmware (V1.0.4)

The firmware for the Recorder is the embedded software that controls all operations of the OSSMM headband. It is designed with two competing requirements: capture physiological signals and transmit them over BLE, while operating at a low-power to prolong battery life. This is achieved through a balancing of minor pre-processing and connection management logic. V1.0.4 of the Arduino code is provided in Appendix C in Section C.1.

4.4.1.1 Data Acquisition and Packetisation

The core function of the firmware is to acquire data from the onboard sensors at a stable rate of 250 Hz. Rather than using a hardware timer interrupt, the firmware implements a polling method using the microcontroller’s *micros()* function. Code using a timer interrupt was tested, and found to be marginally more energy efficient, consuming 0.08 mA less during BLE connection. However, it required direct register access making the code less accessible to future users. Over a 12 hour time span the energy savings equate to less than 1 mAh, therefore, code accessibility was prioritized.

A timing variable, *sampling_interval*, is calculated at startup with a guarding factor. This ensures the sampling rate consistently meets or exceeds the 250 Hz target, even with minor variations in processing overhead.

Within each sampling event, the firmware reads data from all active sensors: one from each axis of the accelerometer and gyroscope (LSM6DS3TR-C IMU), one analog channel for the EOG/EEG signal (AD8232), and one analog channel for the PPG signal (PulseSensor), for a total of 8 values. To optimize the system for low-power wireless transmission, the firmware performs on-device data compression by

converting and scaling raw sensor values, 4-byte floating-point numbers from the IMU, into 2-byte unsigned integers ($uint16_t$). This conversion halves the data footprint for these signals without a significant loss of resolution, which in turn reduces the wireless payload size. A transmission number is also added to each sample as a sanity check for lost samples or lost BLE packets. This is the only meta data created by the MCU.

To further improve the efficiency of wireless transmission, the firmware utilizes packetisation. Instead of transmitting data after each sample, the firmware buffers 10 complete sensor readings (i.e. 10 sample rounds) into a local character array. Once this buffer is full, the entire 180-byte data packet is sent in a BLE transmission. This strategy bundles the high-frequency samples into a single wireless event, resulting in an effective BLE notification frequency of 25 Hz. This approach achieves sufficient data throughput while capitalizing on the energy-saving potential of the BLE radio. The BLE packet layout is described in Table 4.1 and visualized in Figure 4.4.

Table 4.1: BLE packet layout.

Field (per sample)	Type / Bytes
Sample Number	uint16 (2)
Accel X, Y, Z (scaled)	uint16 $\times$ 3 (6)
Gyro X, Y, Z (scaled)	uint16 $\times$ 3 (6)
EOG/EEG (ADC-scaled)	uint16 (2)
PPG (ADC-scaled)	uint16 (2)
Total per sample round	18 bytes
Total per packet (10 sample rounds)	180 bytes
ATT header (opcode+handle)	3 bytes
Complete BLE Packet	183 bytes
Android ATT_MTU	184 bytes

BLE Sleep Data Packet Layout

Single Sample Structure (18 bytes):

Trans # 2 bytes	EOG 2 bytes	Pulse 2 bytes	Accel X 2 bytes	Accel Y 2 bytes	Accel Z 2 bytes	Gyro X 2 bytes	Gyro Y 2 bytes	Gyro Z 2 bytes
--------------------	----------------	------------------	--------------------	--------------------	--------------------	-------------------	-------------------	-------------------

Total: 9 measurements × 2 bytes = 18 bytes per sample

Byte-level Breakdown (Sample 0 detail):

Byte indices [0-17] for Sample 0:

[0]	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]	[12]	[13]	[14]	[15]	[16]	[17]
Trans #		EOG		Pulse		Acc X		Acc Y		Acc Z		Gyro X		Gyro Y		Gyro Z	

Complete BLE Packet (183 bytes total):

BLE Overhead 3 bytes	Sample 0 [0-17] 18 bytes	Sample 1 [18-35] 18 bytes	Sample 2 [36-53] 18 bytes	Sample 3 [54-71] 18 bytes	Sample 4 [72-89] 18 bytes	Sample 5 [90-107] 18 bytes	Sample 6 [108-125] 18 bytes	Sample 7 [126-143] 18 bytes	Sample 8 [144-161] 18 bytes	Sample 9 [162-179] 18 bytes
-------------------------	--------------------------------	---------------------------------	---------------------------------	---------------------------------	---------------------------------	----------------------------------	-----------------------------------	-----------------------------------	-----------------------------------	-----------------------------------

Color Legend:

Transmission Number	EOG	Pulse/HR
Accelerometer	Gyroscope	BLE Overhead

Figure 4.4: BLE Sleep Data Packet Layout. Each measurement uses Little-Endian format (LSB first, MSB second). The total payload consists of 10 sample rounds × 18 bytes = 180 bytes. The complete packet includes 3 bytes of BLE overhead + 180 bytes of payload = 183 bytes total.

4.4.1.2 Bluetooth Low Energy (BLE) Protocol

The firmware implements a custom Generic Attribute Profile (GATT) to structure all wireless communication between the Recorder and the Base Station. This profile defines a primary service, identified by a unique 128-bit UUID, which contains three characteristics for managing data flow and device control.

1. **SleepData Characteristic:** This characteristic is the primary data channel, configured for notifications only. The Base Station subscribes to it to receive the 180-byte packets of sensor data streamed from the Recorder.
2. **SleepModulator Characteristic:** This is a bidirectional characteristic with read and write properties. It allows the Base Station to send commands to the Recorder, such as a command to trigger the haptic feedback motor.
3. **PowerControl Characteristic:** This is also a bidirectional characteristic that enables the Base Station to query the device state and, critically, send a command to initiate a “shutdown” of the Recorder.

Given physiological data, potentially related to health, is recorded, privacy is a fundamental requirement for the BLE implementation. The firmware is configured to use the “Just Works” association model. When a connection is initiated by the Base Station, this configuration prompts an unauthenticated pairing process that establishes an encrypted link between the devices. By specification, the BLE link-layer encryption uses the AES-128 in CCM mode (AES-CCM) cipher. All characteristics are configured with a permission level of *SECMODE_ENC_NO_MITM* (encryption with no man-in-the-middle protection), which mandates that the BLE link must be encrypted before any data can be exchanged. This ensures that all communication, from sensor data streams to control commands, is secured against passive eavesdropping.

Importantly, this implementation of “Just Works” pairing supports bonding, where the encryption keys are stored on both devices after the initial pairing. This feature is critical for research deployment, as it permits a researcher to pair a specific headband to a specific smartphone, creating a trusted and persistent link between the two devices for the duration of a study. That is, in real-world experimental use, devices should be paired in a secure laboratory setting before being used for data collection.

4.4.1.3 Power Management and Connection Logic

Efficient power management permits the +12-hour run-time target while promoting a long battery cycle life. Because the nRF52840’s Cortex-M4 runs at a fixed 64 MHz, typically the main adjustable influence on current draw aside from the radio is the sampling rate. Its effect is measurable, but in this case counterintuitive. As shown in Figure 4.5, higher sampling rates were found to slightly reduce current consumption. This phenomenon was only measured empirically and not investigated further. It is possible in this case that the higher sampling frequencies may align better with the radio wake-ups, resulting in less wasted radio usage over time. Regardless, this behavior, together with the resolution requirements of the physiological signals, supports selecting 250 Hz as the default operating frequency.

Beyond the sampling rate, the underlying firmware leverages the inherent low-power capabilities of the BLE stack. The Nordic Semiconductor BLE stack automatically duty-cycles the radio and places the CPU into a low-power state between wireless transmission events (at 25 Hz). This automatic power-saving minimizes energy use during idle periods in the communication cycle.

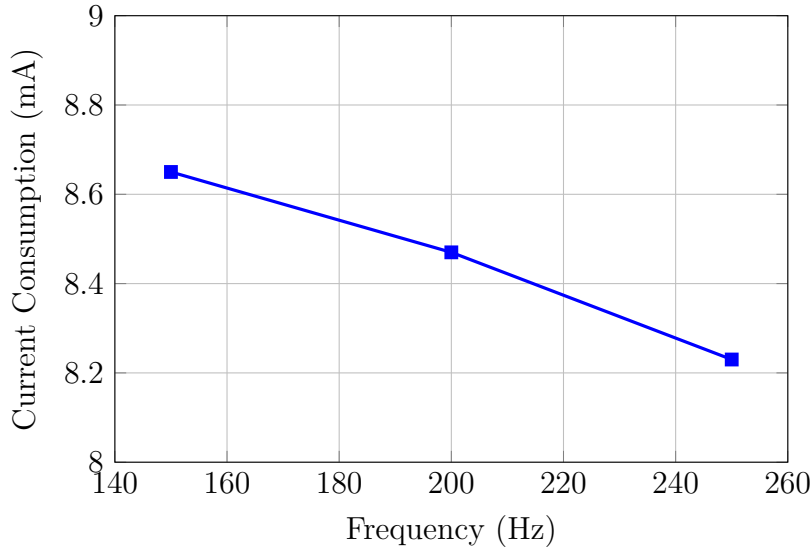


Figure 4.5: Current consumption as a function of the sampling frequency observed in the Seeed Xiao nRF52840 Sense while transmitting over BLE. Current values represent the mean value observed over 60 seconds of transmission using the PPK2. Sampling frequency denotes the set value within code, not measured values.

Figure 4.6 presents a screenshot of the Nordic Power Profiler v4.3.1 during power testing of the OSSMM system with the PPK2. The BLE radio wakes once per 15 ms connection interval, visible as the regularly spaced spikes in current draw. Four connection events appear in this view. During some connection events no data is queued, so only an empty packet is exchanged to maintain the link, producing a brief spike. When a notification is queued (i.e. data), the radio stays awake longer to transmit the 180-byte packet, producing a prolonged current draw. Here, the first event (at the start of the highlighted region) and the fourth (just after it) carry data, while the two intervening events are empty. The highlighted region therefore spans 45.07 ms, three connection intervals, from one data transmission to the next.

This separation reflects the interaction of two timings. The firmware queues a new packet for notification every 40 ms (the 25 Hz transmission cadence, set by buffering 10 sample rounds at ≈ 250 Hz), but a queued packet can only be sent at the next connection event, so transmissions are quantized to whole 15 ms intervals. This produces inter-transmission gaps of predominantly 45 ms (three intervals, shown here) and periodically 30 ms (two intervals), averaging the nominal 40 ms. This quantization is one contributor to the inter-packet timing variability discussed in Section 4.4.3.

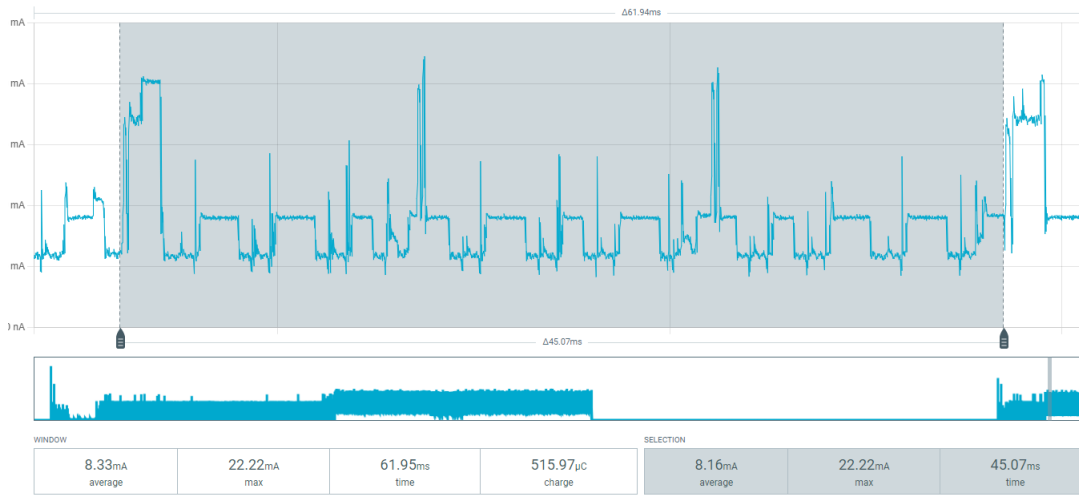


Figure 4.6: Screenshot of Nordic Power Profile v4.3.1 during power testing using the PPK2. Highlighted region spans 45.07ms

In addition to the BLE stack’s automatic power saving, the Arduino code implements several explicit power management strategies. A robust reconnection logic was designed to handle real-world usage scenarios, such as a user temporarily walking out of the Base Station’s range (i.e., out of the sleeping area). Upon an unexpected disconnection, the firmware places the device into a reconnection mode for a 10-minute timeout period. During this time, it advertises its presence to allow the Base Station to reconnect automatically. If no connection is re-established within 10 minutes, the firmware executes a deep-sleep command (`NRF_POWER->SYSTEMOFF = 1`) to power down the microcontroller. This same deep-sleep command is used for intentional shutdowns triggered by a command from the companion app, providing a method for the user to turn off the device.

Finally, the firmware includes intelligent startup logic to prevent accidental power-on and permits a charge-only mode. At boot, the code checks hardware registers to determine if the startup was initiated by a press of the reset button or merely by the connection of a USB power source. If the device detects that it is receiving USB power without an intentional reset-button press and a serial connection, it assumes it has been connected for charging and proceeds to re-enter deep sleep, thus avoiding unnecessary battery waste. This also provides users a method to shut down the headband should it be turned on and the companion app isn’t available for use.

4.4.2 Base Station Companion App (Android)

The OSSMM companion application is built for the Android operating system using the Dart and Flutter framework. The release version is compatible with Android 15 and above. Its main purpose is to serve as a UI, process the data, and store it.

The app’s workflow begins with scanning for and establishing a secure, bonded connection with the user’s designated OSSMM Recorder headband. Once connected, it subscribes to the SleepData characteristic and requests a Maximum Transmission Unit (MTU) size of 184, then begins receiving the 180-byte data packets transmitted by the Recorder. The app parses this incoming byte stream, reconstructing the 10 individual sample rounds within each packet into their respective time series: actigraphy (accelerometer and gyroscope), EOG/EEG, and PPG, and time stamping each sample round. These data streams are simultaneously rendered as real-time plots, allowing the user to confirm that the headband is functioning correctly before beginning a sleep session, for instance by observing saccades in the EOG plot via intentional left-right eye movements.

To ensure data integrity, the app continuously writes all incoming data to a temporary Comma-Separated Values (CSV) file on the smartphone’s local storage, safeguarding the recorded data against loss in the event of an application crash or unexpected device shutdown. Upon the completion of a recording, the temporary CSV is archived into a password-protected ZIP file using the ZipCrypto (ZIP 2.0) legacy encryption standard, after which the original unencrypted file is deleted by default to protect user privacy. ZipCrypto was selected as a temporary measure because it is widely supported and the resulting archives are easy to access on any computer with standard utilities. A unique password is generated when the app is first used, formatted as OSSMM-XXXXXX, where XXXXXX are the final six digits of the creation timestamp. This password is reused for all subsequent recordings, stored on the device using the Flutter Secure Storage library, and accessible to the user within the app’s settings. While suitable for proof-of-concept evaluation, future versions should migrate to AES-standard encryption and restrict the password from participant viewing.

4.4.3 Bluetooth Sampling Cadence and Timestamping

A critical aspect of the OSSMM system’s design is the timing and flow of data from the point of capture on the Recorder to its storage on the Base Station. The architecture must balance the need for high-frequency sampling against energy consumption with the practical limitations of BLE communication and smartphone operating systems. To summarize, the firmware on the Recorder samples all phys-

iological signals at an approximate 250 Hz rate. It then buffers 10 of these sample rounds before transmitting them as a single 180-byte BLE packet. Upon receiving this packet, the Android companion app applies a single, high-resolution DateTime stamp to each sample round before writing it to the local file.

In theory, this process could create two distinct timing domains within the final dataset. Within a packet, the 10 sample rounds are separated by approximately 4 ms ($1 / 250$ Hz) as dictated by the Recorder’s timing loop. Between packets, timing is variable. Packet delivery to the Android device is subject to the BLE connection schedule and Android’s non-deterministic scheduling, which can introduce jitter due to competing system processes. Consequently, the final data stream may not represent a perfectly uniform 250 Hz signal, but rather consists of precisely timed 40 ms “bursts” of data, with a variable time gap between each burst.

This timing characteristic is a deliberate design trade-off, prioritizing low power consumption and high data throughput over perfect temporal uniformity at the Receiver. However, the primary application of the OSSMM system is through automated sleep staging, which relies on data collected in 30-seconds epochs. Therefore, this micro level of jitter between packets is on a scale which does not carry significant impact on performance.

Figure 4.7 shows the number of samples for each complete 30-second epoch during an over night evaluation. The mean number of samples per epoch was 7576.6, averaging a sampling rate of 252.52 Hz. Most sample counts per epoch were between 7570–7580, a range of just 10. Only 9 epochs exceeded this range out of a total of 1047, with the maximum and minimum reaching 25 samples beyond the mean. This marginal discrepancy between epochs does not significantly impact feature selection, or even spectral analysis methods with convolutional neural networks (CNN), as spectrograms are easily treated with padding or removals. However, future analyses which would require precise timing across samples and packet boundaries must take the current programming into account.

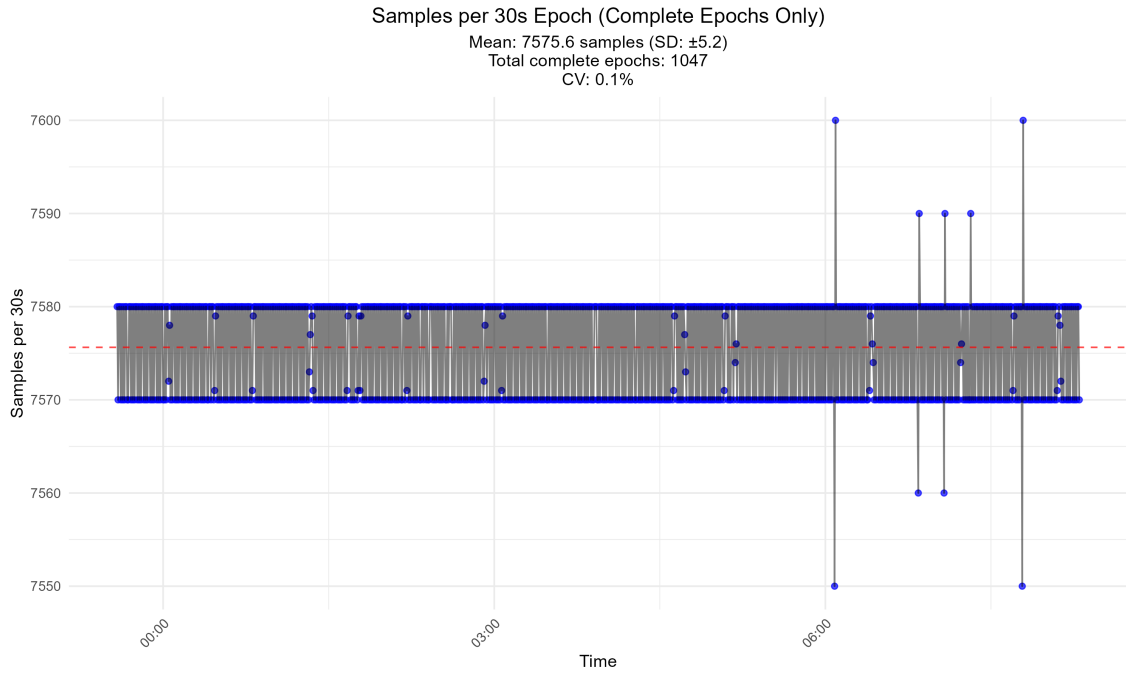


Figure 4.7: Samples count for each complete 30 second epoch throughout Night 2 of the experimental study. Most sample counts fell strictly within a range between 7570–7580, with just 9 observations outside of this window. SD = Standard Deviation, and CV = Coefficient of Variance. An epoch is considered complete if the time difference from the first and last observation is at least 29.5 seconds. Only 2 epochs in Night 2 were not complete, the first and last, which is expected during connection/disconnection.

4.5 Safety, Regulatory, and Ethical Considerations

Prioritizing user safety was fundamental to the OSSMM system’s design. This section details the key safety features concerning the system’s electrical hardware, material selection, and data handling protocols, presenting its suitability for research involving human participants.

4.5.1 Electrical and Battery Safety

The system’s electrical safety is ensured through its inherently low-voltage (3.7V battery, 3.3V MCU) operation and a multi-layered battery protection architecture that prevents over-charge, over-discharge, and thermal runaway. Key electronic components, including the battery and the Seeed Xiao Sense MCU, are CE marked, indicating compliance with health, safety, and environmental protection standards for products sold within the European Economic Area. Other modules, such as the AD8232 board and PPG sensor, are components intended for integration and therefore do not qualify for individual CE markings.

Furthermore, all measurement systems are passive, with the exception of the PPG sensor. However, the PPG uses the same low-intensity, light-based technology found in consumer smartwatches. As a deliberate safety measure, the design avoids any active injection of electrical current into the user (i.e., the removal of the RLD electrode).

4.5.2 Material Biocompatibility and Hygiene

The CTPU electrodes for EOG/EEG capture are sourced from commercial HRMCS. While these are widely used in consumer fitness products, it may be prudent for research purposes, that they are sourced from vendors who can supply material requirement documentation. This documentation should also include information about resistivity values which would be helpful for replication studies.

To ensure hygiene, particularly in multi-participant studies, the system's modularity facilitates effective cleaning. For daily use, the electronic case and receiver can be wiped with a damp cloth, as can the CTPU electrodes to prevent the build-up of skin oils or facial products. For disinfection between participants, the device should be disassembled to its modular components and thoroughly cleaned. The HRMCS headband can be washed and dried, while the 3D-printed components can be treated with appropriate disinfectants.

4.5.3 Data Security and Privacy

Data privacy is protected through a two-pronged strategy. First, all wireless communication is secured using industry-standard BLE link-layer encryption. Second, all data is stored exclusively on the Base Station. The release version of the companion app has no internet connectivity and does not use cloud storage. An additional layer of file encryption is applied to session data after recording is complete.

4.5.4 Risk Analysis and Mitigation

The primary risks of minor skin irritation and battery malfunction were systematically addressed. The risk of skin irritation from the TPU receiver is minimized by the use of a filament with manufacturer-provided biocompatibility data. For the CTPU electrodes, the risk is mitigated by using components common in consumer products designed for skin contact. In line with established experimental protocols for wearable devices, it is recommended that participants be given the opportunity to test the materials on their skin for a short period to ensure no adverse reactions occur before commencing an overnight study.

The risk of fire or overheating due to battery failure is estimated to be very low. This risk is mitigated by the two-tiered protection circuitry (on the battery and

on the MCU) that provides overvoltage, short-circuit, and thermal protection. The firmware also includes logic to prevent accidental power-on during charging, entering a deep-sleep mode if it detects USB power without an intentional startup command from the reset button.

These mitigation strategies, combined with the inherently safe low-voltage design and non-invasive sensing methods, demonstrate a rigorous and responsible engineering approach to user safety.

4.6 The Open-Source Deliverable: Dissemination and Reproducibility

A central tenet of the OSSMM project is that the scientific contribution is not limited to the physical device but equally encompasses its comprehensive, open-source documentation. This documentation serves as the primary vehicle for achieving the project's goals of transparency, accessibility, and reproducibility. By making all design files, source code, and assembly instructions freely available in a clear and structured format, the project transitions from a singular research artifact into a reusable platform for the wider scientific community. This section details the structure of this key research deliverable and the framework established for its dissemination.

4.6.1 Open-Source Repository

All project assets are hosted in a public GitHub repository at: <https://github.com/jvgiordano/OSSMM>, which serves as the central archive for hardware designs, firmware, the companion app, and documentation. The repository is organized to be intuitive, with three principal folders for 3D printable files (STL), microcontroller code (.ino), and the Android application source code (apk).

To ensure posterity, the hardware and software evaluated in this thesis has been version tagged as V1.0.4. This provides a stable baseline for comparison and further development.

Accompanying the repository is a dedicated documentation website generated using GitHub Pages (<https://jvgiordano.github.io/OSSMM/>), which presents the project information in an accessible, web-based format.

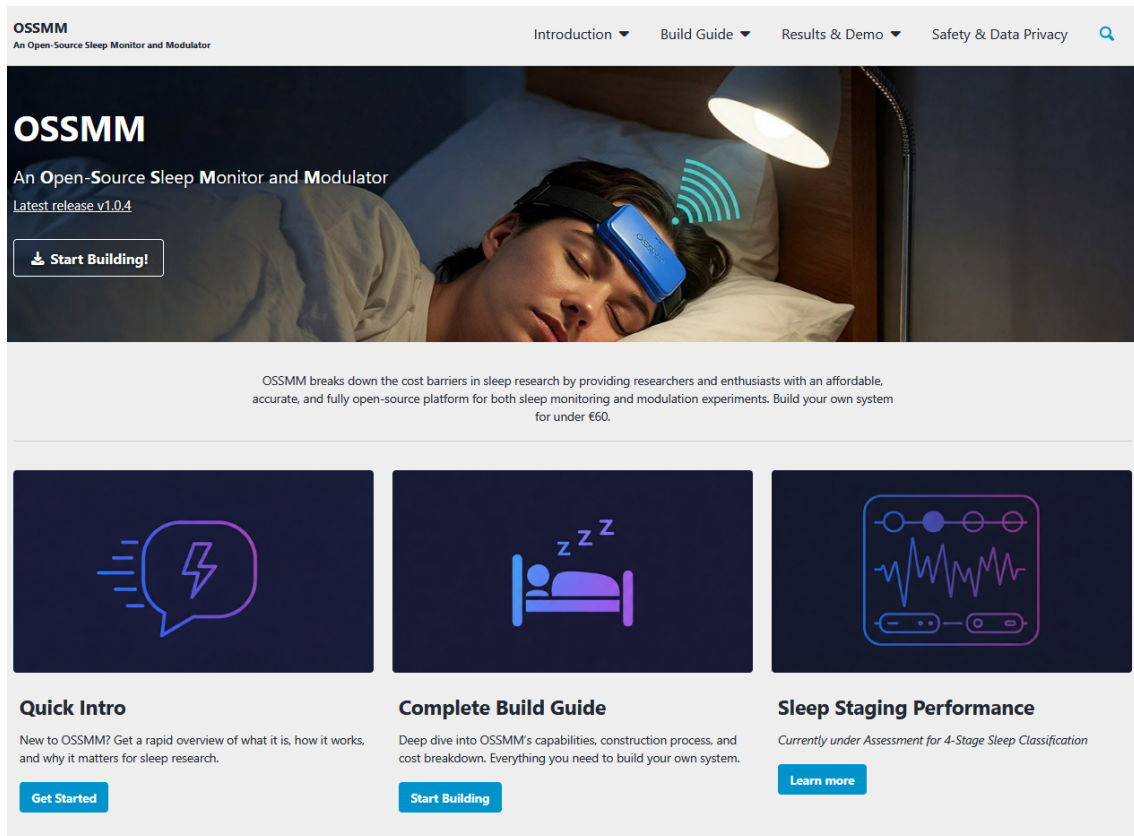


Figure 4.8: The landing page of the OSSMM website, which provides an overview of the project and navigation to the comprehensive build guides. The banner image is an AI rendition.

4.6.2 Comprehensive Build and Configuration Guide

To lower the barrier to entry for replication, a comprehensive build guide was created, detailing every step of the assembly and setup process. The documentation is intentionally written to be accessible for users who may not be experts in electronics or 3D printing. While aimed at beginners, it contains sufficient technical detail for advanced users to understand and modify the system. The construction process is partitioned into five logical stages, each with its own detailed instructions, photographs, and technical notes:

1. **3D Printables:** Instructions for printing the three required components, including recommended slicer settings and material choices.
2. **Electronics Assembly:** A step-by-step guide for soldering the electronic components and installing them into the 3D-printed case.
3. **Major Parts Assembly:** Instructions for integrating the three principal modules: the completed electronics case, the receiver, and the modified heart rate monitor strap.

4. **Software Installation:** Procedures for uploading the Arduino sketch (i.e., firmware) to the MCU and installing the companion app (apk) onto an Android device.
5. **Final Checks:** A protocol to verify all sensors and functionalities of the newly assembled device are operating correctly.

This structured approach ensures that the system can be reliably reproduced by other researchers, a prerequisite for independent validation and collaborative improvement.

The guide's extensive detail, which includes step-by-step instructions, photographs, and short videos, amounting to over 60 A4 pages in length, makes it best viewed in its digital format. Therefore, a link is provided to its online location rather than provided as a static appendix.

4.7 Discussion

This section provides a critical reflection on the release version of the OSSMM project: design, limitations, and overall contribution to the field.

4.7.1 Design Trade-offs and Justifications

The final design of the OSSMM system is the product of numerous engineering trade-offs, where ideal performance was balanced against the core project goals of low cost, accessibility, and ease of assembly. The decision to use a simple, low-cost analog front-end, the AD8232, instead of more complex and costly alternatives, exemplifies this philosophy. While other components might offer a higher signal-to-noise ratio or more advanced filtering capabilities, the AD8232 proved sufficient for acquiring EOG and even frontal EEG signals of a quality suitable for machine learning classification. This choice was fundamental to achieving the sub-€40 unit cost. Similarly, the selection of CTPU electrodes from commercial HRMCS over traditional Ag/AgCl electrodes was a critical decision. This choice directly satisfied the design constraint of avoiding consumables, which can represent a significant recurring cost in long-term studies. Although this introduced challenges related to manufacturing and ergonomics that were addressed over several prototypes, the final implementation successfully capitalizes on these reusable electrodes. The summation of these design selections ultimately led towards the final working release.

4.7.2 Limitations and Critique

The OSSMM system makes several trade-offs in order to achieve its initial design constraints, has limitations inherent in its design, and has several areas which require improvement.

Based on personal observation, the performance of the haptic feedback system is suboptimal. This assessment was consistent with informal user feedback. The vibration motor was selected as a robust, power-intensive stimulus to prove the system's capability, but its tactile feedback is dampened by the combination of the folded HRMCS strap and the soft TPU receiver. The result is a stimulus that is more audible than tactile, and further investigation is required to determine if it can effectively modulate sleep without causing arousal due to sound.

Continued testing of the headband has shown other areas of improvement. The UI of the electronic case could be ameliorated to assist in user access to the LED notifications. These light notifications are only readily apparent while looking at the front of the device. It would be beneficial to the user to know the state of the headband while wearing it. It is however not clear if adjustments to the electronic case can be made given the existing components. From a comfort perspective the metal buckle straps on the HRMCS can become irritating with long usage (i.e., during the night), which is a user comfort problem. This requires future work, but could be improved by providing some kind of soft cover or replacing the metal buckles with 3D printed soft buckles (i.e., TPU).

Concerning the Android app, the data encryption strategy, while convenient for researchers in development, constitutes a notable limitation from a security perspective. The implementation relies on the ZipCrypto standard, a legacy algorithm that is known to be vulnerable to modern cryptographic attacks. While the encryption is sufficient to protect data against casual attacks, if the smartphone were lost, it would be inadequate for research involving sensitive participant data. Future development should prioritize migrating to a stronger, widely supported standard such as AES-256 encryption within the archive format. For proof-of-concept testing with operating protocols where the data is immediately brought to a properly encrypted system after collection the risk is mitigated.

Finally, the scope of the current Base Station software is limited to the Android platform. The decision to focus on Android was a practical necessity. However, platform compatibility remains a possibility for any platform (e.g., iPhone, Raspberry Pi, etc.) which supports BLE and high connection priority.

4.7.3 Contribution to Open-Source Biomedical Hardware

The primary contribution of this work is the delivery of a complete and well-documented open-source system that lowers the barrier to entry for sleep research. By providing a platform that can be built for under €40, this project makes multi-participant, longitudinal sleep studies financially and logistically feasible for a wider range of researchers, particularly sleep research groups operating without large grants for whom conventional monitoring equipment may be prohibitively expensive. The successful use of inexpensive commercial fitness electrodes to capture not only EOG but likely EEG signals of sufficient quality for machine learning is a key finding that challenges common assumptions about the cost of hardware required for effective sleep monitoring.

The open-source documentation, including the detailed build guide is positioned as an important research deliverable of equal importance to the physical device itself. The OSSMM system is therefore not presented as a finished product, but as a foundational platform for future researchers and enthusiasts to build, adapt, and innovate upon.

4.7.4 Constraints and Innovation

The severe resource constraints of this thesis proved challenging but clarifying for the hardware development process. With only a €1,000 budget for all hardware development, the system design necessarily prioritized component accessibility and cost minimization. This constraint in part drove the investigation of CTPU electrodes from commercial fitness bands, a choice that initially stemmed from budget limitations and ultimately revealed an unexpected possibility. Rather than requiring research-grade Ag/AgCl electrodes or conductive gel, the system design demonstrated that sleep monitoring hardware could be built entirely from commercially available and reusable components. With the technical basis of the OSSMM system now established, the subsequent chapters evaluate its capacity for automated sleep stage classification.

Proof-of-Concept Evaluation: Classical Machine Learning

“Talk is cheap. Show me the code.”

— Linus Torvalds, *Linux Kernel Mailing List*, August 25, 2000

Abstract

This chapter presents a proof-of-concept investigation of the Open-Source Sleep Monitor and Modulator (OSSMM) system V1.0.4. The experiment assesses the OSSMM’s performance against a validated, silver-standard device, the Somnofy SM-100. The chapter addresses three primary technical questions: the system’s reliability over multiple nights, its overall data quality for sleep staging, and the specific contribution of features from its PPG, EOG, and IMU sensors to classification performance. This investigation addresses a principal question of this thesis: can the OSSMM design, with its selected hardware components and system programming, reliably collect signals relevant for sleep staging. The results are presented here, supporting the OSSMM’s technical feasibility for longitudinal sleep monitoring. Supplementary results are located within Appendix D.

5.1 Introduction

That sleep is not a homogeneous state, but characterized by periods of distinct neural patterns was first discovered by Loomis, Harvey, and Hobart in 1937 (Loomis et al., 1937). It would not be until 1953, the same year DNA was discovered, that Rapid Eye Movement sleep (REM) was identified (Aserinsky and Kleitman, 1953; Watson and Crick, 1953). Since then sleep research has exploded, and numerous associations have been uncovered between sleep states and facets of cognition and health (Palagini et al., 2013; Shepard et al., 2005; Sikka et al., 2019; Vogel et al., 1975). The accurate identification of sleep stages is therefore essential for sleep research.

Traditionally, sleep staging has relied on manual scoring of polysomnography (PSG) by qualified experts (AASM, 2023). While considered the gold standard, PSG has significant trade-offs. A full polysomnography (Type 1) is not ecological, requiring attendance at a dedicated facility, necessitating trained staff for set-up and monitoring, while demanding the participant tolerate uncomfortable, restrictive equipment throughout the sleep period. Importantly, the cost is nontrivial. In the United States, a single polysomnography for investigating sleep apnea cost on average \$1,840 for an in laboratory evaluation, and an at-home evaluation \$1,575 (Kim et al., 2015). The cost of portable units, such as for research, is also significant. Type III portable systems cost approximately \$3,500 to \$5,500 (e.g., Natus Medical Embletta MPR: \$3,508 USD; Nox Medical T3: $\approx$ \$5,500 USD), while Type II polysomnography systems with full EEG capability have documented costs ranging from approximately \$5,500 to \$9,100 (e.g., Advanced Brain Monitoring Sleep Profiler & PSG2: \$5,495 USD; SOMNOmedics SOMNOscreen Plus PSG+AASM: $\approx$ \$9,100 USD) (CPAP Europa, 2024a,b; Tikkanen, 2025a,b). Furthermore, analysis of the raw PSG data requires dedicated training, and is time intensive.

Many solutions have been put forth to address these drawbacks. Reduced equipment set-ups requiring fewer sensor modalities or electrode placements have been employed, lowering costs and movement restrictions of users (Arnal et al., 2020; Grözinger et al., 2002). Portable or at home devices have been engineered, both for the medical, research, and commercial market which provide a more ecological approach for data collection, allowing users to sleep at home (Arnal et al., 2020; Toften et al., 2020). Automatic sleep scoring using machine learning classification has been employed, eliminating costly manual analysis and reducing variability introduced by human error (Grözinger et al., 2002).

However, it still remains the case that the barrier to sleep research remains high. Low-cost, accessible sleep monitoring devices perform poorly compared to PSG

(Robbins et al., 2024; Schyvens et al., 2025). Dedicated sleep monitoring devices cost more, and it is not unusual for the cost of a single device to be greater than that of a single night of PSG. Without significant funding, sleep researchers are therefore confronted with choosing between low-cost methods with poorer accuracies in order to capture more participants over longer periods of time, or high cost methods with improved accuracies but for fewer participants and short durations. Concerns also remain for the ease-of-use for take-home equipment for participants, with many systems requiring special conductive gels or electrode pads.

As difficult as passive sleep monitoring may be, active sleep monitoring is even more so. Active sleep monitoring is required for sleep modulation research, that is, investigations in modifications of sleep such as affecting a specific sleep state. This research is more demanding because it requires constant present attention particularly where targeting a specific sleep state or neural signature is the aim. The target may require immediate detection as soon as it occurs.

Despite promising results from sleep modulation studies, such as potential therapies for depression, or improving sleep, it remains a domain showing limited movement in improvement (Cartwright et al., 2003; Ketz et al., 2018; Landsness et al., 2011; Vogel et al., 1975; Voss et al., 2014). Consider that nearly 25 years ago one sleep modulation study employed near real-time automatic sleep staging, yet more modern studies completed more than a decade later returned to manual manipulations of sleep, no different than what had been done during the 1970s (Grözinger et al., 2002; Landsness et al., 2011; Rosales-Lagarde et al., 2012; Vogel et al., 1975).

Therefore, a solution is needed which can provide suitable sleep data collection, accurate sleep stage classification, while also being affordable, and completely transparent in how the data is collected and analyzed. Furthermore, since sleep modulation studies focus on various stimulus modalities and sleep related variables, a platform is required whose software and hardware can be modified to purpose.

The OSSMM project (Open-Source Sleep Monitor and Modulator) was born out of the idea for a sleep monitoring and modulation platform that would respond to these issues by being affordable, accessible, modifiable, and transparent. The platform's hardware and software are open-source, and it is designed for assembly with basic electronics knowledge and equipment by researchers or enthusiasts. The system comprises a wearable electronic headband, and a dedicated Android application which permits data storage and serves as a user interface. The headband collects movement, pulse, electrooculography (EOG), and electroencephalography (EEG) data and transmits wirelessly to the connected smartphone. The application allows users to verify the connection and view real-time data collection.

This investigation conducted a proof-of-concept technical investigation of the OSSMM platform, using the Somnofy SM-100 by Vitalthings, a validated non-contact sleep monitor, as a silver-standard reference device (Toften et al., 2020). The investigation was designed to answer several key questions regarding the system’s viability as a research tool.

First, the platform’s operational reliability and robustness over multiple nights of continuous use was assessed. This included evaluation of hardware durability, battery performance, wearability, and the fidelity of wireless data transmission.

Second, an evaluation was performed to determine whether the quality of physiological data captured was sufficient for automated sleep staging into four categories: Wake, Light Sleep, Deep Sleep, and REM, as determined by the Somnofy reference device. Specifically, the investigation examined whether conventional machine learning classifiers (Support Vector Machine, Random Forest, and XGBoost) could detect sleep stage indicators using minimal hyperparameter tuning and feature sets derived from basic signal processing techniques.

Third, to determine the individual contribution of feature selections from the sensor modalities, classification performance was analyzed using features derived from the photoplethysmography (PPG), electrooculography (EOG), and inertial measurement unit (IMU) sensors. This multi-faceted approach represents an initial step toward establishing OSSMM as a viable sleep monitoring solution for research applications, while evaluating design optimization and resource allocation decisions in its release version.

Answers to these three questions permit us to answer the principal research question: can the OSSMM design, with its selected hardware components and system programming, reliably collect signals relevant for sleep staging.

5.2 Methods

5.2.1 Experimental Design

A total of 17 nights of sleep data were captured using the OSSMM and Somnofy devices. The nightly protocol was as follows: The CTPU (conductive Thermoplastic Polyurethane) electrodes were wiped clean with a damp cloth and dried to remove any oils that may have accumulated from the previous night. The device was donned, powered on, and connected to the companion application. The application’s real-time plots were used to verify that movement, pulse, and EOG data were being detected and recorded. A typical bedtime activity (e.g. reading) was then carried out for at least 15 minutes to augment the number of “Wake” observations.

Immediately before attempting sleep, a modified consensus sleep diary was partially completed before sleep ([Carney et al., 2012](#)).

A Google Pixel 9 smartphone running Android 16 ran V1.0.4 of the OSSMM application. Shortly after waking the following morning, data recording was halted, and the headband powered off and removed. The remaining sections of the consensus sleep diary were completed for the respective night’s sleep, and the headband was connected to a USB-C charger.

The sleep diary was augmented with two self-report questions: “Did you notice the headband move during your sleep?” and “Did the headband come off during sleep?” These questions were included to assess device stability during overnight wear. If the device shifted position or came off completely during the night, its position was restored without verification of proper signal capture.

This proof-of-concept study utilized a single-participant design (N=1), with the researcher serving as participant across all 17 nights of data collection. This approach was selected for initial evaluation to assess basic system functionality, hardware reliability, and data quality before broader testing.

5.2.2 Ethics, Health and Safety

5.2.2.1 Ethics

This study received Tier 3 ethics approval from the Biomedical & Life Sciences Research Ethics Subcommittee at Maynooth University, and the approval can be found in [Appendix F](#).

This proof-of-concept investigation employed a single-participant design with the researcher as participant, consistent with established principles of responsible engineering practice ([IEEE, 2020](#)). The researcher-participant’s comprehensive understanding of the system’s construction, operation, and potential failure modes would enable immediate identification and response to technical issues during overnight testing. This approach follows the engineering principle that designers should test their own prototypes before broader deployment, ensuring preliminary safety validation occurs under conditions of maximal technical expertise and informed consent.

For this technical evaluation investigating device capability and data quality, the study aimed to establish system feasibility (whether the hardware reliably collects physiological data suitable for machine learning classification) rather than to quantify population-level performance metrics.

5.2.2.2 Health and Safety

The greatest priority in OSSMM design and construction was safety. The chosen filaments met dermatological safety standards (Jiangsu Kerbio, 2024a,b,c). Importantly, the OSSMM headband has no exposed wiring and delivers no current to the user. This investigation followed a set of Standard Operating Procedures (SOPs) which were reviewed by the Health & Safety officer of the Hamilton Institute at Maynooth University as part of ethical approval. As part of this approval process, the device design was reviewed by an independent expert.

5.2.3 Experimental Equipment Setup

The SM-100 device was placed on a bedside table with its distance setting configured appropriately through the SM-100 companion application. Once powered on and configured, the SM-100 continuously collected data without requiring daily initiation for recording. Data collected by the SM-100 were retrieved from the SM-100 server via an Application Programming Interface (API) provided by VitalThings. Each classified SM-100 observation captured approximately 30 seconds of data, with epochs labeled as “Wake,” “Light Sleep,” “Deep Sleep,” “REM,” or “Not Detected.”

5.2.4 Data Methods

Data from the OSSMM device were time locked to the 30-second bins of Somnify data which provided classification labels. Features were derived from the OSSMM data from these 30-second epochs for machine learning classification.

Of the 17 nights, 2 were discarded due to mains power interruptions (Nights 2 and 10), which prevented the Somnify device from achieving full nightly recording. Of the remaining 15 nights, three (Nights 3, 13, and 17) were randomly withheld for final testing using a random number generator. Test nights were not allowed to be consecutive in order to promote longitudinal representation of the data.

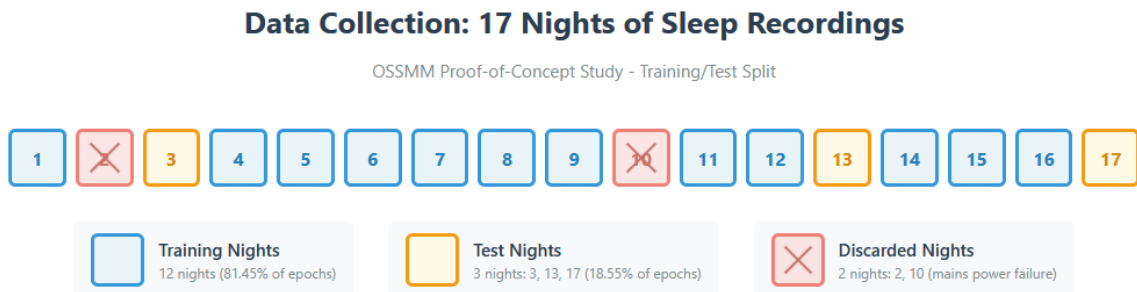


Figure 5.1: Breakdown of 17 experimental nights.

5.2.5 OSSMM System Hardware and Software

The OSSMM system was described thoroughly in earlier chapters 3 and 4. Explicitly, the OSSMM System Hardware and Software V1.0.4 was utilized in this technical evaluation. Arduino code utilized the Seeed nRF52 V1.1.8 board manager. The OSSMM V1.0.4 Android application was written using Dart Version 3.7.0 with Flutter Version 3.29.0.

5.2.6 Signal Processing and Feature Selection

All sensor data were digitized at 250 Hz with 10-bit ADC resolution. IMU data underwent baseline shifting to ensure positive values for optimized BLE transmission. No further signal processing occurred on the MCU. The classification models utilized 42 features.

A full descriptive list of features can be found in Appendix D under Section D.1.

5.2.6.1 Pulse Features

Pulse data contributed 8 features, principally capturing heart rate and heart rate variability through frequency domain and harmonic-based detection approaches. Signal preprocessing consisted of a 4th-order Butterworth low-pass filter with a 10 Hz cutoff frequency. For frequency-domain analysis, pulse data were additionally preprocessed by mean-subtraction detrending followed by application of a Hamming window to address spectral leakage.

5.2.6.2 IMU Features

IMU data provided 12 features exclusively characterizing movement rather than positional characteristics. Positional characteristics were avoided to improve generalizability as positional correlations with sleep states might be specific to the limited participant sample. Movement quantification consisted of three-dimensional magnitude calculations, basic statistical parameters, and adaptive statistical thresholding to identify sudden movement changes.

5.2.6.3 EOG and EEG Features

Finally, EOG/EEG data contributed 22 features. Fluctuations in impedance due to compression or movement of electrodes, and therefore unwanted signal changes, are common in wearable bio-signal recording, particularly EEG. To differentiate ocular movement from such motion artifacts, several adaptive thresholding features were employed. Basic statistical parameters such as standard deviation (STD) and Interquartile Range (IQR) were also calculated.

Although the headband was not originally designed for EEG, five power features in neurophysiologically relevant ranges were calculated: delta (<4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–100 Hz). Spectral density estimation utilized periodograms with a Tukey window (10% cosine taper).

5.2.7 Classifiers

Three out-of-the-box machine learning classifiers were examined in a comparative analysis: Support Vector Machine (SVM), Random Forest (RF), and XGBoost (XGB) (Chen and Guestrin, 2016; Pedregosa et al., 2011). For this experiment, no extensive search of optimal hyperparameters or feature selection was carried out.

Basic hyperparameter optimization was carried out on training data using Group K-fold cross-validation, grouped by night. Five hyperparameter configurations for each classifier type were tested, providing sufficient parameter space exploration to demonstrate system capability without extensive optimization. This conservative approach was chosen to establish that the OSSMM captures sleep-relevant signals detectable through basic tuning, thereby validating signal quality and system design rather than classifier optimization.

The five hyperparameter configurations for each classifier type were tested using a six-group K-Fold on the 12 nights of training data, such that on any given fold 10 nights served as training and 2 as validation. For Random Forest, configurations varied ensemble size (50–500 trees), tree depth (6–15 levels, plus unlimited), and splitting criteria (*min_samples_split*: 3–10, *min_samples_leaf*: 2–5), exploring the trade-off between model complexity and generalization. XGBoost configurations similarly varied tree depth (4–10), learning rate (0.05–0.1), ensemble size (50–500 estimators), and regularization through subsampling parameters (*subsample*: 0.7–0.9, *colsample_bytree*: 0.7–0.9). Support Vector Machine configurations explored regularization strength (C: 0.01–10.0) and kernel influence (gamma: 0.01, plus scale default) while maintaining RBF kernels and balanced class weights throughout. A complete view of these hyperparameter settings can be found in Appendix D under Section D.2.

One configuration for each classifier type was selected for testing on the withheld data. The selection criterion was the highest F1-score, with accuracy used as a tiebreaker for configurations achieving the same F1-score. Three qualifying classifier configurations were trained on the 12 training nights and tested on the three withheld nights. Class balancing was handled using Synthetic Minority Over-Sampling Technique (SMOTE) during training (Chawla et al., 2002).

5.2.8 Performance Metrics

All results are computed at the 30-second epoch level for $K = 4$ classes (Deep, Light, REM, Wake). For each class i , let TP_i denote true positives, FP_i false positives, FN_i false negatives, and N the total number of epochs.

Model performance was evaluated using macro-averaged metrics to ensure equal weighting across sleep stages regardless of class imbalance. This approach is important given the variation in sleep stage observations in the test set (Light Sleep: 45.9%, REM: 21.7%, Deep Sleep: 16.6%, Wake: 15.8%).

Cohen’s Kappa (κ) quantifies agreement between classifier predictions and reference labels while correcting for chance agreement:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where $p_o = \frac{1}{N} \sum_{i=1}^K TP_i$ is the observed agreement (overall accuracy) and $p_e = \frac{1}{N^2} \sum_{i=1}^K (TP_i + FN_i)(TP_i + FP_i)$ is the expected agreement by chance. Interpretation of kappa values for strength of agreement are as follows: 0.00–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1.00 almost perfect ([Landis and Koch, 1977](#)).

Balanced Accuracy (equivalent to Macro Recall) computes the unweighted mean of per-class recall values:

$$\text{Balanced Accuracy} = \frac{1}{K} \sum_{i=1}^K \frac{TP_i}{TP_i + FN_i}$$

This metric prevents majority class dominance in performance assessment, as Light Sleep constitutes nearly half the observations.

Macro F1-Score was selected as the primary criterion for hyperparameter search and final classifier selection. It is calculated as the arithmetic mean of per-class F1 scores:

$$F1_{macro} = \frac{1}{K} \sum_{i=1}^K F1_i \quad (5.1)$$

where $F1_i$ is the F1-score for class i , and K is the number of classes.

The F1 score for class i is the harmonic mean of Precision (Positive Predictive Value) and Recall (True Positive Rate, also known as Sensitivity):

$$Precision_i = \frac{TP_i}{TP_i + FP_i} \quad (5.2)$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (5.3)$$

$$F1_i = 2 \cdot \frac{Precision_i \cdot Recall_i}{Precision_i + Recall_i} = \frac{2 \cdot TP_i}{2 \cdot TP_i + FP_i + FN_i} \quad (5.4)$$

The F1-Score balances Recall (the ability to detect a sleep stage) with Precision (the correctness when that sleep stage is predicted), effectively balancing false negatives and false positives. The Macro F1-Score treats each class as equally important, which is essential for sleep staging applications where accurate detection of all sleep stages, including minority classes like Deep Sleep and Wake, is clinically relevant. Unless otherwise indicated, all accuracy and F1 metrics presented are balanced/macro-averaged.

5.3 Results

5.3.1 Durability, Wearability, and Power

During the 17 nights of usage, including 2 nights of interrupted mains power, the OSSMM headband and Android device maintained 100% uptime. On one night, the headband was not charged during the day according to the experimental protocol, but still functioned throughout the following night. The device therefore functioned for over 16 hours.

No parts on the headband broke or had to be replaced during testing. The headband never came off or fell below the face during all nights of testing. Displacement of the device was detected on 6 out of 17 nights. Three times the device shifted upwards, and three times a rotation was observed. None of these displacements were significant enough for the receiver and electronic case to leave the forehead area.

5.3.2 Connection Stability

To assess connection fidelity, the first and last two epochs of each recording session were excluded to avoid periods of BLE connection initialization and shutdown. Across 17 nights, including 2 nights of interrupted mains power, no packet loss was detected in BLE transmission among the remaining epochs. Of 16,907 analyzed epochs, 16,906 (99.99%) met the strict excellent quality criterion ($\pm 1\%$ of measured sampling rate). Only one epoch exhibited over-sampling (108.5% of expected samples), likely due to brief buffering, but maintained complete temporal coverage.

5.3.3 Data

From 15 nights of sleep data, 15,285 30-second epochs qualified for analysis. 40 epochs were excluded due to insufficient data samples (<95% of expected samples), all occurring during BLE connection initialization or shutdown. These boundary epochs reflect the stabilization period of the wireless connection described in Section 5.3.2. Additionally, 10 epochs were removed due to a Somnify device classification of ‘Not Detected’.

The dataset was split into 12 nights for training (12,449 epochs, 81.45% of total) and 3 nights for testing (2,836 epochs, 18.55% of total). Table 5.1 shows the sleep stage distribution across both sets.

Table 5.1: Sleep stage distribution in training and test sets

Sleep Stage	Training Set (%)	Test Set (%)
Deep Sleep	14.93	16.64
Light Sleep	48.28	45.91
REM	23.58	21.65
Wake	13.21	15.80
Total Epochs	12,449	2,836

Figure 5.2 shows a 30-second epoch labeled as REM sleep during Night 16. This observation was chosen because of its clear correspondence to its respective sleep label, and because it was noted to occur minutes before waking from a dream. The epoch presents hallmark characteristics of REM sleep: rapid eye movements, increased heart rate variability, and muscle atonia. Most of the segment is dominated by a series of saccades shown in the EOG plot. The PPG plot shows a change in heart rate variability, with the first half indicating a mean heart rate of 64 BPM, and the second half 72 BPM. This increased heart rate variability is indicative of REM sleep. No head movement is indicated by the accelerometer axes. The sleep diary notes waking from a dream, with wake onset occurring approximately 10 minutes after this observation.

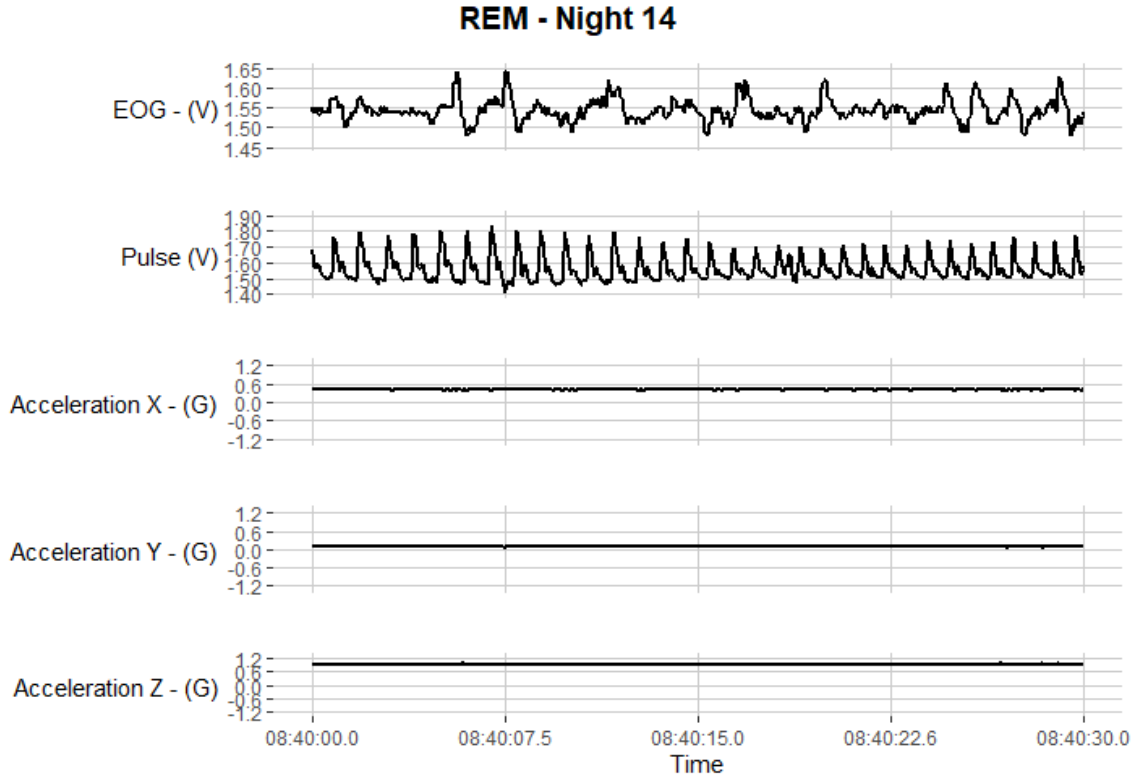


Figure 5.2: REM-labeled observation during Night 16, capturing portions of epochs 1002 and 1003.

5.3.4 Classifier Configuration Selection — Group K-Fold

Five hyperparameter configurations were evaluated for each classifier type using six-fold group cross-validation on the training data. Table 5.2 presents the selected configurations and their performance metrics. Random Forest achieved the highest performance (F1-score: 0.766), followed closely by XGBoost (F1-score: 0.762), while the RBF kernel SVM substantially underperformed (F1-score: 0.551).

Table 5.2: Selected Classifier Configurations and Cross-Validation Performance

Classifier	Hyperparameters	F1	Acc.
SVM	C=1.0, gamma=scale, kernel=rbf	0.551	0.548
Random Forest	n_estimators=150, max_depth=15, min_samples_split=3, min_samples_leaf=2	0.766	0.779
XGBoost	max_depth=6, learning_rate=0.1, n_estimators=100, subsample=0.8, colsample_bytree=0.8	0.762	0.776

Note: Configurations selected from 5 candidates per classifier based on highest macro F1-score during 6-fold group cross-validation on training data. Complete configuration details provided in Appendix D under Section D.2.

5.3.5 Classifier Performance on Test Data

The final evaluation on the withheld test data, shown in Figure 5.3, revealed a clear performance hierarchy. The SVM model was the least effective, yielding an F1-score of 0.598 and accuracy of 0.590. In contrast, both Random Forest and XGBoost demonstrated strong and marginally different performances.

The XGBoost model achieved an F1-Score of 0.741 and accuracy of 0.762. Random Forest obtained the highest performance on both metrics, with an F1-Score of 0.770 and an accuracy of 0.776.

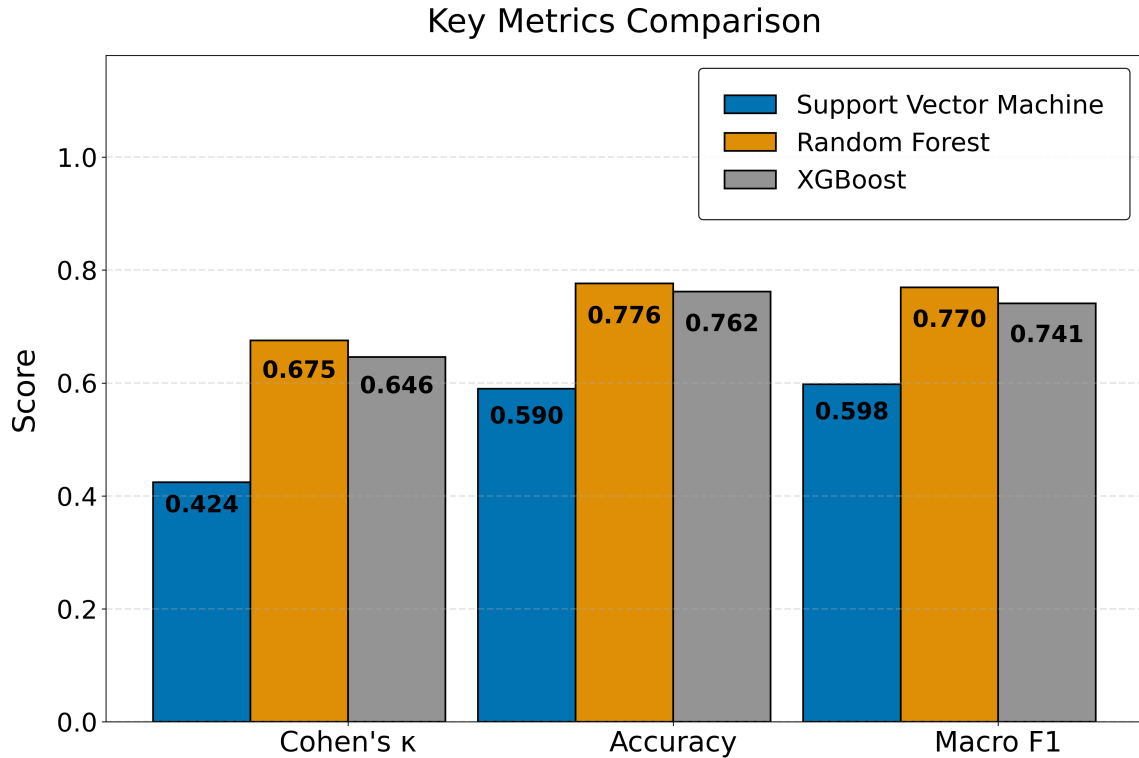


Figure 5.3: Classification Performance Comparison between Support Vector Machine, Random Forest, and XGBoost on Withheld Test Data. Random Forest and XGBoost demonstrated nearly equivalent performances, while SVM with RBF kernel performed noticeably worse.

Figure 5.4 shows the normalized confusion matrix for the Random Forest classifier. XGBoost and Random Forest demonstrated similar performance patterns, except for Deep Sleep where XGBoost struggled significantly. For the remaining sleep-wake states, both models achieved strong performance across all sleep stages. Both decision tree methods exhibited a clear distance gradient in misclassifications, whereby adjacent sleep stages were more frequently confused than distant stages. For example, both models never misclassified REM or Wake as Deep Sleep, and similarly, neither misclassified Deep Sleep as REM. Deep Sleep was misclassified as Wake by both algorithms just 1.1–1.3% of the time.

All classifiers demonstrated difficulty with Deep Sleep. Specificity for Deep Sleep was high, but recall was low. For XGBoost and Random Forest Deep Sleep was often misclassified as Light Sleep. Furthermore, the SVM classifier struggled particularly with REM versus Light Sleep differentiation, often misclassifying one as the other. SVM also over-predicted Wake, misclassifying other sleep stages as Wake more than 11% of the time, including for Deep Sleep.

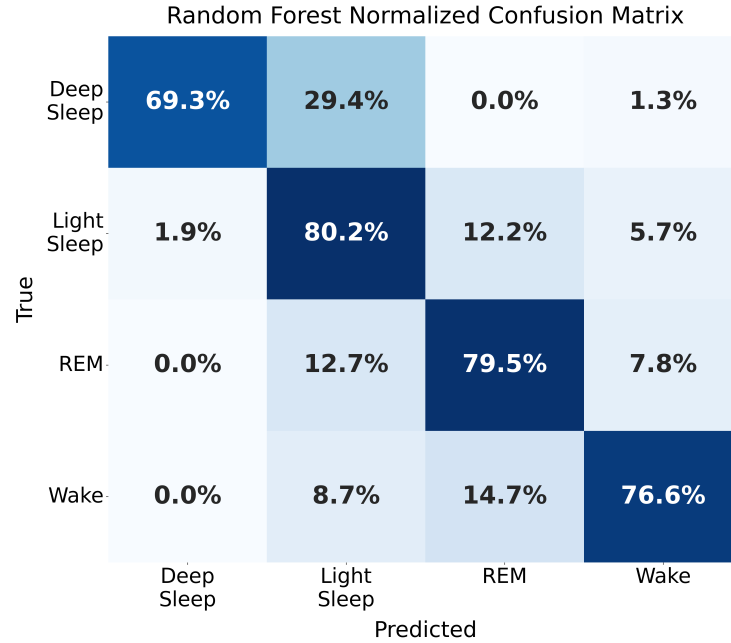


Figure 5.4: Normalized Confusion Matrix for Random Forest Classifier on Test Data.

Both Random Forest and XGBoost provide feature importance scores, and these were calculated with default settings. Explicitly, Random Forest used Mean Decrease in Impurity (MDI) and XGBoost used Gain Importance. The top 10 features for Random Forest are shown in Figure 5.5.

The two models demonstrated strong agreement in feature importance rankings. EOG-derived features dominated both models' top 10 lists, with all five EEG frequency band power features (alpha, beta, theta, gamma, and delta) appearing in both rankings. Both models identified alpha power as the most important feature, with Random Forest assigning it an importance of 0.153 and XGBoost 0.148. The order of frequency band importance differed between models: Random Forest ranked them (1) alpha, (2) theta, (3) beta, (5) delta, (6) gamma, while XGBoost ordered them (1) alpha, (3) beta, (4) theta, (5) gamma, (6) delta. An additional EOG statistical feature, EOG IQR (the interquartile range of EOG value), also ranked highly in both models, at (4) fourth in Random Forest and (2) second in XGBoost.

IMU derived features appeared in both top 10 lists, though with lower importance

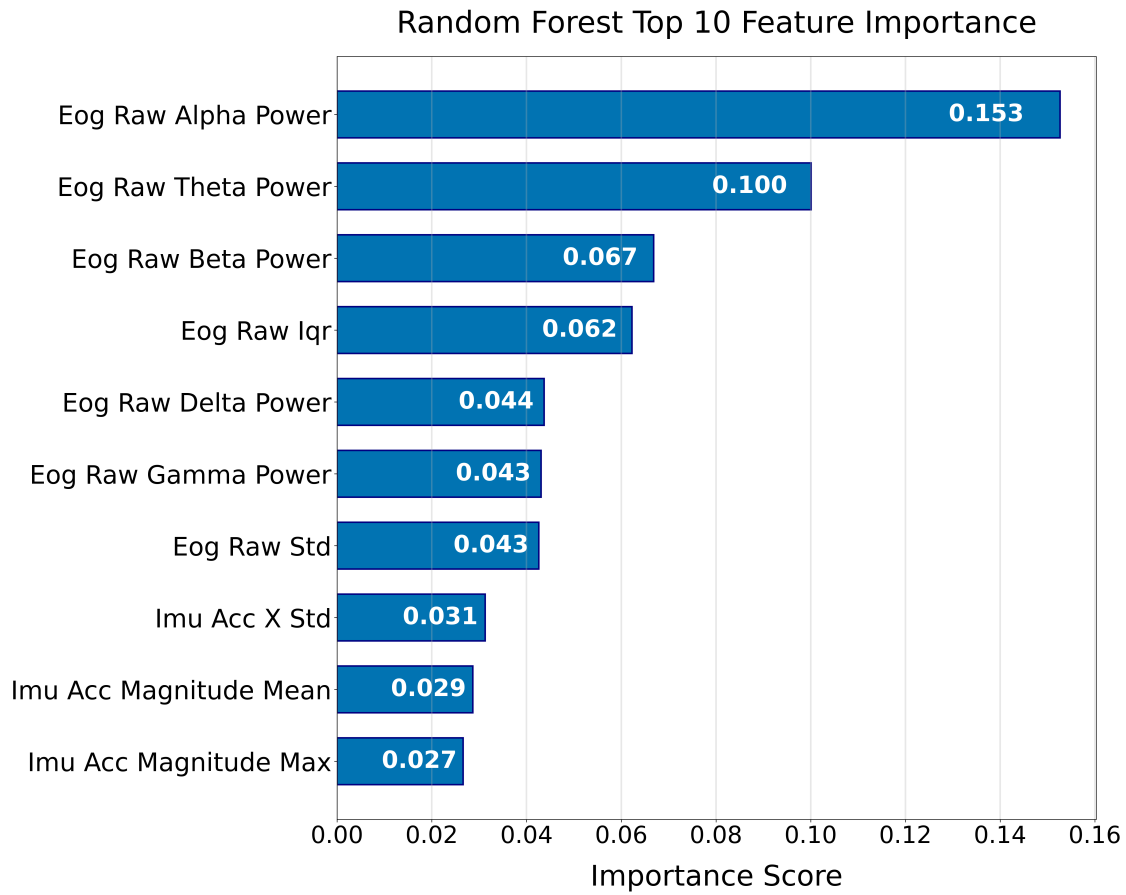


Figure 5.5: Top 10 Feature Importance Rankings for Random Forest Classifier using Mean Decrease in Impurity.

scores, while pulse derived features were absent from the top rankings in both models. The models showed some divergence in their lower-ranked features, with XGBoost including adaptive threshold features related to eye movements (local movement percentage and signal bursts) and Random Forest including more IMU derived accelerometer features.

5.3.6 Night-to-Night Classifier Performance

In order to assess the consistency of classification performance across test nights with varying conditions (e.g., sleep, data collection quality, etc.), the Random Forest classifier’s performance was examined for each of the three test nights. Table 5.3 presents these results together with contextual information from modified consensus sleep diary entries and the Somnify stage distribution for each night.

Random Forest performance varied by approximately 6 percentage points in F1-score across the three test nights, ranging from 0.738 on Night 13 to 0.796 on Night 17. This variation appears related to sleep and data quality conditions rather than representing arbitrary fluctuations. Night 13, which exhibited the lowest classifica-

Table 5.3: Random Forest Performance Across Test Nights with Sleep Context

	Night 3	Night 13	Night 17	Overall
Sleep Context				
Sleep Quality	Good	Very Poor	Fair	—
Device Stability	✓	×	✓	—
Total Observation Time	8h 45m	8h 33m	6h 20m	23h 38m
Somnofy Stage Distribution (%)				
Deep Sleep	17.7	13.4	19.5	16.6
Light Sleep	47.7	42.5	48.1	45.9
REM	25.9	20.6	17.3	21.7
Wake	8.7	23.6*	15.2	15.8
Random Forest Performance				
Accuracy	0.790	0.736	0.812	0.776
Macro F1-Score	0.778	0.738	0.796	0.770
Cohen’s Kappa	0.686	0.626	0.718	0.675

Note: Sleep Quality from Question 8 of Modified Consensus Sleep Diary (options: Very Poor, Poor, Fair, Good, Very Good). Device Stability indicates whether displacement was observed during sleep; displaced device remained on forehead area. Somnofy Stage Distribution shows percentage of 30-second epochs classified by reference device as Deep Sleep, Light Sleep, REM, or Wake. Overall column shows aggregate metrics across all three test nights (2,836 epochs total).

*Night 13 sleep diary included the comment “Onset insomnia. Trouble staying asleep.”

tion performance, was noted as being “Very Poor” in self-rated sleep quality, with comments citing, “Onset insomnia. Trouble staying asleep.” Indeed, the Somnofy platform labeled 23.6% of the recorded epochs as Wake, substantially higher than the 13.21% in the training and 15.80% in the test set. Night 13 was the only test night with device displacement, noted in the sleep diary as “Right shift.”

Night 17 demonstrated the highest classification performance with an F1-score of 0.796, despite having only “Fair” self-rated sleep quality and the shortest observation period at six hours and twenty minutes. Night 3, with good self-rated sleep quality, achieved strong performance with an F1-score of 0.778.

5.3.6.1 Visual Inspection of Classifier Performance with Hypnograms

A visual inspection of hypnogram overlays, shown in Figure 5.6, highlights that prediction-reference disagreements occur throughout the recording period rather than clustering at specific times, like towards the end of recording. Night 13 shows

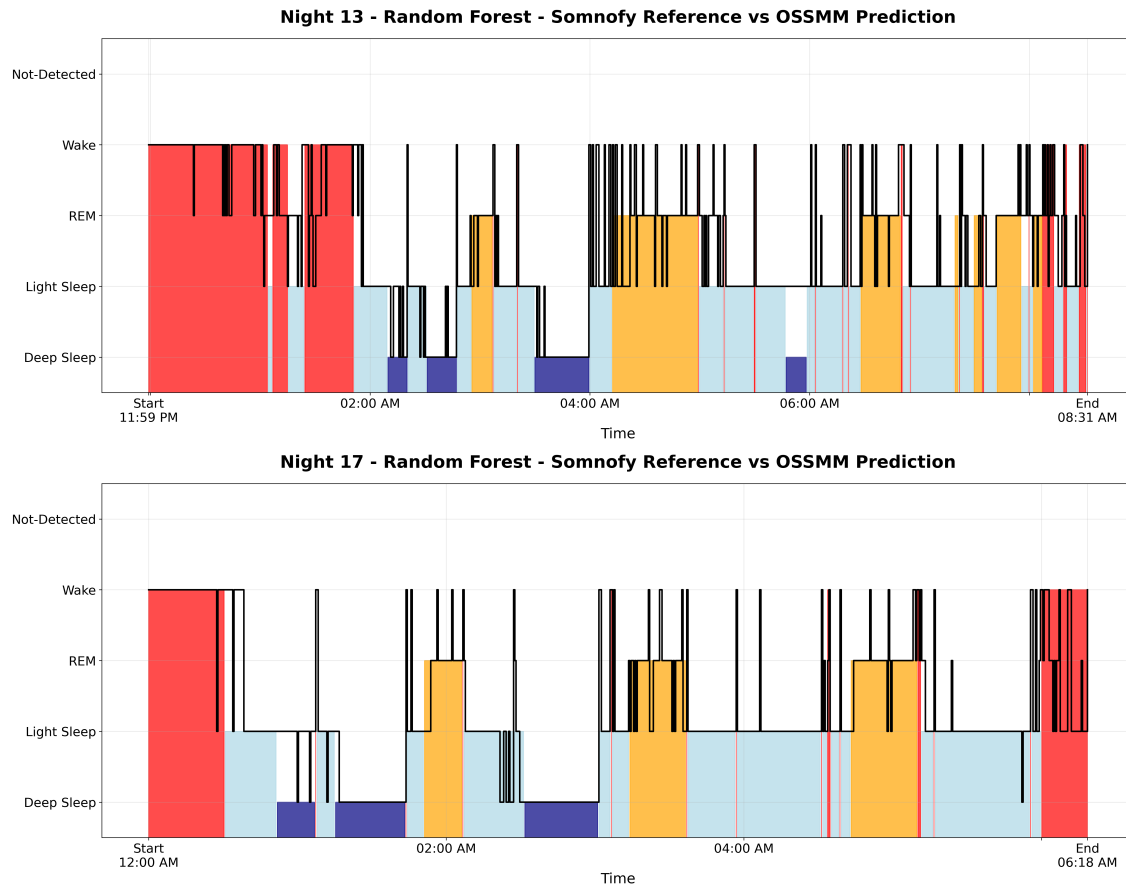


Figure 5.6: Hypnogram overlays comparing Somnify reference classifications (colored regions) with Random Forest predictions (black line). Shown are Test Night 13 (worst performance, $F1=0.738$) and Test Night 17 (best performance, $F1=0.796$). Night 13’s fragmented sleep architecture and device displacement correspond to increased prediction-reference disagreement, while Night 17 shows close tracking between methods. “Wake” is shown in red, “Light Sleep” in light blue, “Deep Sleep” in dark blue, and “REM” in orange.

increased mismatches corresponding to its fragmented sleep architecture and elevated wake percentage. Of note, both recordings show the OSSMM fail to predict virtually entire periods of deep sleep. For Night 13, this was the last deep sleep period, and for Night 17, this was the first. This corresponds to the confusion matrices which showed that “Deep Sleep” had the weakest F1 and accuracy performance, being confused with “Light Sleep” nearly a third of the time.

5.3.7 Feature Selection Importance by Ablation

The Random Forest classifier, having achieved the highest F1-score on testing data, was selected for an ablation study evaluating 16 feature combinations across the three sensor modalities: IMU, PPG, and EOG. Figure 5.7 shows the top 10 performing configurations, with complete performance metrics provided in Table 5.4.

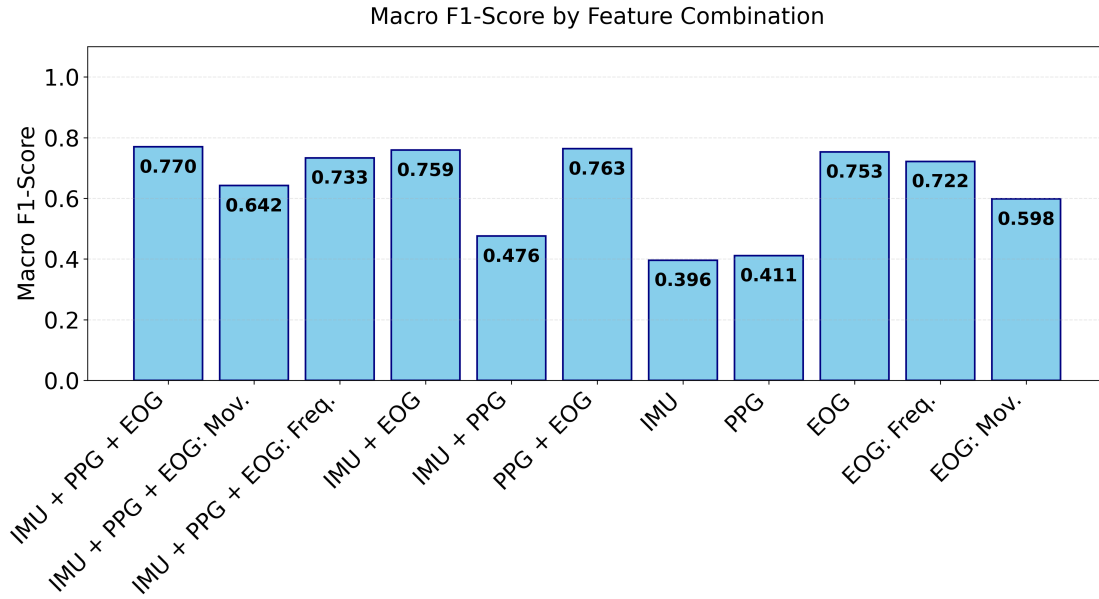


Figure 5.7: Weighted F1-score comparison for selected feature ablation configurations. Complete performance metrics for all 16 configurations are provided in Table 5.4.

The ablation study revealed EOG as the dominant sensor for sleep stage classification. EOG features alone achieved 97.8% of the full three-sensor system’s Macro F1 while using approximately half the features. Just five EOG frequency band features achieved 93.8% of full system performance (F1: 0.722 vs. 0.770), demonstrating that EEG-like spectral information captured substantial sleep architecture dynamics.

Among dual-sensor combinations, PPG + EOG and IMU + EOG performed comparably to the full system (F1: 0.763 and 0.759), while IMU + PPG achieved substantially lower absolute performance (F1: 0.476). However, IMU + PPG demonstrated the strongest synergy, improving approximately 15.8% when combined compared to either sensor alone. In contrast, adding IMU or PPG to EOG provided minimal synergistic gains (<2% improvement over EOG alone), suggesting that EOG features already capture most of the discriminative information, and offer limited complementarity with movement or cardiac features.

A notable finding emerged when comparing feature types: EOG frequency features alone (F1: 0.722) outperformed the combined IMU, PPG, and EOG movement features (F1: 0.642). This indicates that spectral characteristics of frontal EEG provide more discriminative information for sleep staging than head movement, cardiac dynamics, and eye movement detection combined, so far as the features which represent these measures.

IMU and PPG sensors showed limited individual utility. Both achieved similar baseline performance (F1: 0.396 and 0.411, respectively), with progressive degradation

when using feature subsets. PPG spectral features performed near chance levels (F1: 0.277), while heart rate variability features were only marginally better (F1: 0.327).

Table 5.4: Performance metrics for Random Forest classification across feature combinations. All models trained with SMOTE on training set and evaluated on withheld test set. Complete metrics including per-class precision and recall are provided in Appendix D under Section D.5.

Feature Combination	N	Accuracy	F1-Score	Cohen's Kappa	Macro AUC
<i>Multi-sensor combinations</i>					
IMU + PPG + EOG	42	0.776	0.770	0.675	0.941
IMU + PPG + EOG: Mov.	37	0.642	0.642	0.476	0.858
IMU + PPG + EOG: Freq.	25	0.748	0.733	0.632	0.927
<i>Two-sensor combinations</i>					
IMU + EOG	34	0.767	0.759	0.661	0.937
IMU + PPG	20	0.529	0.476	0.307	0.732
PPG + EOG	30	0.770	0.763	0.670	0.939
<i>IMU features</i>					
IMU (all)	12	0.448	0.396	0.206	0.681
IMU: Accelerometer	8	0.409	0.356	0.144	0.651
IMU: Gyroscope	4	0.361	0.325	0.082	0.583
<i>PPG features</i>					
PPG (all)	8	0.435	0.411	0.196	0.673
PPG: HR Mean	1	0.378	0.348	0.126	0.589
PPG: HR Spectral	3	0.300	0.277	0.053	0.548
PPG: HRV	2	0.353	0.327	0.106	0.600
<i>EOG features</i>					
EOG (all)	22	0.761	0.753	0.658	0.934
EOG: Frequency	5	0.734	0.722	0.620	0.914
EOG: Movement	17	0.604	0.598	0.422	0.799

5.4 Discussion

This proof-of-concept study evaluated the Open-Source Sleep Monitor and Modulator (OSSMM) system across three primary objectives: operational reliability over multiple nights, data quality sufficient for automated sleep staging, and the con-

tribution of individual sensor modalities to classification performance. The results demonstrate that the OSSMM platform successfully achieved its core design goals as a viable, accessible sleep monitoring solution, while also revealing important considerations regarding validation methodology and design optimization.

5.4.1 System Reliability

The OSSMM headband demonstrated robust operational performance throughout 17 nights of continuous use, maintaining 100% uptime with no hardware failures or component replacements required. The device exceeded its specified operational duration, functioning for over 16 hours on a single charge when inadvertently left uncharged during the protocol. The battery-powered system provided a concrete advantage during the study: when mains power interruptions occurred, the OSSMM continued data collection while the mains-powered Somnify SM-100 failed, resulting in the loss of two nights of reference data. While modern PSG systems may incorporate uninterruptible power supplies, this incident underscores the practicality of battery-powered systems for sleep monitoring.

Wireless data transmission via Bluetooth Low Energy proved highly reliable, with 99.99% of analyzed epochs meeting quality criteria ($\pm 1\%$ of expected sampling rate per epoch). The single over-sampling event likely reflected transient buffering rather than a systematic flaw.

The device’s physical stability during overnight wear was acceptable, with displacement detected on 6 of 17 nights but no instances of the headband falling off or leaving the forehead area entirely. These displacements were not significant enough to compromise data collection,

The combination of reliable hardware operation, stable wireless transmission, and successful overnight wearability confirmed the first steps necessary for the OSSMM platform to meet the fundamental requirements for longitudinal sleep monitoring. Importantly, these steps were achieved with a system costing under €40 in parts and designed for assembly by researchers with basic electronics knowledge.

5.4.2 Data Quality for Automated Sleep Staging

The OSSMM collected physiological data of sufficient quality for automated sleep stage classification, with the Random Forest classifier achieving an F1-score of 0.770 and Cohen’s κ of 0.675 on withheld test data. However, interpreting this performance requires critical consideration of the evaluation framework employed.

The Reference Ceiling Problem. The Somnify SM-100, used as the reference standard, achieved $\kappa=0.63$ (SD=0.10) when validated against PSG, while PSG inter-

rater reliability in the validation study was $\kappa=0.82$ (SD=0.10) (Toften et al., 2020). In other words, the Somnify SM-100 achieved 77% of PSG inter-rater reliability.

The OSSMM achieved $\kappa=0.675$ against the SM-100, exceeding Somnify’s own agreement with PSG. This result indicates a potential Reference Ceiling: the OSSMM system may have saturated the information content available in the Somnify labels. Without direct PSG validation, it is not possible to distinguish whether the OSSMM captures genuine sleep architecture better than Somnify, or the classifiers learned to replicate Somnify’s systematic biases.

This fundamentally affects interpretation of the results. The observed performance likely represents a lower bound constrained by the reference standard’s limitations rather than the OSSMM’s true capabilities. While this study does not aim to determine a performance metric of the OSSMM device, only that it can reliably collect signatures useful for automated sleep staging, this tempers the conclusions we can draw about potential absolute performance values.

Sleep stage misclassifications that agree with Somnify errors cannot be distinguished from genuine OSSMM limitations without gold-standard PSG validation.

Generalization and Robustness. Cross-validation estimates during hyperparameter search closely matched held-out test results. Test performances differed from CV by about -0.02 to +0.06 F1 points across models. This provides evidence that the feature set captured genuine sleep-related signals rather than overfitting to idiosyncrasies in the training data. With only 12 training nights and straightforward feature engineering, this generalization is non-trivial. The selection of non-consecutive test nights guarded against temporal confounds such as equipment degradation or physiological changes (e.g., stress, minor illness, etc.), ensuring that model evaluation reflected consistent performance across the study duration.

Night-to-Night Performance. The night-to-night performance analysis provides evidence that the OSSMM system responds appropriately to genuine variations in data quality and sleep architecture. The fluctuation in F1-scores demonstrates reasonable consistency across nights. Nights with moderate-to-good sleep quality, and no noted disturbances yield stable F1-scores: 0.778 and 0.796. In comparison, Night 13’s reduced performance (0.738) occurred under challenging circumstances: poor sleep quality, sleep fragmentation evidenced by elevated wake percentages, and device displacement during recording. Critically, the system continued to function and still achieved substantial agreement ($\kappa = 0.626$).

The system’s behavior, maintaining functional performance under poor conditions while achieving higher accuracy when conditions are optimal, indicates expected

sensitivity to data quality rather than random performance fluctuations. This supports the platform’s viability for longitudinal sleep monitoring where night-to-night variations in sleep quality and device wear are inevitable.

Classification Patterns and Sleep Architecture. Both Random Forest and XGBoost classifiers exhibited a “distance gradient” in misclassifications: adjacent sleep stages (e.g., Light and Deep sleep) were confused more frequently than distant stages (e.g., REM and Deep sleep). This is physiologically sensible because adjacent sleep states are more similar, and present some of the same characteristics despite being classified apart. Sleep stages represent points along a continuum of brain activity rather than being discrete categories. For example, even light sleep (N2) may contain delta waves, a classic signature of deep sleep ([AASM, 2023](#)). The specificity of certain misclassifications further supports this physiological interpretation. Neither model ever misclassified REM as Deep Sleep or vice versa. Deep Sleep was misclassified as Wake only 1.1–1.3% of the time, and Wake was never misclassified as Deep Sleep. These patterns suggest that the feature set captured genuine sleep architecture dynamics rather than arbitrary correlations.

The primary classification challenge occurred at the Light/Deep sleep boundary, where both decision tree methods showed reduced accuracy. Notably, the Somnofy device exhibits similar difficulties with this distinction in its original validation ([Toften et al., 2020](#)), suggesting this may reflect the reference standard’s limitations rather than deficiencies in the OSSMM’s sensor capabilities. The RBF kernel SVM classifier’s poor performance ($F1=0.598$) and tendency to over-predict Wake highlights the importance of algorithm selection for this application, though this comparison also suffered from minimal hyperparameter optimization.

Conservative Proof-of-Concept Approach. The deliberate use of limited hyperparameter exploration (Five configurations per classifier) served a specific validation purpose: demonstrating that OSSMM signals contain sufficient sleep signal information to be detected by conventional methods without extensive optimization. Had classification required exhaustive hyperparameter search or sophisticated feature engineering to achieve acceptable performance, this would raise concerns about the systems physiological signal capture capabilities. The success with minimal tuning provides confidence in the fundamental system design, that the hardware reliably captures signals with intrinsic discriminative power for sleep staging.

This interpretation is further supported by the stability of decision tree performance across hyperparameter configurations. Random Forest configurations varied by only 2.8 percentage points ($F1: 0.738\text{--}0.766$), while XGBoost demonstrated even tighter clustering with 1.4 percentage points of variation ($F1: 0.748\text{--}0.762$). This stabil-

ity indicates that the signals contain easily extractable discriminants that decision tree architectures can consistently identify regardless of parameter variations. In contrast, SVM performance varied dramatically across configurations (F1: 0.479–0.551), yet even its best configuration (F1=0.551) underperformed the worst decision tree method. This suggests the SVM’s struggles reflect an algorithmic mismatch to the problem of sleep staging structure rather than inadequate signal quality. Given that the SVM exploration was limited to RBF kernels, linear or polynomial kernels may yield different results.

5.4.3 Feature and Sensor Contribution

The ablation study revealed notable findings about sensor importance that have both scientific and practical implications for system design.

Dominance of Frontal EEG Information. The frontal EEG information by the EOG electrode configuration dominated classification performance. EOG-derived features alone achieved 97.8% of the full system’s F1-score (0.753 vs 0.770) while using only 22 out of 42 of the features. More remarkably, just the five EEG frequency band power features (alpha, beta, theta, delta, gamma) achieved F1=0.72, representing 93.8% of full-system performance with only 12% of the feature set.

The agreement of feature importance rankings between Random Forest and XGBoost strengthens confidence in these findings. Despite using different approaches to quantify importance (Mean Decrease in Impurity vs Gain), both algorithms identified the five EEG frequency band features as being within the six most important features. The rank order of frequency features was also equivalent, except for a reversal of delta and gamma, the last two features. This increases confidence that the rankings reflect genuine signal properties rather than artifacts.

This finding of dominant EEG features requires careful interpretation of what “EOG” means in this context. The OSSMM’s electrode configuration captures potential differences approximately equivalent to $(F7+Fp1) - (Fp2+F8)$ in the 10-20 system, using a differential amplifier. While designed for eye movement detection, this frontal montage evidently captures sufficient spectral EEG information to dominate sleep staging performance.

In contrast, features explicitly designed to detect eye movements (adaptive thresholds for saccades, local movement percentages, burst detection) contributed far less to classification accuracy. When EOG spectral features were excluded, leaving only movement-based EOG features, performance dropped substantially (F1=0.642). This pattern suggests that “EOG” as implemented in the OSSMM functions primar-

ily as a simplified frontal EEG channel that happens to also capture eye movements, rather than an eye movement detector that incidentally picks up some brain activity.

The Surprising Success of CTPU Electrodes. A notable finding is that these dominant EEG signatures were captured through CTPU electrodes with no conductive gel, and surface resistivity substantially higher than standard Ag/AgCl electrodes used in clinical PSG. The system successfully extracted neurophysiologically relevant frequency bands across the full spectrum from delta (<4 Hz) to gamma (30–100 Hz) despite these non-ideal electrode characteristics.

This result has significant implications for accessibility in sleep research. Traditional EEG recording requires conductive gel or paste, and trained technicians for electrode application. These contribute to PSG’s high cost, participant burden, and unsuitability for long-term home use. The CTPU electrodes integrated into the OS-SMM headband eliminate these requirements entirely. Users simply don the device and begin recording. As is noted in Chapter 4, visual detection of EEG signatures as would be typical in standard PSG data is not a given. However, that this approach yields sufficient signal quality for the most important classification features suggests a path toward more accessible EEG-based sleep monitoring.

Limited Contribution of Cardiac and Movement Sensors. In stark contrast to EEG’s dominance, the PPG and IMU sensors showed limited individual utility. Both achieved similar baseline performance in isolation ($F1 \approx 0.43$), and their combination (IMU+PPG) achieved similar results to “sleep monitoring capable” devices with these same sensors ($F1=0.476$) (Robbins et al., 2024; Schyvens et al., 2025). Adding either PPG or IMU to EOG provided minimal synergistic gains ($<2\%$ improvement), while the IMU+PPG combination showed the strongest synergy.

EOG frequency features alone ($F1=0.722$) substantially outperformed the combined IMU, PPG, and EOG movement features ($F1=0.642$). In other words, spectral characteristics of frontal EEG provide more discriminative information for sleep staging than head movement, cardiac dynamics, and eye movement detection combined. This in itself is not surprising, given that sleep states are primarily defined by EEG signatures (AASM, 2023). That the CTPU electrodes captured these signatures to sufficient quality for conventional Decision Tree methods for performant classification is.

Implications for System Design and Economics. From an engineering perspective, these ablation results provide guidance for design optimization. EOG sensor-derived features, particularly EEG frequency bands, dominated classification performance while PPG and IMU contributed minimally when EOG data was

available. The PPG sensor added hardware cost, assembly complexity, and potential failure points while contributing less than 2% performance improvement when combined with EOG.

These findings suggest that for applications focused specifically on sleep staging, simplified designs emphasizing robust EOG/EEG capture may be more cost-effective than multi-sensor approaches. However, design decisions should account for research objectives beyond sleep staging. PPG provides heart rate and heart rate variability information valuable for other sleep research applications (e.g., autonomic nervous system activity, detecting cardiac events, or studying sleep-cardiovascular interactions). The OSSMM’s modular design philosophy supports both configurations: researchers can include or exclude sensors based on their specific experimental requirements.

5.4.4 Limitations and Proof-of-Concept Scope

This study’s primary limitation is its single participant design ($N=1$). While all data derive from one individual, the 15-night dataset captured substantial variability across different physiological states, sleep quality levels, and data collection conditions, as demonstrated by the night-to-night performance analysis. The feature importance rankings, sensor ablation results, and absolute performance metrics reflect that the OSSMM approach can work rather than establishing a population level metric for system performance.

The investigation was constrained to single-participant testing due to the project timeline and resource limitations. The extensive engineering development across three prototype iterations over four years necessarily limited the scope for population-level validation within the PhD time frame. However, the open-source release of all hardware designs, software, and documentation explicitly enables the broader research community to conduct the multi-participant validation studies that represent the logical next phase of this work.

The single-participant design is, however, appropriate and sufficient for the study’s proof-of-concept objectives. The investigation sought to answer: (1) Can the hardware reliably collect data over multiple nights? (2) Is that data of sufficient quality for machine learning classification? (3) Do the different sensors contribute useful information? These questions were successfully answered. Establishing the system’s performance characteristics across diverse populations, age groups, sleep disorders, and user expertise levels remains an essential goal for future work but falls outside the scope of validating basic system viability.

The validation against Somnify rather than gold-standard PSG, while practical

and ethically appropriate for initial testing, introduces the Reference Ceiling problem discussed above. The OSSMM’s true capabilities relative to PSG-defined sleep stages remain unknown.

5.4.5 Conclusions

This proof-of-concept study demonstrates that the OSSMM platform successfully achieves its core design objectives: reliable hardware operation over multiple nights and data quality sufficient for automated sleep staging. These findings provide an affirmative answer to a principal question of this thesis: Given the design implementation of the OSSMM headband, the selected hardware and system programming, can signals useful for automated sleep staging be reliably collected?

The investigation establishes technical viability across three dimensions. The system maintained complete operational reliability throughout 17 nights of testing, with no hardware failures, robust wireless data transmission achieving 99.99% epoch quality, and acceptable overnight wearability despite minimal device displacement. The Random Forest classifier achieved Cohen’s $\kappa = 0.675$ on withheld test data using only basic hyperparameter tuning and straightforward feature engineering, demonstrating that the captured signals contain readily extractable sleep-relevant information. The ablation study indicated the relative contribution of each sensor modality, with findings that have implications beyond the immediate classification task.

A notable finding emerged from the sensor ablation analysis: frontal EEG features captured through inexpensive CTPU electrodes without conductive gel dominated classification performance, with just five frequency band features achieving an F1-score of 0.722 compared to the full system’s 0.770 (93.8% of full system performance). This result was unexpected given that CTPU electrodes have substantially higher surface resistivity than standard Ag/AgCl electrodes used in clinical and research polysomnography. The system successfully extracted neurophysiologically relevant frequency bands across the complete spectrum from delta to gamma despite these non-ideal electrode characteristics. While this finding requires validation across diverse participants to establish generalizability, it suggests that accessible EEG-based sleep monitoring may require less stringent electrode specifications than conventional approaches assume. The implications extend beyond the OSSMM platform itself, potentially informing future developments in wearable neurophysiological monitoring.

The reference system employed in this study, while appropriate for proof-of-concept evaluation, introduces constraints that temper conclusions about potential perfor-

mance values. The OSSMM achieved $\kappa = 0.675$ against the Somnify SM-100. This suggests the system may have saturated the information content available in the reference labels, such that observed performance represents a lower bound constrained by reference standard limitations rather than the OSSMM's true capabilities. Direct validation against manually scored polysomnography remains necessary to establish whether the system can exceed the performance of substantially more expensive alternatives. The component cost of under 40 euros represents not more than 2.3% of the Somnify system, positioning the platform as economically viable for research contexts where cost has previously been prohibitive.

The combination of demonstrated reliability, signal quality suitable for conventional machine learning methods, and significant cost reduction establishes the OSSMM as a viable platform for continued development. The success of frontal EEG features captured through CTPU electrodes suggests that more sophisticated analysis of this signal may yield further improvements. The next chapter investigates this possibility by applying convolutional neural networks directly to spectrogram representations of the EOG/EEG and PPG data, bypassing hand-crafted features entirely and potentially capturing temporal-spectral patterns not accounted for by the classical machine learning methods employed here.

Proof-of-Concept Evaluation: Deep Learning Approaches

“Give a man a fish and you feed him for a day. Teach a man to fish and you feed him for a lifetime. Give a neural network some fish and it will learn to identify fish on its own.”

— *Unknown*

Abstract

This chapter extends the proof-of-concept technical evaluation of the OSSMM platform by investigating whether convolutional neural networks (CNNs) applied to spectrograms can substantially improve upon the classification performance of conventional machine learning methods presented in Chapter 5. The analysis addresses two key questions. First, if spectrograms derived from EOG and PPG sensor data contain discriminative information beyond hand-engineered features. Second, if the sensor contribution ranking observed in Chapter 5, where EOG features dominated while PPG contributed minimally, persists when evaluated through deep learning architectures. The results demonstrate that a CNN classifier using EOG-derived spectrograms alone achieves stronger performance than conventional classifiers trained on hand-engineered features from all three OSSMM sensor modalities (EOG, PPG, IMU), providing further evidence that the OSSMM’s frontal electrode configuration captures genuine neurophysiological signatures suitable for sleep staging. Supplementary results are located within Appendix E.

6.1 Introduction

Chapter 5 demonstrated in a single participant design that the OSSMM platform captures physiological signals of sufficient quality for automated sleep stage classification, with Decision Tree classifiers (Random Forest $\kappa = 0.675$ and XGBoost $\kappa = 0.646$) achieving substantial agreement when evaluated against the Somnify SM-100 reference device. The ablation study revealed a dominance of EOG sensor derived features, particularly the five EEG frequency band power features (alpha, beta, theta, delta, gamma), which alone achieved 93.8% of full-system performance despite representing only 12% (5/42) of the total feature set. This finding suggests that the OSSMM’s frontal electrode configuration, while designed primarily for electrooculography, successfully captures spectral EEG information sufficient for sleep stage discrimination.

However, the conventional machine learning approach employed in Chapter 5 relied on hand-engineered features derived from basic signal processing techniques: spectral power features, statistical measures, and adaptive thresholding. Though these features demonstrated effectiveness, their design was constrained by assumptions about which signal characteristics would prove discriminative. Therefore, this chapter investigates whether convolutional neural networks (CNNs) applied directly to spectrogram representations can learn feature patterns not captured by hand-engineered approaches, potentially revealing discriminative information that conventional methods missed.

6.1.1 Deep Learning for Sleep Stage Classification

Convolutional Neural Networks (CNNs) have proven effective at automatically learning hierarchical feature representations from spatially-structured data, with applications spanning image classification, speech recognition, and time-frequency analysis of physiological signals (Krizhevsky et al., 2012; Simonyan and Zisserman, 2014). When applied to spectrograms (two-dimensional time-frequency representations of physiological signals), CNNs can leverage computer vision architectures to extract patterns from what are fundamentally time-series data (Li et al., 2022).

For sleep stage classification, this approach offers a compelling alternative to hand-engineered features. Rather than defining which frequency bands or statistical measures to extract, CNNs can potentially learn optimal feature representations directly from spectrogram images. Several studies have demonstrated CNNs’ effectiveness on polysomnography data, including single-channel EEG and spectrogram-based approaches (Li et al., 2022; Mousavi et al., 2019; Supratak et al., 2017). These studies consistently found that CNNs could extract relevant features supporting the

hypothesis that comprehensive multichannel recordings may not be necessary for sleep staging when deep learning is employed.

CNN-based classifiers applied to spectrograms offer certain advantages for the OS-SMM evaluation. Spectrograms provide a compact view of frequency content over time, potentially capturing transitions between sleep stages that may be difficult to characterize through basic feature calculations. More importantly, this approach allows direct assessment of whether Chapter 5’s key findings (the dominance of five frequency band features and the sensor hierarchy) reflect fundamental signal characteristics or limitations of the feature engineering methodology.

6.1.2 Study Objectives

Given the advantages of CNN-based classifiers, this chapter pursues two primary objectives. First, to investigate whether the EOG sensor (potentially capturing frontal EEG) and PPG sensor contain discriminative information beyond that captured by the hand-engineered features employed in Chapter 5. Specifically, can CNNs trained on spectrograms learn features that support improved classification performance compared to conventional classifiers using basic hand-engineered features?

Second, to assess whether the sensor modality importance ranking identified in Chapter 5, where EOG-derived features overwhelmingly dominated classification performance while PPG contributed minimally, persists when evaluated through deep learning architectures.

By training separate CNN models on spectrogram derived from the EOG sensor (now referred to as EOG/EEG spectrograms), the PPG sensor, and both modalities combined, this analysis can help determine whether the sensor hierarchy reflects the value of the physiological information content or artifacts of feature engineering.

These objectives serve a broader purpose within the proof-of-concept evaluation: establishing whether the OSSMM platform’s performance is fundamentally limited by the reference device’s hypothesized reference device ceiling (as described in Chapter 5) or whether more sophisticated feature extraction might overcome this limit. The convergence or divergence of conventional and deep learning approaches provides insight into this question.

6.2 Methods

6.2.1 Experimental Data

This investigation utilized the same 15-night dataset collected in Chapter 5. However, the conventional classifiers in Chapter 5 used predetermined hyperparameters identified through a basic search on the full training set. A validation set was used to enable fair hyperparameter comparison and early stopping based on generalization performance. Therefore, the original 12 training nights (12,449 epochs) were further subdivided: 9 nights for CNN training (9,461 epochs) and 3 nights for validation (2,988 epochs). The test set remained unchanged; the same 3 randomly selected non-consecutive nights (2,836 epochs), isolated from all model development.

All data were collected simultaneously from the OSSMM headband and Somnify SM-100 reference device, with Somnify classifications (Wake, Light Sleep, Deep Sleep, REM) serving as ground truth labels for 30-second epochs.

6.2.2 Spectrogram Generation

Spectrograms were generated separately for EOG and PPG channels to create visual representations of frequency content evolution over each 30-second epoch. This approach transforms the 1-D time-series signals into 2-D images suitable for CNN input while preserving temporal-spectral structure.

6.2.2.1 EOG/EEG Spectrograms

EEG data was collected through an EOG sensor, sampled at approximately 252 Hz, and filtered with hardware (AD8232 ‘Cardiac Monitor Configuration’) through a 0.5 Hz high-pass and 40 Hz low-pass. An additional digital 50 Hz notch filter (quality factor=30) was applied to remove potential mains electrical interference. Spectrograms were computed using the Short-Time Fourier Transform (STFT), with parameters shown in Table 6.1.

The resulting spectrograms, examples of which are shown in Figure 6.1d, captured frequency content from 0–60 Hz over the 30-second epoch duration, encompassing neurophysiologically relevant frequency bands: delta (<4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–60 Hz). Spectrograms were converted from amplitude to decibels, cropped (122 frequency bins), and normalized per-epoch to [0, 1] before saving as 126×122 pixel images.

Table 6.1: EEG Spectrogram Generation Parameters

Parameter	Value
Sampling Rate	≈ 252 Hz
Preprocessing	50 Hz notch filter ($Q=30$)
FFT Size (n_fft)	512 points
Window Length	504 samples (≈ 2.0 seconds)
Hop Length	56 samples (≈ 222 milliseconds)
Center Padding	False (no padding at signal boundaries)
Frequency Range	0–60 Hz
Frequency Resolution	≈ 0.49 Hz per bin
Time Windows per Epoch	≈ 126 (per 30-second epoch)
Dynamic Range	80 dB (top_db parameter)
Normalization	Per-epoch min-max normalization

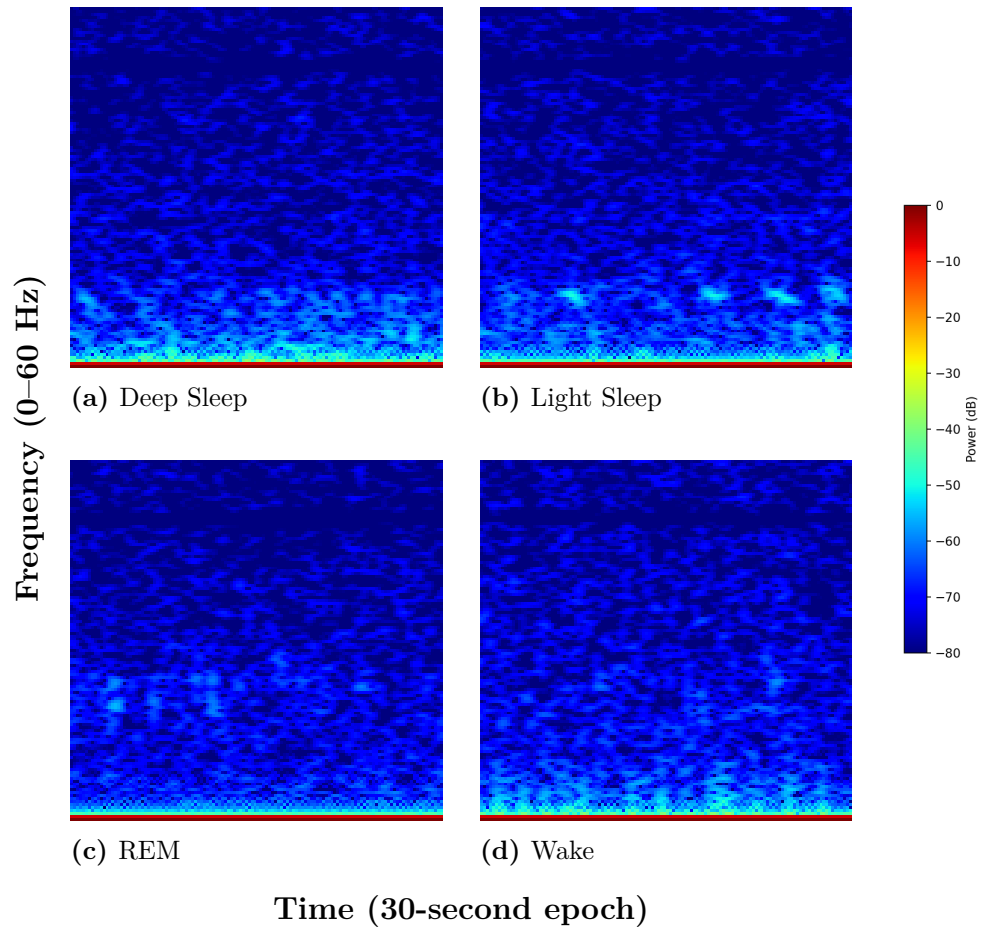


Figure 6.1: EOG Spectrograms from Night 17. Color bar represents power. Evidence of the 50Hz Notch filter can be seen near the top of each spectrogram. Epoch numbers for each observation are (a) 173 (b) 100 (c) 609 (d) 31.

Following spectrogram generation, a preliminary visual inspection was conducted to assess whether EEG sleep features were identifiable in the EOG/EEG spectrograms, potentially providing qualitative evidence of neurophysiological signal capture. Distinct patterns are evident in the spectrograms with Deep and Light Sleep showing the greatest similarity, notably high power at lower frequency ranges. REM and Wake show faint, higher frequency signatures which are absent in the NREM observations.

6.2.2.2 PPG Spectrograms

PPG data, sampled at approximately 252 Hz, underwent digital bandpass filtering between 0.3–12 Hz, using a 4th-order Butterworth. The bandpass range was selected to capture heart rate (0.5–3 Hz, corresponding to 30–180 BPM) and heart rate variability signatures. Following filtering, signals were downsampled to 50 Hz to reduce computational demands while maintaining adequate temporal resolution for cardiac features. Spectrograms were computed using STFT with parameters optimized for cardiac signal analysis, with parameters shown in Table 6.2.

The 10-second window length captures 6–7 cardiac cycles at even at low heart rates (40 BPM, 0.67 Hz), while 1-second hops provide temporal resolution for heart rate variability tracking. Spectrograms were converted to decibel scale, cropped to 0–12 Hz (123 frequency bins) to retain cardiac fundamentals and first harmonics, normalized per-epoch, and saved as 20x123 pixel PNG images.

The difference in spectrogram dimensions between EOG (126×122) and PPG (20×123) reflects the different temporal characteristics between neural and cardiac signals. EEG requires finer temporal resolution to capture rapid state transitions, while PPG benefits from longer integration windows for heart rate estimation.

Table 6.2: PPG Spectrogram Generation Parameters

Parameter	Value
Original Sampling Rate	≈ 252 Hz
Preprocessing	0.3–12 Hz bandpass filter (4th-order Butterworth)
Downsampled Rate	50 Hz
FFT Size (n_fft)	512 points
Window Length	500 samples (≈ 10.0 seconds)
Hop Length	50 samples (≈ 1.0 second)
Center Padding	False (no padding at signal boundaries)
Frequency Range	0–12 Hz
Frequency Resolution	≈ 0.098 Hz per bin
Time Windows per Epoch	≈ 20 (per 30-second epoch)
Dynamic Range	80 dB (top_db parameter)
Normalization	Per-epoch min-max normalization

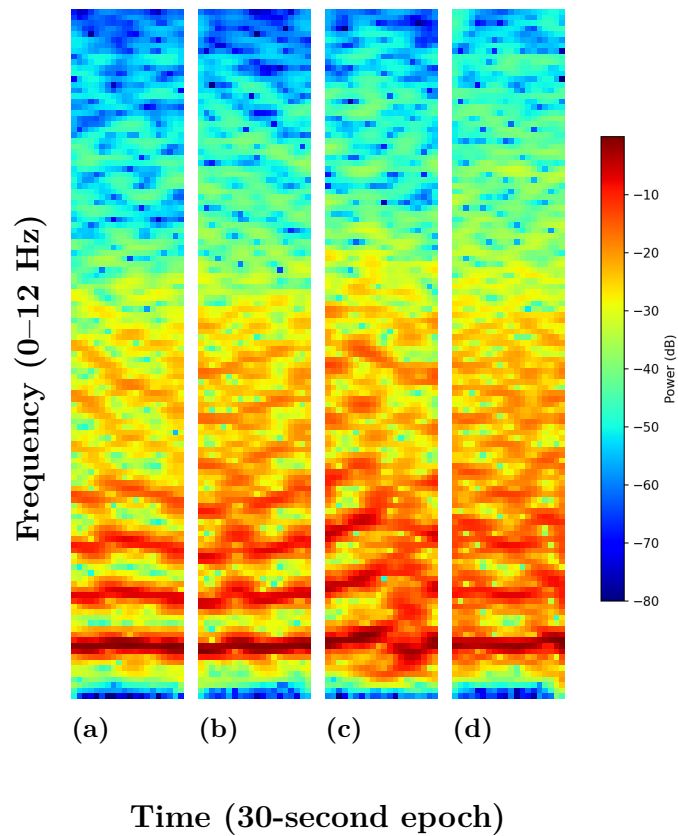


Figure 6.2: PPG Spectrograms from Night 17. The colorbar represents power. Sleep labels are for (a) Deep Sleep, (b) Light Sleep, (c) REM Sleep, and (d) Wake. Epoch numbers for each observation are (a) 173 (b) 100 (c) 609 (d) 31.

6.2.3 Convolutional Neural Network Architecture

Two model architectures were developed: a single-input CNN for processing either EEG or PPG spectrograms independently, and a dual-input CNN for processing both modalities combined.

The single-input branches are a simplified, modernized version of VGG (Simonyan and Zisserman, 2014). Each branch consists of 3 blocks with 6 convolutional layers using 3×3 filters ($padding = 1$), and adapted for smaller input dimensions (126×120 and 20×123 pixels). Modernizations include BatchNorm (Ioffe and Szegedy, 2015) after each convolutional layer and spatial dropout (Dropout2d) after each block. This lightweight architecture supports potential deployment on edge devices (e.g., the OSSMM Android app) while maintaining classification performance.

6.2.3.1 Single-Input CNN Architecture

The single-input architecture (Figure 6.3) consists of three convolutional blocks followed by fully connected classification layers. Each convolutional block employs the pattern: Conv2D $\rightarrow$ BatchNorm $\rightarrow$ ReLU $\rightarrow$ Conv2D $\rightarrow$ BatchNorm $\rightarrow$ ReLU $\rightarrow$ MaxPool2D $\rightarrow$ Dropout. This design progressively increases channel depth while reducing spatial dimensions, enabling the network to learn hierarchical feature representations from low-level to high-level patterns.

Table 6.3: CNN Architecture for EEG and PPG Sleep Stage Classification

Layer	EEG	PPG
Input Dimension	$126 \times 122 \times 1$	$20 \times 123 \times 1$
Convolutional Blocks (Shared):		
Block 1	32 filters (3×3), MaxPool (2×2), Dropout (0.4)	
Block 2	64 filters (3×3), MaxPool (2×2), Dropout (0.4)	
Block 3	128 filters (3×3), MaxPool (2×2), Dropout (0.4)	
Extracted Features	28,800	3,840
Classification Head (Shared):		
FC Layer 1	$\rightarrow$ 512 units (BatchNorm, ReLU, Dropout 0.4)	
FC Layer 2	$\rightarrow$ 256 units (BatchNorm, ReLU, Dropout 0.4)	
Output Layer	$\rightarrow$ 4 classes (Softmax)	

The EEG model contains 15,167,332 trainable parameters, while the PPG model contains 2,387,812 trainable parameters due to the smaller input feature dimension. All convolutional layers use same-padding to preserve spatial dimensions before

pooling. Batch normalization stabilizes training by normalizing activations, while dropout provides regularization by randomly deactivating neurons during training. Identical training procedures were applied to spectrograms from both sensors, allowing direct assessment of their relative information content for sleep stage classification, and not performance due to hyperparameter selection.

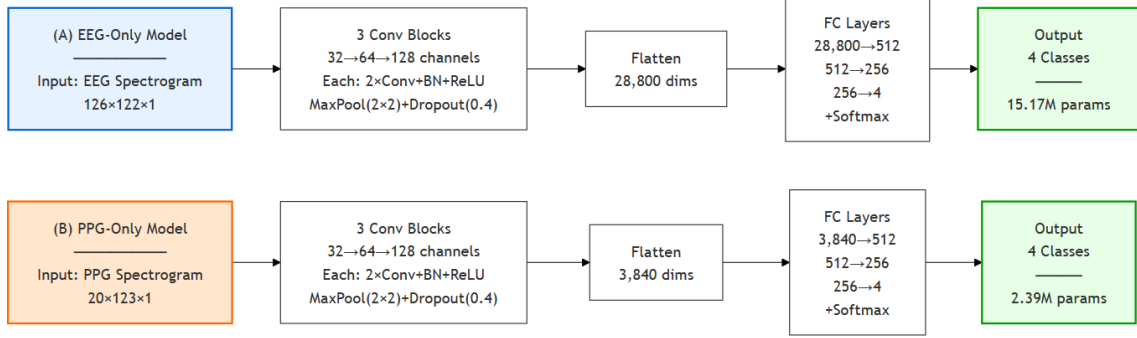


Figure 6.3: Single-input CNN architecture for processing EEG or PPG spectrograms. The network consists of three convolutional blocks that progressively increase channel depth (32→64→128) while reducing spatial dimensions through max pooling. Each block includes batch normalization and dropout (0.4) for regularization. Feature maps are flattened and passed through two fully connected layers for classification into four sleep stages. The different input dimensions (EEG: 126×122 , PPG: 20×123) result in different feature counts after convolution but identical classifier architectures.

6.2.3.2 Dual-Input CNN Architecture

The dual-input architecture processes EEG and PPG spectrograms through separate but identical convolutional branches before fusing features for classification (Figure 6.4). Each branch follows the same three-block convolutional architecture as the single-input model, with feature concatenation occurring after flattening: This architecture contains 17,420,740 trainable parameters. The dual-branch design allows each modality to develop specialized feature representations before integration, potentially capturing complementary information. The architecture preserves the dimensionality of each spectrogram type rather than forcing uniform input sizes. This allows each branch to process its respective signal at an appropriately suited temporal-spectral resolution.

Table 6.4: Multi-Modal Fusion Architecture

Stage	Configuration
Parallel Feature Extraction:	
EEG Branch	$126 \times 122 \times 1 \rightarrow 3 \text{ Conv Blocks} \rightarrow 28,800 \text{ features}$
PPG Branch	$20 \times 123 \times 1 \rightarrow 3 \text{ Conv Blocks} \rightarrow 3,840 \text{ features}$
Feature Fusion:	
Concatenation	$28,800 + 3,840 = 32,640 \text{ features}$
Joint Classification:	
FC Layer 1	$32,640 \rightarrow 512$ (BatchNorm, ReLU, Dropout 0.4)
FC Layer 2	$512 \rightarrow 256$ (BatchNorm, ReLU, Dropout 0.4)
Output Layer	$256 \rightarrow 4 \text{ classes}$ (Softmax)

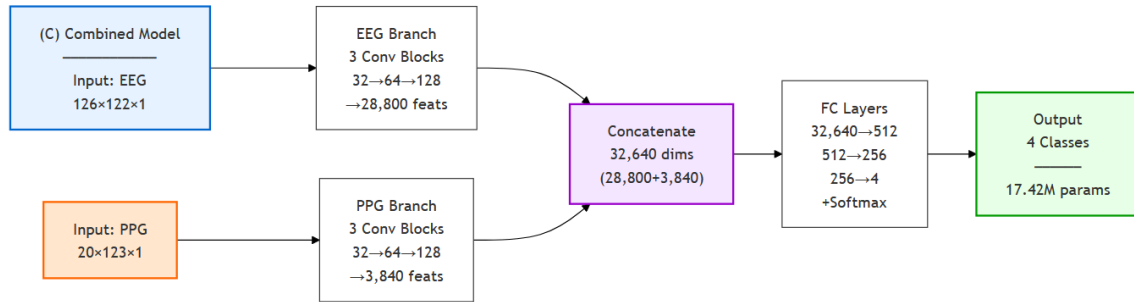


Figure 6.4: Dual-input CNN architecture for processing EEG and PPG spectrograms simultaneously. Two identical convolutional branches (in structure, not parameters) process each modality independently, preserving their natural temporal-spectral resolutions. The EEG branch outputs 28,800 features while the PPG branch outputs 3,840 features, reflecting their different input dimensions. Features are concatenated (32,640 total) before passing through fully connected layers for classification. This architecture contains 17.4 million trainable parameters.

6.2.4 Hyperparameter Selection and Regularization

The two research objectives of this chapter were at tension due to different optimization targets. The first objective (i.e., can CNNs exceed the RF results?) requires optimized performance. The second objective (i.e., does a sensor hierarchy persist?) requires a fair comparison between modalities where the performance differences are attributed to information capture, not hyperparameter optimization. To address these competing objectives, a single regularization configuration was selected through a simple systematic hyperparameter exploration, ensuring fair comparison between sensor modalities while optimizing for generalization to unseen data.

The systematic hyperparameter exploration was conducted on the training data. Regularization strength was evaluated at three levels: low, medium, and high. Table 6.5 presents the detailed configurations. Low regularization employed minimal constraints to allow detection of weak signals, Medium regularization employed moderate constraints as a balanced baseline, and Heavy regularization employed strong constraints to suppress potential overfitting to noisy data.

Table 6.5: Hyperparameter configurations tested for regularization sensitivity analysis. All configurations used a learning rate of 0.0005, label smoothing of 0.1, and batch sizes optimized for GPU memory (256).

Configuration	Dropout Rate	Weight Decay	Color Jitter
Low Regularization	0.2	1×10^{-5}	0.15
Medium	0.4	1×10^{-4}	0.2
Heavy Regularization	0.6	1×10^{-3}	0.3

For each modality and configuration, separate single-branch CNN models were trained using the architecture described in Section 6.2.3 with balanced sampling through class-weighted random sampling. Model selection and hyperparameter evaluation were performed using four-fold group cross-validation by recording night, with each fold containing nine training nights and three validation nights. This group-based splitting strategy prevented data leakage by ensuring all epochs from a given recording session appeared in either the training or validation set, but never both.

All models were trained using the AdamW optimizer with an initial learning rate of 0.0005 and label smoothing of 0.1 to improve generalization. A ReduceLROnPlateau learning rate scheduler reduced the learning rate by a factor of 0.5 after five epochs without improvement in validation loss, with a minimum learning rate of 6.25×10^{-5} . Training proceeded for a maximum of fifty epochs with early stopping based on validation F1 score using a patience of five epochs. Gradient clipping with a maximum norm of 1.0 was applied during backpropagation to stabilize training.

To enable a fair comparison between sensor modalities, a single regularization configuration was selected to use across all three architectures (EEG-only, PPG-only, and EEG+PPG). Hyperparameter selection balanced three considerations: Macro F1-Scores (mean cross-validation performance), F1-Score Standard Deviations (consistency across folds), and training-validation accuracy gap (generalization to validation data). Table 6.6 presents the four-fold cross-validation results for all tested

configurations. The Gap column reports the mean training-validation accuracy difference across folds.

Table 6.6: Four-fold cross-validation performance across hyperparameter configurations. Low, Medium, and Heavy describe the configuration’s relative regularization.

Configuration	Val Acc (Std)	F1 (Std)	κ (Std)	Gap (Std)
EEG				
Low	0.788 (0.012)	0.779 (0.015)	0.691 (0.019)	0.061 (0.068)
Medium	0.782 (0.011)	0.773 (0.013)	0.684 (0.018)	0.024 (0.036)
Heavy	0.777 (0.009)	0.768 (0.012)	0.678 (0.017)	0.009 (0.019)
PPG				
Low	0.569 (0.059)	0.548 (0.047)	0.379 (0.058)	0.199 (0.060)
Medium	0.536 (0.063)	0.532 (0.060)	0.360 (0.069)	0.132 (0.085)
Heavy	0.489 (0.087)	0.482 (0.093)	0.314 (0.086)	0.077 (0.042)

Note: Gap = Train Acc – Val Acc at best validation epoch. Low: dropout = 0.2, weight_decay = 0.00001, color_jitter = 0.15. Medium: dropout = 0.4, weight_decay = 0.0001, color_jitter = 0.2. Heavy: dropout = 0.6, weight_decay = 0.001, color_jitter = 0.3.

Given the limited test set size (three nights), configurations demonstrating stable generalization were prioritized over marginal validation performance gains. While Low Regularization achieved the highest mean F1 scores (EEG: 0.779, PPG: 0.548), it exhibited the most variable fitting behavior, with mean training-validation accuracy gaps of 6.1 percentage points on EEG and 19.9 percentage points on PPG. Individual fold gaps reached as high as 25.1 percentage points (PPG Fold 1), indicating the model was memorizing training-specific patterns rather than learning generalizable signatures. These large fold-to-fold variations raised concerns about reliable generalization. Heavy Regularization provided the most consistent fitting behavior (EEG Gap: 0.009, PPG Gap: 0.077) but notably reduced mean performance to on PPG data (PPG F1: 0.482).

Medium regularization balanced these considerations, achieving F1 scores within 0.006 of Low Regularization on EEG (0.773 vs 0.779) and 0.016 on PPG (0.532 vs 0.548) while maintaining substantially more stable fitting patterns across folds with training-validation accuracy gaps of 2.4 percentage points on EEG and 13.2 percent on PPG. The Medium configuration (*dropout* = 0.4, *weight_decay* = 0.0001, *color_jitter* = 0.2) was therefore selected for all subsequent experiments.

6.2.5 Training Protocol

Using the selected Medium regularization configuration, three final models (EEG-only, PPG-only, and EEG+PPG) were trained on all 12 training nights with identical optimization settings as described in Section 6.2.4. The training set was split 80-20 by recording night (random seed=42) for training and validation monitoring. Class imbalance was addressed through weighted random sampling, with sampling weights computed as the inverse of class frequencies. Here, a *training epoch* refers to one complete pass over the training set using the CNN, distinct from the 30-second *recording epochs* that make up the data. Each CNN model was trained for up to 50 such training epochs, with early stopping based on validation performance.

Models were trained separately for each sensor configuration, with the single-branch CNN architecture used for EEG-only and PPG-only models, and the dual-branch architecture with feature concatenation used for the combined EEG+PPG model. The best model checkpoint (highest validation F1-score) was selected for evaluation on the three withheld test nights.

6.2.6 Performance Metrics

To maintain continuity with the conventional models (see Section 5.2.8), all evaluations are computed at the 30 second epoch level over $K = 4$ classes (Deep, Light, REM, Wake) and are macro-averaged to mitigate class imbalance. Macro F1, Balanced Accuracy (Macro Recall), and Cohen’s κ are reported as the primary performance metrics. The Macro F1-Score is used for the primary selection criterion. Definitions and formulas are identical to those in Section 5.2.8. Unless otherwise stated, all accuracy and F1 values are macro-averaged.

6.2.7 Implementation and Reproducibility

All models were implemented in PyTorch with torchvision, and trained on a single GPU using automatic mixed precision (AMP). Each 30-second recording epoch was converted to two spectrogram images, one EEG and one PPG, which served as the possible model inputs; therefore the dataset sizes in Table 6.7 are counts of 30-second epochs, each represented by an image. Class imbalance was handled with inverse-frequency weighted random sampling, and a fixed random seed (42) was applied to Python, NumPy, and PyTorch (including CUDA) to support reproducibility. The compute environment, dataset partition, and training configuration are summarized in Table 6.7.

Table 6.7: Compute environment, dataset partition, and training configuration for the CNN experiments. Training times are elapsed real times, including data loading, evaluation, and plotting.

Item	Value
Compute environment	
GPU	NVIDIA GeForce RTX 4080 Laptop GPU (12 GB)
CPU	Intel Core i7-13700HX (13th Gen, 2.10 GHz)
RAM	32 GB (5600 MT/s)
Operating system	Windows 11
Software	Python 3.12.7, PyTorch 2.6.0, CUDA 12.6
Dataset	
Total valid epochs	15,285 (15 nights)
Training / Validation	9,461 (9 nights) / 2,988 (3 nights)
Withheld test	2,836 (3 nights)
Inputs per epoch	EEG, PPG, or both spectrograms (PNG; model-dependent)
Training configuration	
Optimizer	AdamW (decoupled weight decay)
Learning rate	5×10^{-4} , ReduceLROnPlateau (factor 0.5)
Loss	Label-smoothing cross-entropy (0.1)
Batch size	256
Max training epochs / early stopping	50 / patience 5 (search: 3)
Gradient clipping	max-norm 1.0
Mixed precision	Enabled (AMP)
Random seed	42
Training (single GPU)	
EEG model	10 epochs (early stopping)
PPG model	20 epochs (early stopping)
EEG+PPG model	9 epochs (early stopping)
All three models (train + evaluation + plots)	16 min 51 s
Hyperparameter search	1 h 56 min (24 model fits)

6.3 Results

6.3.1 Visual Inspection of EEG/EOG Spectrogram

A preliminary visual inspection of the EOG/EEG spectrograms revealed events with typical EEG sleep spindle characteristics (Nayak and Anilkumar, 2023). Figure 6.5 shows an epoch scored as Light Sleep (N1+N2) together with the corresponding EOG time-series. In the EOG/EEG spectrogram here, these phenomena appear as brief bursts of activity (approximately 0.5–2 seconds), centered around 12–13 Hz. As expected for spindle-like activity, these events are predominant in epochs labeled as Light Sleep (N1+N2), and less so in epochs labeled as Deep Sleep.

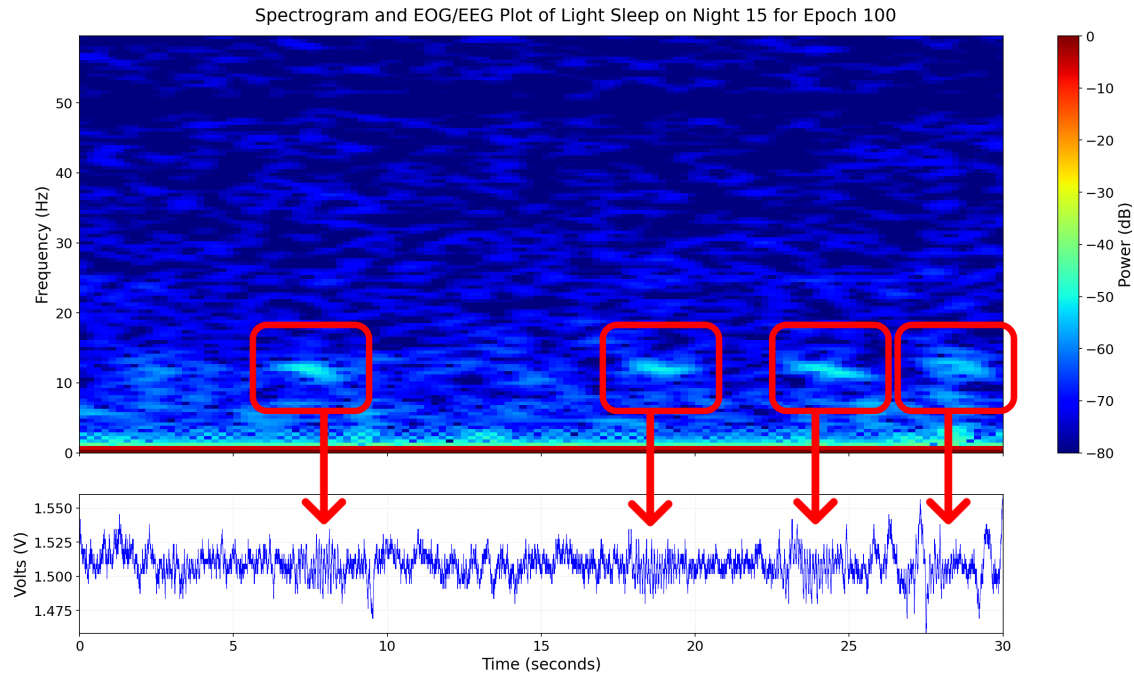


Figure 6.5: EOG channel spectrogram and time-series plots of epoch 100 from Night 17 using the OSSMM headband. The y-axis shows frequency (0–60 Hz) and the x-axis shows time (30-second epoch). Four sleep spindles are highlighted in the spectrogram, with corresponding arrows to the location in the time-series plot below. The 50 Hz notch to remove potential mains line interference can be seen at the corresponding frequency.

These phenomena are considered as sleep spindle-like activity based on (1) their frequency, (2) short duration, (3) predominance in Light Sleep labelled epochs, and (4) qualitative agreement with a published PSG EEG example (Li et al., 2022; Nayak and Anilkumar, 2023). For comparison, Figure 6.6 reproduces a PSG EEG spectrogram and time-series plot from Li et al. (2022), where spindles are highlighted and exhibit similar time-frequency structures to those observed in the OSSMM derived spectrograms (Li et al., 2022). This supports the inference that the OSSMM EOG channel can capture EEG signatures, including spindles.

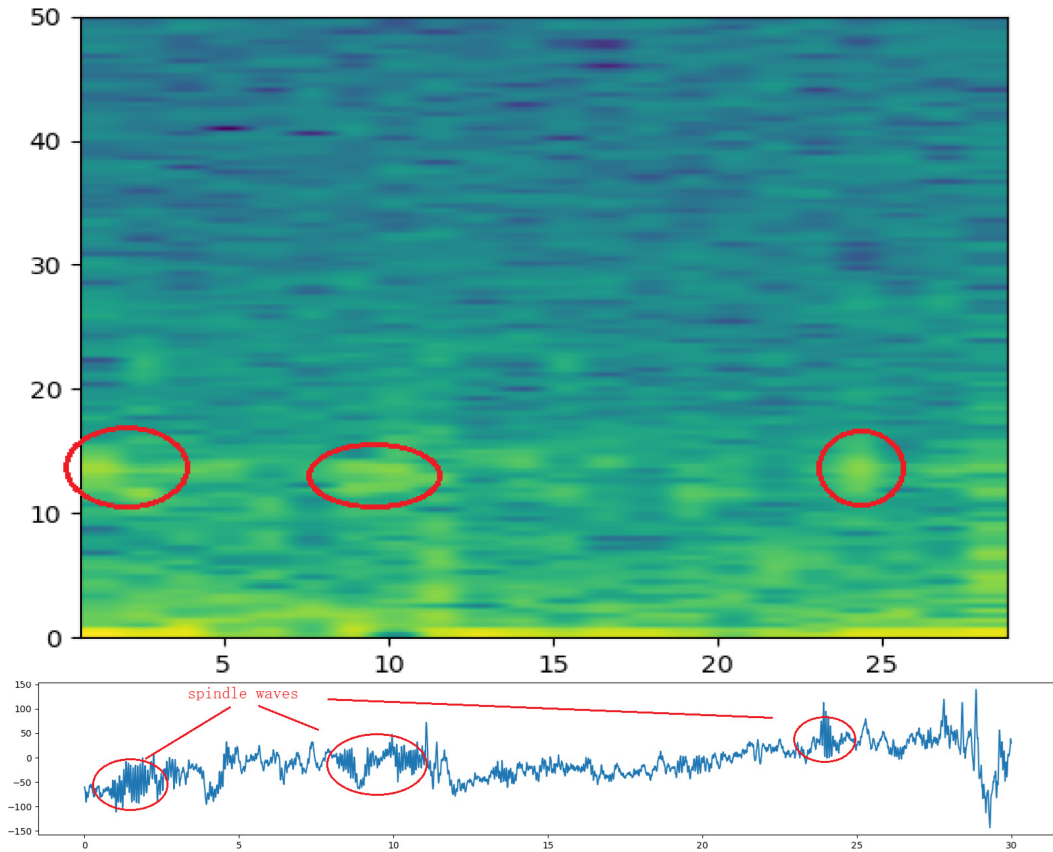


Figure 6.6: PSG EEG spectrogram with spindle detection highlighted in red ovals, and EEG time-series plot below. Figures combined and temporally aligned from Figures 1 and 2 in (Li et al., 2022).

6.3.2 Classification Performance on Test Set

Three CNN models (EEG-only, PPG-only, and EEG+PPG dual-input) were evaluated on the combined test set comprising 2,836 epochs across three withheld nights (Night 3, 13, and 17). Models converged rapidly through early stopping at epochs 9 (EEG+PPG), 10 (EEG), and 20 (PPG). Each of the three CNN models displayed higher performance over its Random Forest counterpart from Chapter 5. Table 6.8 presents CNN model performance alongside those of Random Forest for direct comparison.

The EEG-only CNN achieved an F1-score of 0.805 and Cohen’s κ of 0.727 on the test set, representing improvements of 5.2 and 6.9 percentage points respectively over the Random Forest classifier (F1=0.753, κ =0.658). The combined EEG+PPG CNN achieved the highest overall performance with F1=0.810 and κ =0.733, exceeding the full Random Forest system using all 42 features from three sensor modalities (IMU, PPG, EOG) by 4.0 F1 points and 5.8 κ points. Notably, the CNN’s Cohen’s κ of 0.733 exceeds the Somnify SM-100 reference device’s agreement with polysomnography (κ =0.63) reported in its validation study (Toften et al., 2020).

Table 6.8: Comparison of CNN and Random Forest performance on test set

Model Type	Input	Acc.	Macro F1	κ
<i>EEG/EOG</i>				
CNN	EOG/EEG spectrogram	0.807	0.805	0.727
Random Forest	EOG (22 features)	0.761	0.753	0.658
Change		+0.046	+0.052	+0.069
<i>PPG</i>				
CNN	PPG spectrogram	0.486	0.481	0.304
Random Forest	PPG (8 features)	0.435	0.411	0.196
Change		+0.051	+0.070	+0.108
<i>EEG/EOG + PPG</i>				
CNN	EEG+PPG spectrograms	0.809	0.810	0.733
Random Forest	EOG+PPG (30 feat.)	0.770	0.763	0.670
Change		+0.039	+0.047	+0.063
<i>Full System</i>				
CNN	EEG+PPG spectrograms	0.809	0.810	0.733
Random Forest	IMU+PPG+EOG (42 feat.)	0.776	0.770	0.675
Change		+0.033	+0.040	+0.058

Note: CNN results from combined test set (three withheld nights, 2,836 epochs). Random Forest results from Chapter 5. CNN models use spectrograms; RF uses hand-engineered features. Changes in performance metrics show CNN advantage over corresponding RF configuration.

PPG performance remained poor across both methodologies. The PPG-only CNN achieved F1=0.481 and κ =0.304, representing improvements of 7.0 and 10.8 percentage points over Random Forest (F1=0.411, κ =0.196). However, this barely exceeded simple baselines on the imbalanced test set (class ratio approximately 9:4:3:3; Light Sleep 45.9%, REM 21.7%, Deep Sleep 16.6%, Wake 15.8%). Stratified random guessing would achieve 31.0% accuracy and a naive majority-class classifier 45.9%; the PPG-only CNN’s 48.6% exceeded these by 17.6 and 2.7 percentage points respectively.

Figure 6.7 presents the normalized confusion matrix for the EEG+PPG model. The model achieved strong per-class F1-scores: Deep Sleep (0.865), Light Sleep (0.806), REM (0.801), and Wake (0.767). Deep Sleep demonstrated the highest classification recall (95.6%), followed by REM (85.2%) and Wake (84.2%). Light Sleep exhibited the lowest recall (72.4%). This is a shift from Chapter 5’s Random Forest analysis: Deep Sleep now shows the highest recall (95.6%) instead of the lowest (69.3%). This

Normalized Confusion Matrix - EEG + PPG - All Nights

True	Deep Sleep	95.6%	4.2%	0.0%	0.2%
	Light Sleep	9.0%	72.4%	9.3%	9.3%
	REM	0.2%	8.8%	85.2%	5.9%
	Wake	0.4%	4.7%	10.7%	84.2%
		Deep Sleep	Light Sleep	REM	Wake
		Predicted			

Figure 6.7: Normalized confusion matrix for the combined EEG+PPG CNN model on the combined test set (2,836 epochs across three withheld nights). Deep Sleep achieved the highest classification accuracy (95.6%), while Light Sleep showed the most distributed confusion pattern with misclassifications primarily to REM (9.3%), Wake (9.3%), and Deep Sleep (9.0%).

improvement stems from a reduction in misclassifications with Light Sleep, which was previously at 29.4% of Deep Sleep epochs. However, it is now Light Sleep which has a lower recall (72.4%) compared to RF (80.2%), and this reduction is principally due to increased misclassifications as Deep Sleep.

6.3.3 Sensor Modality Hierarchy

A central objective of this investigation was to determine whether the sensor contribution hierarchy observed in Chapter 5, where EOG-derived features dominated classification performance while PPG contributed minimally, persists when evaluated through deep learning architectures. This comparison directly addresses whether the sensor hierarchy reflects genuine physiological information content or artifacts of hand-engineered feature selection.

Figure 6.8 demonstrates the clear performance hierarchy across combined and individual testing nights. The EEG-only model achieved $F1=0.805$ on all test data, representing 99.4% of the combined EEG+PPG model's performance ($F1=0.810$). The addition of PPG spectrograms provided only 0.5 percentage points improvement,

despite adding 2.3 million parameters (15% increase) and requiring dual-branch architecture with separate feature extraction pathways.

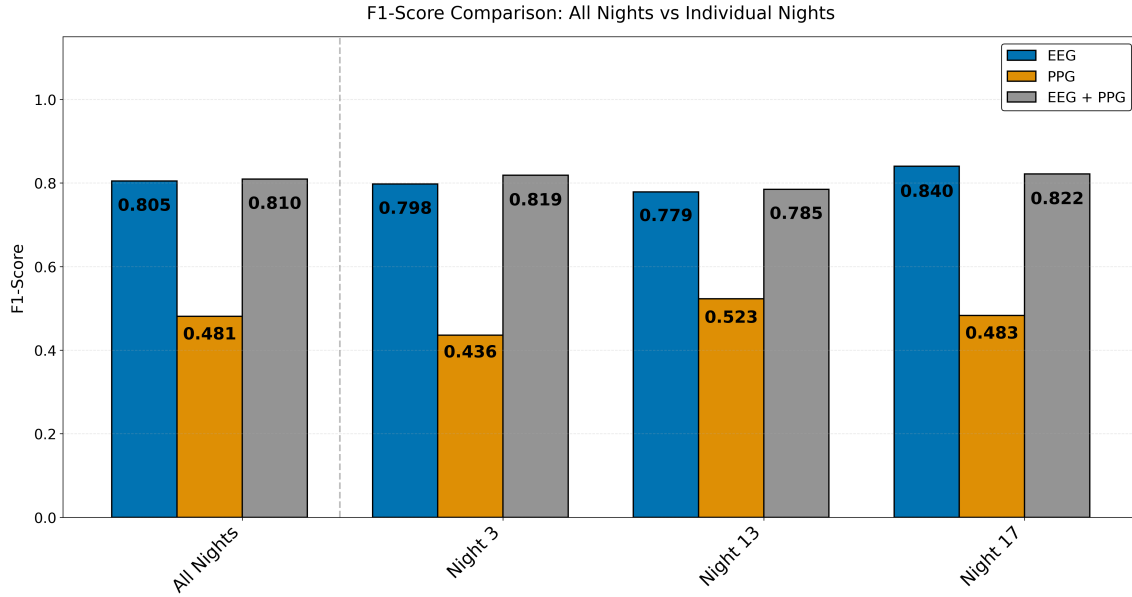


Figure 6.8: F1-score comparison across combined test set and individual test nights for EEG, PPG, and EEG+PPG CNN models. The vertical dashed line separates aggregate performance (left) from individual night performance (right). EEG and EEG+PPG models show consistent performance across nights, while PPG exhibits greater variability with high performance on Night 13.

6.3.4 Night-to-Night Performance Variability

To assess consistency across different sleep conditions and data quality scenarios, model performance was evaluated separately on each of the three withheld test nights. Figure 6.8 presents F1-scores for all three models across individual nights and the complete test set.

The EEG-only model demonstrated robust performance across test nights, with F1-scores ranging from 0.779 (Night 13) to 0.840 (Night 17). Night 13, which exhibited the poorest EEG performance, corresponds to the same night that showed the lowest Random Forest performance in Chapter 5 (F1=0.738). That night was documented in the sleep diary as having "Very Poor" sleep quality with sleep diary comments noting "Onset insomnia. Trouble staying asleep," along with observed device displacement. This consistency across methodologies suggests genuine night-specific factors (sleep fragmentation, data quality, device positioning) rather than model specific instability.

The EEG+PPG combined model showed similar robustness, with F1-scores spanning 0.785 (Night 13) to 0.822 (Night 17), a 3.7 percentage point range. The com-

binned model’s relatively tighter performance range compared to EEG-only suggests that the dual-input architecture may provide modest stability benefits, though the improvement is marginal. It is worth noting that compared with Random Forest performance on Night 13 (F1=0.738), the EEG+PPG model improved performance by 4.7 percentage points to 0.785.

PPG performance exhibited greater night-to-night variability, with F1-scores ranging from 0.436 (Night 3) to 0.523 (Night 13), representing a 20.0% relative variation. Notably, PPG achieved its highest performance (F1=0.523) on Night 13, the night with the poorest sleep quality and greatest EEG performance degradation. However, even PPG’s “best” night-specific performance (F1=0.523) remained considerably below EEG’s worst night (F1=0.779).

Table 6.9 presents comprehensive metrics for each model across all test nights, demonstrating that the sensor modality hierarchy observed in aggregate results (EEG » PPG) holds consistently across individual nights despite varying sleep quality and data conditions.

Table 6.9: Night-by-night performance metrics for all CNN models

Model	Accuracy	Macro F1	Cohen’s κ
<i>Night 3</i>			
EEG	0.816	0.798	0.728
PPG	0.441	0.436	0.190
EEG + PPG	0.833	0.819	0.757
<i>Night 13</i>			
EEG	0.768	0.779	0.679
PPG	0.533	0.523	0.369
EEG + PPG	0.772	0.785	0.687
<i>Night 17</i>			
EEG	0.846	0.840	0.780
PPG	0.484	0.483	0.331
EEG + PPG	0.825	0.822	0.753
<i>All Test Nights</i>			
EEG	0.807	0.805	0.727
PPG	0.486	0.481	0.304
EEG + PPG	0.809	0.810	0.733

Note: Models were tested on 2,836 epochs (all test nights): 1,050 epochs (Night 3), 1,027 epochs (Night 13), and 759 epochs (Night 17). Night 13 corresponds to documented poor sleep quality with onset insomnia and device displacement. Performance patterns show consistency with Chapter 5 Random Forest results, where Night 13 also exhibited poorest performance.

6.4 Discussion

6.4.1 CNN Performance Relative to Random Forest Classifier

6.4.1.1 EEG-Only CNN versus Random Forest

The EEG-only CNN achieved an F1-score of 0.805 and Cohen’s κ of 0.727 on the test set, representing meaningful improvements of 5.2 and 6.9 percentage points respectively over the Random Forest classifier trained on 22 hand-engineered EOG features (F1=0.753, κ =0.658). This demonstrates that spectrograms contain discriminative information beyond that captured by basic frequency band power features and statistical measures employed in Chapter 5.

The improvement, while meaningful, is modest rather than transformative. The CNN did not fundamentally alter classification performance but rather extracted incrementally more information from the same underlying signals. Several factors likely contribute to this advantage. Spectrograms provide complete spectral-temporal information across the 30-second epoch, capturing frequency dynamics and transitions that single summary statistics cannot represent. The CNN can learn to recognize patterns in how spectral content evolves over time, for example, the gradual emergence of delta power during NREM sleep or the spectral signatures of K-complexes in N2 sleep. Additionally, CNNs learn hierarchical feature representations, combining low-level spectral patterns into higher-level sleep stage signatures without requiring explicit engineering of these relationships.

However, the modest magnitude of improvement suggests that the hand-engineered features from Chapter 5 already captured much of the discriminative information present in the signals. The five EEG frequency band power features (delta, theta, alpha, beta, gamma) that dominated Chapter 5’s feature importance rankings represent a reasonable approximation of the spectral patterns that CNNs learn from spectrograms. The CNN’s advantage lies in finer temporal resolution and automatic feature interaction learning, but these benefits provided only an incremental rather than a substantial gain for this classification task.

6.4.1.2 PPG-Only CNN versus Random Forest

The PPG-only CNN achieved an F1-score of 0.481 and Cohen’s κ of 0.304, representing marginal improvements of 7.0 and 10.8 percentage points over Random Forest (F1=0.411, κ =0.196). Despite this improvement, PPG’s poor standalone performance reinforces the interpretation of having limited discriminative information content rather than inadequate feature engineering. Despite providing the CNN ac-

cess to complete spectral-temporal information through spectrograms, PPG barely exceeded majority-class baseline performance (45.9%) and remained far below EEG performance across all test nights.

The CNN’s ability to learn optimal feature representations directly from raw spectrograms did not overcome PPG’s fundamental limitation: cardiac dynamics captured at the finger provide minimal discriminative information for distinguishing sleep stages. This result aligns with the performance of commercial devices incorporating "sleep staging" capabilities, such as smartwatches (Robbins et al., 2024; Schyvens et al., 2025), which achieve similar performance levels using PPG-based approaches. This convergence demonstrates consistency between OSSMM hardware capabilities and commercially available and proven techniques, establishing that PPG’s poor performance reflects sensor limitations rather than OSSMM-specific issues.

6.4.1.3 Higher PPG Performance on Night 13

PPG achieved its highest individual-night test performance ($F1=0.523$) on Night 13, which also exhibited the poorest EEG performance and documented poor sleep quality with onset insomnia and device displacement. This inverse, paradoxical relationship warrants consideration of several possible explanations. First, PPG showed the greatest variability in performance scores, both during group-fold cross-validation (standard deviations of 0.047–0.093) and on the final test set. Such variability may account for the performance difference alone.

Another possibility is that this performance paradox reflects PPG’s sensitivity to movement artifacts that could serve as inadvertent arousal markers during fragmented sleep. However, it is important to note that even PPG’s "best" night-specific performance ($F1=0.523$) remained only marginally above the majority-class baseline (45.9%) and considerably below EEG’s worst night ($F1=0.779$), reinforcing that the sensor hierarchy persists under all tested conditions regardless of sleep quality or data characteristics.

6.4.2 Persistence of Sensor Hierarchy Across Methodologies

The results closely replicate Chapter 5’s ablation study, where EOG features alone achieved 97.8% of the full three-sensor system’s performance using Random Forest classification. The EEG-only CNN achieved 99.4% of the combined EEG+PPG system’s performance (F1-Score), demonstrating remarkable consistency across fundamentally different methodologies. EOG/EEG spectrograms or features alone achieved 99.4% and 93.8% of combined system performance (F1-score) for CNN

and RF respectively. This consistency across methodologies that learn features in fundamentally different ways (i.e., manually designed features versus learned convolutional filters) provides strong evidence that the sensor dominance reflects underlying physiological information content rather than limitations of the feature engineering approach.

The consistency with EEG/EOG achieving 93.8–99.4% of combined performance in both CNN and Random Forest approaches establishes that the OSSMM’s frontal electrode configuration captures the primary discriminative information for sleep stage classification. This finding has direct implications for system design: the added hardware complexity, cost, and failure points introduced by the PPG sensor provide negligible benefit for sleep staging when frontal EEG is reliably captured, regardless of whether classification employs conventional or deep learning approaches.

However, this conclusion applies specifically to sleep stage classification and to the single-participant dataset examined here. PPG may retain value for other research applications, such as autonomic nervous system assessment through heart rate variability, detection of sleep-disordered breathing, or investigations of cardio-respiratory coupling during sleep. The sensor’s presence in the current design does not impede performance and preserves research flexibility.

6.4.3 Reference Device Ceiling Question

In Chapter 5, a reference device ceiling imposed by the Somnify SM-100 reference device was discussed as a possibility. The CNN investigation strengthens arguments that this ceiling is indeed present.

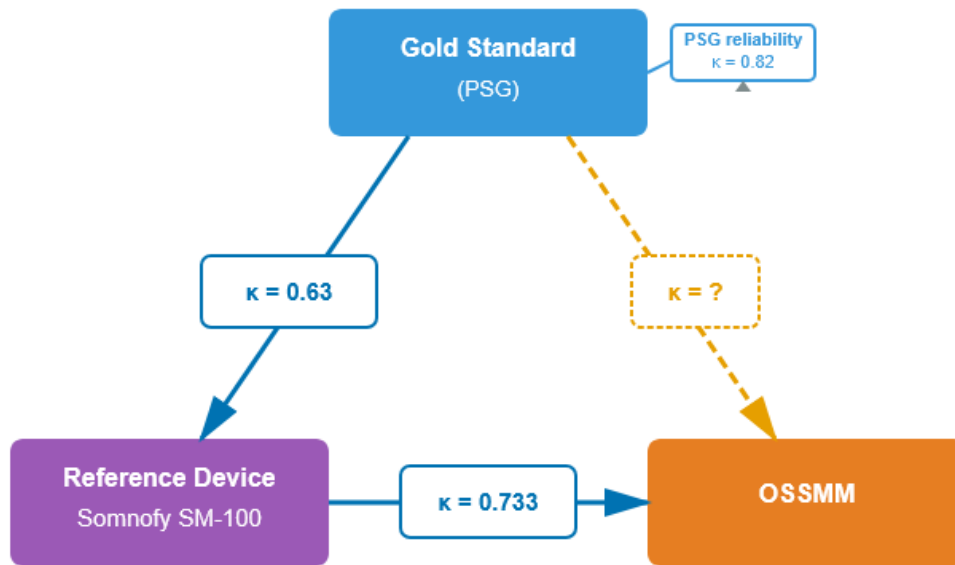


Figure 6.9: Conceptual diagram of the reference device ceiling problem.

Though both the EEG-only CNN and EEG+PPG CNN models improved performance relative to Random Forest models, and these improvements are meaningful, they cannot be called substantial. During training, both CNN models achieved rapid learning which plateaued, exiting training early at epochs 10 (EEG) and 9 (EEG+PPG). This suggests the architecture was able to discern discriminative patterns within the data quickly. By the second epoch, the EEG+PPG model’s validation F1-score had already reached 0.779, and likewise for the EEG-only model, 0.755.

During EEG hyperparameter search, despite testing Heavy regularization, all group-fold configurations produced average F1 scores between 0.768–0.779, with standard deviations ranging from 0.012–0.015. The tight clustering of performance across diverse regularization strategies, combined with rapid convergence during final training, suggests the models quickly saturated the available discriminative information in the Somnify-labeled data.

The CNN’s achievement of $\kappa=0.733$ exceeds Somnify’s agreement with PSG ($\kappa=0.63$) reported in its validation study (Toften et al., 2020). This creates an interpretive challenge: the OSSMM system either genuinely captures sleep architecture more accurately than the commercial reference device, or the CNN learned to predict Somnify’s classifications, including its systematic errors, with higher fidelity than Somnify agrees with PSG. The evidence suggests the former is more likely. The OSSMM system demonstrably registers EEG signals, which are not correlates of sleep stages but how sleep stages are principally defined in polysomnography. The confusion patterns observed in both CNN and Random Forest analyses mirror patterns reported in Somnify’s own PSG validation, suggesting our models learned Somnify’s classification tendencies. The ceiling most likely reflects the Somnify SM-100’s limitations rather than OSSMM capability. However, only gold-standard PSG validation can determine with certainty whether the ceiling reflects OSSMM capability or Somnify limitations.

This ambiguity does not diminish the proof-of-concept findings. As described in Chapter 5, the purpose of this evaluation was not to establish a performance metric of the device, but to demonstrate its technical feasibility as capable of sleep staging. The investigations thus far have established that: (1) OSSMM hardware reliably captures physiological signals across multiple nights, (2) these signals contain information sufficient for sleep stage classification at levels matching or exceeding commercial alternatives, (3) frontal electrodes dominate information content across analytical approaches, and (4) the system’s <€40 cost positions it as economically viable for research contexts. One benefit of Somnify over PSG-type validation at

this stage is its non-contact nature. There was no contention over electrode or device placements, and the full focus remained on device wearability, fidelity, and signal-acquisition.

6.4.4 Implications for CTPU Electrode Performance

Perhaps the most notable finding across both Chapters 5 and 6 is that conductive thermoplastic polyurethane (CTPU) electrodes without conductive gel successfully captured neurophysiologically relevant EEG signatures sufficient for automated sleep staging. These electrodes—sourced from commercial heart rate monitor chest straps and having higher impedance than research-grade Ag/AgCl electrodes—reliably recorded spectral content from delta through gamma (0.5–60 Hz, limited by hardware filtering) across 17 nights of testing.

This finding challenges conventional assumptions about electrode requirements for EEG recording. Clinical polysomnography and sleep research typically employ low-impedance electrodes with conductive gel or paste, carefully prepared skin surfaces, and multiple redundant channels to ensure signal quality. The OSSMM’s frontal electrode configuration, designed primarily for EOG, captured sufficient EEG information that five frequency band power features alone achieved 93.8% of full-system performance in Chapter 5, and CNN analysis of EOG/EEG spectrograms achieved 99.4% of dual-sensor performance in the current investigation.

The convergence of evidence across conventional and deep learning approaches is particularly compelling. Both Random Forest classifiers and CNNs learning identified frontal EEG as the dominant discriminative signal. This consistency across fundamentally different analytical approaches establishes that CTPU electrodes capture genuine neurophysiological signatures.

Several factors likely contribute to this success. First, the frontal electrode placement ((F7+Fp1)–(Fp2+F8) equivalent) captures strong signals due to proximity to neural generators. Sleep-related EEG patterns—particularly the high-amplitude slow waves of deep sleep and the distinctive spectral signatures of different sleep stages—may be less sensitive to electrode impedance than subtle waking EEG phenomena (e.g., event-related potentials) that demand pristine recordings. Second, the differential amplifier configuration with hardware filtering (0.5–40 Hz band-pass) provides common-mode noise rejection, compensating partially for higher electrode impedance. Third, sleep stage classification requires detection of robust state changes (wake vs sleep, NREM vs REM) rather than precise quantification of microvolt-level features, potentially tolerating signal quality inadequate for research requiring finer temporal or spectral resolution.

The practical implications are notable. Traditional EEG electrode preparation—skin abrasion, conductive gel application, careful impedance verification—requires trained technicians, creates participant discomfort, and complicates multi-night home use. CTPU electrodes integrated into a simple headband eliminate these barriers entirely. Users don the device and begin recording without special preparation. This accessibility enables longitudinal sleep monitoring that would be impractical with traditional EEG approaches.

However, important caveats apply. This investigation involved a single participant who designed and assembled the device. Generalization to diverse participants, different physiques (head shape, skull thickness, etc.) and skin conditions, and varying user expertise requires explicit validation. The CTPU electrodes’ performance likely varies across individuals in ways not characterized by single-participant testing. Additionally, we cannot determine whether signal quality would remain adequate for research questions requiring finer EEG resolution than sleep staging demands, such as quantifying sleep spindle characteristics and analyzing sleep microstructure.

6.4.5 Limitations

This investigation’s limitations largely mirror those discussed in Chapter 5, with some additional considerations specific to the deep learning approach.

The single participant design limits generalizability, though the sensor hierarchy findings which were consistent across both conventional and deep learning approaches appear robust. Validation against the Somnify SM-100 rather than gold-standard PSG creates the reference device ceiling problem discussed above. How OSSMM-CNN performance would change when validated against manually-scored PSG remains unknown.

To maintain a focused scope and manageable study design, this investigation focused on EEG and PPG. This followed from the Chapter 5 Random Forest results, where EEG features performed strongly while PPG performed poorly despite adding to the device’s cost and complexity. IMU was not explored in this chapter, as the three axes each from the gyroscope and accelerometer would have added complexity to the investigation without a design reason for doing so. Even though IMU did not show strong classification performance, it is an on-board component of the selected microcontroller and incurs no additional hardware design complexity.

The test set comprised three nights (2,836 epochs, representing 18.55% of the 15-night dataset). While this is proportionally appropriate for the available data, the small absolute size constrains statistical power for detecting small performance differences.

The CNN architectures employed represents a straightforward adaptation of the established VGG-inspired designs rather than extensive architecture search. Other architectures (ResNet, attention mechanisms, recurrent layers for temporal modeling) might extract additional information, though given the modest improvement of CNNs over conventional methods and evidence of reference device ceiling effects, substantial gains from architectural optimization appear unlikely.

Finally, the investigation focused exclusively on four-class sleep staging (Wake, Light Sleep, Deep Sleep, REM). Clinical polysomnography distinguishes additional states (N1, N2, N3, REM and Wake separately). Whether the OSSMM system and CNN approaches could support finer-grained classification remains unexplored, as the Somnify reference device provides only four-class outputs.

6.4.6 Conclusion

This investigation conclusively answered both research objectives. CNNs applied to spectrograms improved classification performance over conventional methods (Cohen’s κ of 0.733 vs 0.675), demonstrating that spectrograms contain discriminative information beyond hand-engineered features. However, the improvement was incremental rather than transformative. Rapid convergence on training data, and tight performance clustering across regularization strategies, suggests the models quickly saturated the discriminative information available in Somnify-labeled data.

More significantly, the sensor hierarchy replicated almost exactly across methodologies: EEG alone achieved 99.4% of the combined system with CNNs and 93.8% of the full three-sensor system with RF, while PPG contributed negligibly in both cases. This consistency across fundamentally different analytical approaches, manually designed features versus learned representations, establishes that the hierarchy reflects genuine physiological information content rather than feature engineering artifacts. Importantly, the convergence supports the Chapter 5 finding that inexpensive CTPU electrodes without conductive gel reliably capture neurophysiologically relevant EEG signatures sufficient for automated sleep staging.

From a broader research perspective, Chapter 5 and the investigation here, sought to establish the technical viability of the OSSMM system. These investigations have revealed that the hardware reliably captures signals suitable for sleep stage classification using multiple analytical paradigms, performance matches or exceeds more expensive alternatives, and the frontal electrode configuration dominates information content regardless of methodology. Critical validation steps remain for future work; multi-participant studies and direct PSG validation, but the convergence of evidence across conventional and deep learning approaches demonstrates

that accessible, open-source sleep monitoring capable of research-quality classification is achievable with appropriate hardware design and signal processing strategies. The proof-of-concept for technical feasibility is established, and the path to broader validation is clear.

Future Work and Conclusion

“Too many men work on parts of things. Doing a job to completeness satisfies a man.”

— *Richard Proenneke, Alone in the Wilderness, June 25th, 1968*

“Hardware is hard.”

— *Unknown*

Abstract

This chapter concludes the thesis, presenting a summary of contributions, three concrete projects for future work, and clarifying its scope and timeline.

7.1 Overview

This thesis investigated the abandonment of REM-R as a potentially therapeutic sleep intervention for depression, and advanced a solution for the prohibitive cost barriers that restrict access to sleep research. These were achieved through a targeted literature review, hardware development, and technical evaluation.

The REM sleep reduction (REM-R) literature deep dive revealed that abandoning this research area lacks justification. All three controlled human studies reported improvement in depressive symptoms (DS) despite methodological limitations. The convergence between REM suppressing properties of traditional monoaminergic antidepressants and behavioral REM-R findings suggests a shared mechanistic pathway worthy of renewed investigation.

The primary contribution of this thesis is the Open-Source Sleep Monitor and Modulator (OSSMM) system, which demonstrates the technical and economic feasibility of a research-grade sleep monitoring platform at a component cost under €40. A technical validation demonstrated that inexpensive conductive thermoplastic polyurethane (CTPU) electrodes, sourced from consumer heart rate monitor chest straps (HRMCS) likely captures frontal EEG signals with adequate reliability. Conventional machine learning classifiers and a convolutional neural network (CNN) classifier successfully discriminated the OSSMM sensor data using labels from a silver-standard sleep monitoring device (Somnify SM-100), confirming that OSSMM captured data is suitable for sleep staging.

7.2 Contributions

This thesis delivered contributions for each of its three objectives. The literature review reviewed the evidence base for REM sleep reduction as a treatment for depression; the system development produced the OSSMM platform itself; and the system evaluation established the platform’s technical feasibility. Each is considered in turn below:

7.2.1 Literature Review

The deep dive identified numerous methodological confounds in existing REM-R studies that warrant caution in interpretation. These lessons learned represent the investigation’s primary contribution: a roadmap to prevent recurrence of these errors and strengthen future research design. While the list is extensive, the most important lessons are distilled into the research roadmap presented in Section 7.3.

7.2.2 System Development

The OSSMM platform’s primary contribution lies in opening the door to sleep research, demonstrating that research-grade sleep monitoring does not require expensive or complex hardware. By reducing the barrier to entry from thousands of euros per unit to under €40 in component costs, this work makes multi-participant, longitudinal sleep studies economically and logistically feasible for research groups that previously could not afford quality monitoring equipment. As illustrated in Section 1.4.2, the high cost of portable PSG units restricts research groups in the number of participants and the duration of observation for studies. A four-week study with $N = 30$ participants, and 4 PSG units would require at least seven months to complete, assuming constant use of devices. This would also likely represent an equipment cost of at least €20,000 (assuming each PSG unit is €5,000, not accounting for consumables). Even with the purchase of a suitable Android device, plus the component cost of an OSSMM headband (€100 + €40), the same study design would have an equipment cost of €560 (a reduction of 97.2%). Alternatively, all $N = 30$ participants could be run concurrently for an equipment cost of just €4,200. This transformation in cost and time scale makes previously impractical study designs viable. Universities and research groups without large equipment budgets would be able to conduct large-scale, longitudinal investigations previously restricted to well-funded sleep laboratories.

The complete transparency of hardware designs, software code, and assembly instructions addresses reproducibility concerns that plague closed, proprietary systems. All project assets have been freely available in a GitHub repository since May 2025, enabling researchers worldwide to replicate, validate, adapt, and improve upon this work (<https://jvgiordano.github.io/OSSMM/>). The OSSMM system is not presented as a finished product, but as a foundational platform for continued innovation.

7.2.3 System Evaluation

The multi-night technical evaluation is the third contribution of this thesis and the first assessment of whether the platform meets its design intent. Across a 17-night single-participant dataset, the OSSMM system demonstrated the wearability, hardware reliability, and signal fidelity targeted by the system evaluation objective, operating reliably across diverse sleep-quality levels and physiological states. Critically, the captured signals carried sufficient discriminant information for sleep staging: conventional classifiers (XGBoost and Random Forest) and a lightweight CNN successfully discriminated the OSSMM sensor data against labels from a silver-

standard reference device (Somnify SM-100), confirming that the captured data are suitable for automated sleep staging.

The successful use of CTPU electrodes from commercial fitness chest straps for capturing EEG signals represents an unexpected finding with greater implications. The demonstration that dry, flexible electrodes with substantially higher surface resistivity than standard Ag/AgCl electrodes can capture neurophysiologically relevant frequency bands across delta through gamma for sleep staging is significant. While validation across diverse participants remains necessary to establish generalizability, this finding suggests that accessible EEG-based monitoring may require less stringent electrode specifications than traditionally employed.

7.3 Future Directions

The future work arising from this thesis is best understood not as a list of isolated tasks but as three interlocking research programmes, each of a scale suitable for a doctoral project. The first would rigorously validate and generalise the OSSMM platform; the second would extend it from a passive monitor into a closed-loop interventional system; and the third would deploy it in the clinical investigation of REM sleep reduction for depression that motivated this work. Together they describe a path from the proof-of-concept established here to the targeted study of sleep modulation as a therapy.

7.3.1 Project 1 — Validation and Generalisation of the OSSMM Platform

The most immediate programme concerns establishing the platform’s true capabilities and the conditions under which they generalise.

The immediate priority is validation against gold-standard PSG to establish whether the OSSMM platform can exceed the performance of substantially more expensive alternatives when provided with ground-truth reference labels. The possible validation ceiling imposed by the Somnify SM-100 reference device ($\kappa = 0.63$ against PSG) indicates the observed performance likely represents a lower bound rather than the system’s true capabilities, albeit in $N = 1$. PSG validation would also enable assessment of 5 stage sleep staging (N1/N2/N3/REM/Wake distinction). Multi-participant studies across diverse populations are essential to establishing generalizability of the findings, particularly regarding CTPU electrode performance across different scalp and skin types, and sleep pathologies. However, detailed characterization of CTPU electrode impedance characteristics, including frequency-dependent impedance and contact pressure effects, was beyond the scope of this

proof-of-concept validation. Such characterization would be valuable for future device optimization and establishing quality control criteria for HRMCS selection to ensure replicability across device builds.

Extending evaluation to include IMU data through multi-modal deep learning architectures would complete the comparison with Chapter 5’s three-sensor ablation study. While Chapter 5 established that movement features contributed minimally, CNN approaches might extract different information from IMU spectrograms of movement dynamics.

Hardware refinements should focus on device stability to minimize potential displacement during sleep. The sensor ablation findings suggest that future designs could be simplified by prioritizing EOG and EEG capture over the need for a pulse sensor, likely achieving equivalent performance at lower cost, improved ergonomics, and design simplification. However, future work should consider whether inclusion of this sensor is necessary in the event a smaller headband can be designed without it, as a smaller design may lead to less disturbance of the CTPU electrodes during sleep, which evidence shows captures the most critical information.

The open-source nature of the OSSMM platform presents both opportunities and challenges for reproducibility. While complete hardware designs, software, and analysis code are publicly available, variations in component sourcing (specifically HRMCS and their CTPU electrodes from different manufacturers), assembly quality, and 3D print parameters may affect system performance in ways not fully characterized by this single-unit study. Establishing the sensitivity of system performance to these manufacturing and assembly variables, and developing quality control metrics to ensure consistent device performance across replications, represents important future work for the platform’s maturation as a research tool. Devices should also be constructed by researchers other than the designer. Such studies should include researchers assembling devices, evaluating whether the open-source design achieves its accessibility goals.

7.3.2 Project 2 — Closing the Loop: Real-Time Staging and Selective Modulation

Realising the system’s modulation potential defines a second programme, transforming OSSMM from a passive monitor into an interventional device.

Software development should prioritize improved security features, like AES encryption, and integration of automated sleep staging algorithms directly into the Android app would enable real-time classification for closed-loop interventions. The next step

would be implementation of selective sleep stage modulation capabilities for interventional research. Experiments in the area of lucid dreaming research would be an ideal stepping stone before more serious, clinical investigations are considered. The vibration motor hardware already present in the system provides a foundation for this work. Nonetheless, under the lens of REM-R studies, development of automated REM detection with appropriate specificity thresholds and protective logic to prevent false positive targeting would be a critical requirement for future REM-R investigations.

Testing the utility of the current modulation stimulus, a vibration motor, during sleep remains a priority. Improved electronic case design may permit additional stimulus modalities (e.g., lighting), and better vibration transmission. It will also be of interest if a method can be developed where a stimulus may be applied to the sleeping participant, and affect REM sleep without causing arousal, similar to approaches used in SWA reduction research.

7.3.3 Project 3 — Clinical Investigation of REM-R for Depression

The culminating programme is the controlled clinical study that this thesis was ultimately built to enable, conducted according to the methodological lessons drawn from the literature review.

Controlled studies are necessary, and where possible, cross-over designs should be employed. This eliminates a key bias undermining psychiatric research: the heterogeneity in diagnosis labels. Any study should include a sufficient sample size, and sufficient measurement periods. Baseline measurements are required to determine comparison effects between interventions and rebounds (Vogel et al., 1975, 1977, 1980). Due to the First Night Effect, where participants in sleep studies show altered sleep measures due to the novel environment and stimuli, and the potential for spontaneous remission, participants should undergo at least one week of baseline monitoring (Vogel et al., 1975). Similarly, post-intervention monitoring should also continue for at least one week, if not longer, in order to understand the potential longer term change in sleep dynamics and mood caused by interventions (Grözinger et al., 1997).

Intervention mechanisms must be considered with diligence. Automated systems, such as future OSSMM iterations, should be strongly considered for the longitudinal study required for such an undertaking. This will reduce both the cost and labor burden of such an experiment, and standardize the interpretation of sleep states through removal of human scorer variability. In such an experiment, great care must

be taken to ensure target specificity. Classification algorithms and control logic must not permit the targeting of other states (Grözing et al., 2002). Delivery schedule of the intervention must be considered. For example, if a replication of Vogel et al. (1975) is intended, then NREM interruptions should not be evenly separated among NREM sleep states. Future studies could employ more granular branch interventions: REM-R, NREM-R for N2, and NREM-R for N3. The purpose of selective targeting of both state and timing is to avoid confounds caused by error in interfering with the wrong state, and accounting for collateral changes which might happen to overall sleep architecture, like changes to the NREM-REM cycle length. Importantly, careful control must be done for use of pharmacological agents, and other life factors, to avoid confounds (Cartwright et al., 2003; Grözing et al., 2002).

Regarding analyses, all intended measures should be pre-registered. This will permit clear understanding of what aspects the study is testing and which effects may be observed with further data investigation. Furthermore, given the improved understanding of depression case-control, and potential sleep architecture biomarkers for responsivity to REM-R, any future study should attempt to undertake a prediction based consideration with early identification. There are two potential methods which may both be utilized. First, if an investigation is intended to cause an intervention over a given duration (e.g. four weeks), then early measurement compared with baseline (e.g., one week after baseline completion), should be used to predict end study responders (i.e., remitters, depending on intervention intensity and duration) and non-responders. This will help to establish if early signs to treatment can be used for future sub-group identification, and potentially have implications for clinical use. Second, machine learning methods should be used to determine if responders and non-responders can be differentiated using repeatable markers (sleep measures, cognitive task performance, etc.). In this way, a portion of the participants should be randomly separated as a test set, and the rest training.

Finally, it may be of great theoretical and practical use to consider the tonic and phasic events of REM separately as no REM-R study for depression has yet to do.

The long-term goal remains clinical investigation into sleep modulation research to shed light on the unclear findings of REM-R and improvement of depressive symptoms. The methodological lessons identified in the literature review provide a road map for improved experimental design. The OSSMM platform provides the technical foundation necessary to overcome the cost and labor barriers that have helped prevent systematic investigation of sleep modulation interventions for over 25 years.

7.4 Context and Scope

While this thesis achieved its core objectives, important limitations must be acknowledged, particularly when contextualizing the work within the constraints faced during its execution. The technical validation employed a single-participant design (N=1), with the researcher serving as participant across all nights. This approach was appropriate for establishing basic system viability but limits claims of generalizability. The findings reflect that the OSSMM system can work rather than establishing population-level performance metrics.

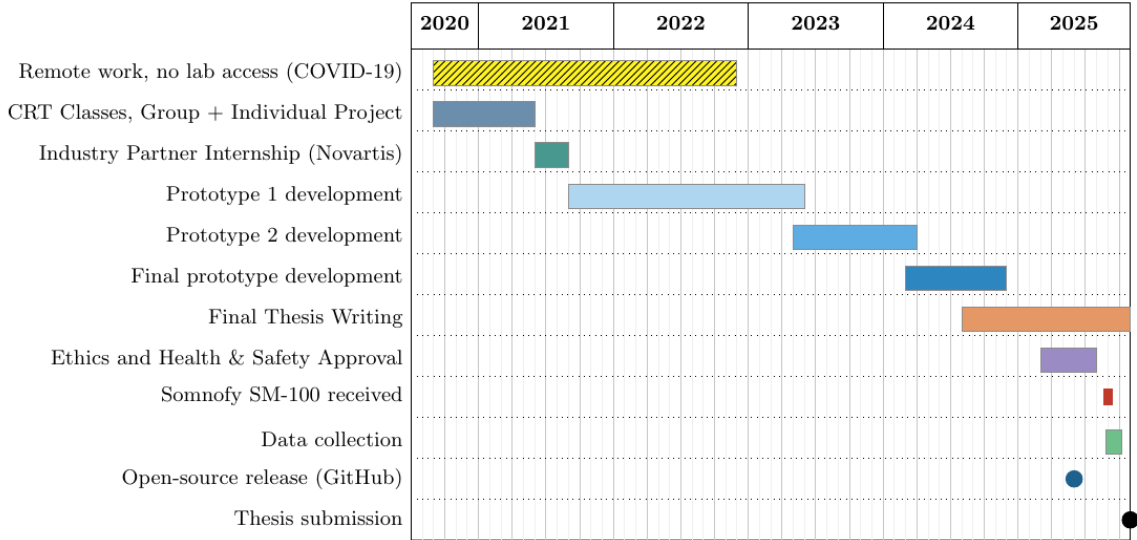


Figure 7.1: Gantt chart of PhD timeline

This constraint resulted from the timeline pressures and resource limitations rather than methodological choice. As shown in Figure 7.1, the extensive hardware development consumed the majority of the available time and resources. The first two years of the PhD were completed remotely due to the Covid-19 pandemic and lockdowns occurring in Ireland, preventing access to laboratory equipment, such as the 3D printer. Second, the entire hardware and consumables budget for this thesis was €1,000. It was only because of unexpected grant funding during the extension of the thesis (in 2025), that funding was found to purchase a Somnify with a six month subscription to serve as a reference device for technical validation. However, to the author’s knowledge, this amount would not have been sufficient for a Type II polysomnography system.

The single-participant constraint does not fundamentally undermine the technical contributions of this thesis. The 17-night dataset captured substantial variability across different sleep quality levels, data collection conditions, and physiological states. The convergence of findings across conventional and deep learning ap-

proaches, the clear sensor hierarchy replicated across methodologies, and the operational reliability demonstrated across diverse conditions all support technical feasibility. Multi-participant validation remains necessary to establish population-level performance, but the proof-of-concept establishes that the fundamental approach is sound.

Validation against the Somnofy SM-100 rather than gold-standard polysomnography introduces a validation ceiling problem. The OSSMM achieved $\kappa = 0.733$ against Somnofy, which itself achieved $\kappa = 0.63$ against PSG in its validation study. This raises the question: Does the OSSMM genuinely capture sleep architecture more accurately than Somnofy, or did the classifiers learn to predict Somnofy’s classifications including its systematic errors? The evidence strongly suggests the former is more likely because the OSSMM demonstrably registers EEG signals that define sleep stages rather than merely correlating with them, but only direct PSG validation can resolve this ambiguity.

7.5 Conclusion

This ambitious PhD project demonstrated that democratizing sleep research through open-source, low-cost hardware is technically feasible. The OSSMM platform establishes proof-of-concept for research-grade sleep monitoring at a fraction of traditional costs, while the REM-R literature review provides a road map for renewed investigation of this potentially therapeutic intervention.

The convergence of findings across multiple analytical approaches, the operational reliability demonstrated over extended testing, and the unexpected effectiveness of inexpensive CTPU electrodes for EEG capture provide strong evidence that the fundamental approach is sound. While multi-participant validation and direct PSG comparison remain necessary to establish population-level performance, the technical foundation now exists.

The complete open-source release of all hardware designs, software, and assembly instructions ensures that this development can serve as a foundation for continued innovation. The OSSMM system stands as proof that purpose-built sleep monitoring devices need not be expensive or complex to be effective, opening new avenues for large-scale sleep research. The barriers that have restricted quality sleep research to wealthy institutions for decades are no longer insurmountable. Whether this platform enables renewed investigation of sleep modulation interventions for depression, supports novel sleep research in resource-limited settings, or inspires further developments in neurophysiological monitoring, the tools and road map now

exist for any motivated researcher to use. The goal of democratizing sleep research has taken a concrete step forward.

Appendix — REM Sleep Reduction and Depression: A Deep Dive Review

Appendix A provides supplementary information for Chapter 2

A.1 Normality and Homogeneity Tests for Parametric Testing for Endogenous Patient Data

A.1.1 Normality

Shapiro-Wilk

All values per group and time point (baseline, crossover) are unique.

Shapiro-Wilk Normality Tests

Variable	W	p-value
control_baseline	0.90661	0.0877
control_crossover	0.96279	0.6844
experimental_baseline	0.9523	0.4938
experimental_crossover	0.96657	0.756

D'Agostino-Pearson

For the Skewness test: Alternative hypothesis — data have a skewness.

Variable	skew	z	p-value
control_baseline	0.23589	0.49810	0.6184
control_crossover	-0.19682	-0.41620	0.6773
experimental_baseline	-0.26202	-0.55267	0.5805
experimental_crossover	-0.24657	-0.52043	0.6028

A.1.2 Homogeneity of Variance

Levene's Test

	Df	F value	Pr(>F)
group	1	0.3893	0.5371
	32		

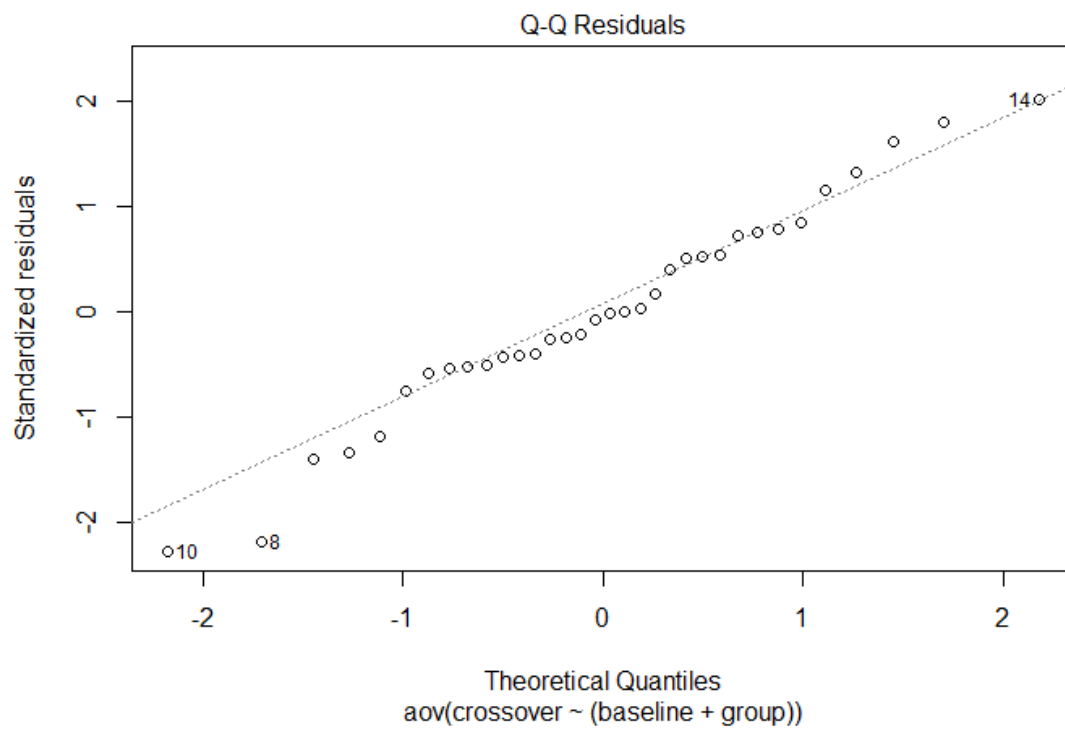


Figure A.1: Normal Q-Q plot of standardized residuals for the ANCOVA predicting crossover scores from baseline and group. Points fall close to the reference line with minor tail deviations, indicating approximate normality of residuals.

A.2 Additional Methodological Considerations for Cartwright et al. 2003

Further considerations for Cartwright et al. 2003 experiment are presented here.

Diurnal Variation Score (HAMD-21) versus POMS-D. Although the DV scores (derived from HAMD-21) suggest that remitters were “evening-worse,” the twice daily POMS-D data contradict this. According to POMS-D, non-remitters not only experienced a greater mood worsening from morning to evening, they always experienced this worsening. Remitters on the other hand either demonstrated no worsening in Month 1, very mild worsening in Month 2, or improvement by Month 5. It is possible this inconsistency is due to differences between these measures. HAMD-21 captures more trait-like features and POMS-D captures more momentary states. However, the way HAMD-21 items 18A and 18B were combined to form a single DV score conflates direction and severity, further complicating the interpretation of paradoxical result.

Diurnal Variation Score. The authors claim that at screening, participants who remit are most depressed at night, while those who will not remit are most depressed in the morning. The diurnal variation score (DV) was created by collapsing HAMD-21 items 18A (direction) and 18B (severity) onto a single 1–4 scale, described as “1 = mildly worse in the morning” to “4 = severely worse in the evening.” This creates a “*Heads I win, Tails you lose*” scenario. This single axis conflates direction (morning- vs evening-worse) with severity (mild vs severe), such that higher scores simultaneously indicate evening worsening and greater intensity. As a result, observed associations between higher DV and remission may reflect direction, severity, or their combination, and cannot be disentangled with this metric.

Seasonality. The study spanned five months during which seasonality changes could affect mood and sleep. Start dates or adjustments for time-of-year effects are not mentioned in the methodology. Including month as a covariate in testing Group X Month interactions would help address this potential confound.

A.2.1 A Note on Heterogeneity and the Specificity Question

Before proceeding to detailed analysis, a conceptual clarification is necessary. The question “Is REM-R specific?” may be improperly framed. Depression, as discussed in Chapter 1, is not a singular condition but heterogeneous, likely comprising multiple potential pathophysiologies. The Diagnostic and Statistical Manual of Mental Disorders, now in its fifth edition (DSM-V), permits 945 unique symptom profiles to

qualify for Major Depressive Disorder. Explicitly, two patients meeting diagnostic criteria may share no overlapping symptoms beyond anhedonia or depressed mood ([American Psychiatric Association, 2013](#); [Fried and Nesse, 2015a](#)).

Therefore, the appropriate question is: Does a subset of depressed individuals respond specifically to REM-R? The evidence reviewed below consistently reveals responder/non-responder patterns rather than uniform effects. This heterogeneity suggests that biological subtypes, identifiable through sleep architecture biomarkers, may be the appropriate level of analysis, not diagnostic categories based on symptomology.

This framing has important implications for interpreting the studies below. Mixed results at the group level may mask strong effects in responder subsets. Apparent “non-specificity” when comparing REM-R to NREM-I may reflect different individuals responding to different mechanisms. The challenge is not proving that REM-R works for all depression, but identifying which patients it works for, and why.

Before examining these studies in detail, the heterogeneity discussed in Chapter 1 merits specific consideration here. Given that depression comprises multiple potential pathophysiologies rather than a singular condition, the question “Is REM-R specific?” may be improperly framed. The more appropriate question is whether distinct responder subsets exist — whether some depressed individuals respond specifically to REM-R while others do not. This framing, focusing on which patients respond and why rather than whether all patients respond uniformly, will guide the analysis below.

It has been suggested that REM-R cannot share mechanisms with TSD given their different timescales of response (weeks versus single night) ([Wirz-Justice and Van den Hoofdakker, 1999](#)). However, this objection warrants scrutiny. First, the same timescale difference applies equally to NREM-R, which also requires weeks to produce sustained effects distinct from TSD’s rapid but transient impact. If differing timescales invalidate REM-R as sharing mechanisms with TSD, they should equally invalidate NREM-R—yet NREM-R findings were accepted as evidence against REM-R specificity. This inconsistency suggests selective application of skepticism. Second, while the mechanisms underlying these timescale differences remain unclear, different intervention intensities or dosing schedules of the same underlying process could plausibly produce distinct temporal dynamics. Chronic pharmacological REM suppression over weeks shares REM-R’s gradual timescale while differing fundamentally from TSD’s immediate reversal upon sleep. Third, sustainability differs critically: TSD effects vanish immediately upon sleeping or napping, whereas REM-R effects persist beyond individual nights. The relevant comparison may therefore

concern sustainable versus acute interventions rather than shared versus distinct mechanisms. The timescale question merits investigation, but does not provide grounds for dismissing REM-R's therapeutic potential.

Appendix — Design Evolution of the OSSMM System

Appendix B provides supplementary information for Chapter 3

B.1 Foundational Prototyping Detail

This section provides extended detail on the foundational prototyping work introduced in Section 3.4.1, specifically the evolution of the companion Android application through MIT App Inventor, Kivy, and Dart/Flutter.

B.1.1 MIT App Inventor and the Foundational Smartphone Application

For the foundational experiment, MIT App Inventor was used to create a basic smartphone app which could receive and display values transmitted over a BC connection. MIT App Inventor is a web-based, click and drag “block based” utility for building smartphone applications, and requires very limited programming experience (Patton et al., 2019). At this time in the project, the author did not have any Android app programming experience which is the principal reason for testing MIT App Inventor.

An anticipated shortcoming of MIT App Inventor for app development was confirmed during the foundation experiment. Although MIT App Inventor allows for relatively quick construction of simple programs, block programming ultimately struggles and becomes cumbersome with complex functionality. As a rule, it can be useful for rapid prototyping and introduction to programming logic, but lacks the flexibility and speed necessary for production level development. It is noted

that other immediate hurdles with MIT App Inventor hastened departure from this utility including difficulties with BC connectivity and data storage.

B.1.2 Kivy Detour

Kivy was chosen next as it seemed a promising approach for Android development, without requiring learning a new language as the coding is in Python (which the author is familiar with). Kivy is an open-source Python framework for mobile app development. Ultimately, Kivy was abandoned despite several weeks of work, as no working API for Bluetooth could be found in 2021–2022. Despite apparent evidence that such an API did exist for Bluetooth, personal correspondence with the publisher confirmed the example in question used WiFi and not BC. Therefore, it was decided that a new programming framework was needed.

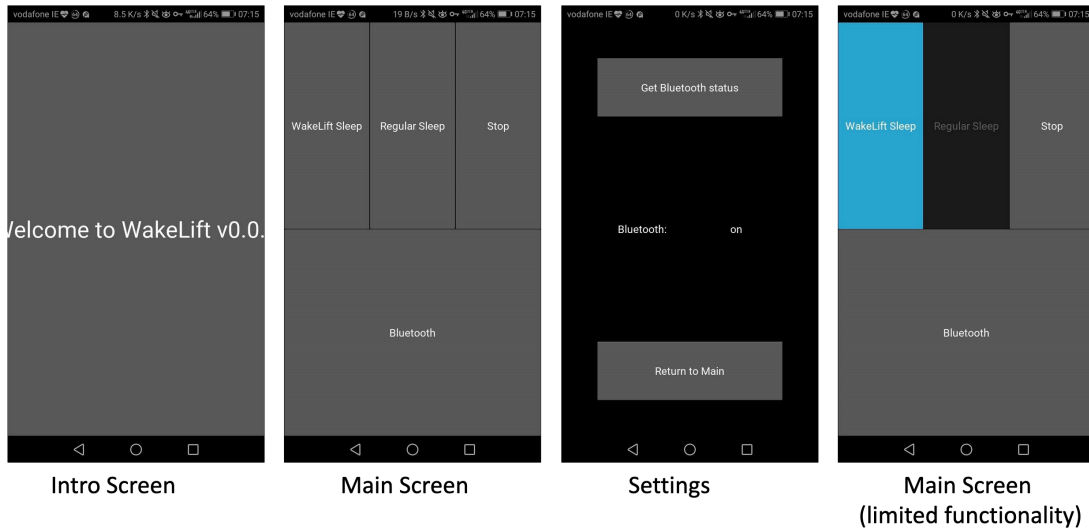


Figure B.1: Various screens in the Kivy Android app, the second app iteration built for the OSSMM project.

B.1.3 Transition to Dart and Flutter

Upon various considerations, app development began in Dart and Flutter. Dart is a programming language typically used for mobile and web development, and Flutter is an open-source software development kit which uses Dart.

Figure B.2 shows several of the screens from the first version of the app built with Flutter, specifically for Android 9. A repeat of the foundation experiment was conducted, and a result is shown in the far right screen. Here a time plot of the sampled voltage as rounded to the nearest whole number is displayed. This early version of the app had limited capabilities but allowed users to enable/disable Bluetooth on their device, search for nearby Bluetooth devices, connect and receive data from a Bluetooth device, and display the connected device's name.

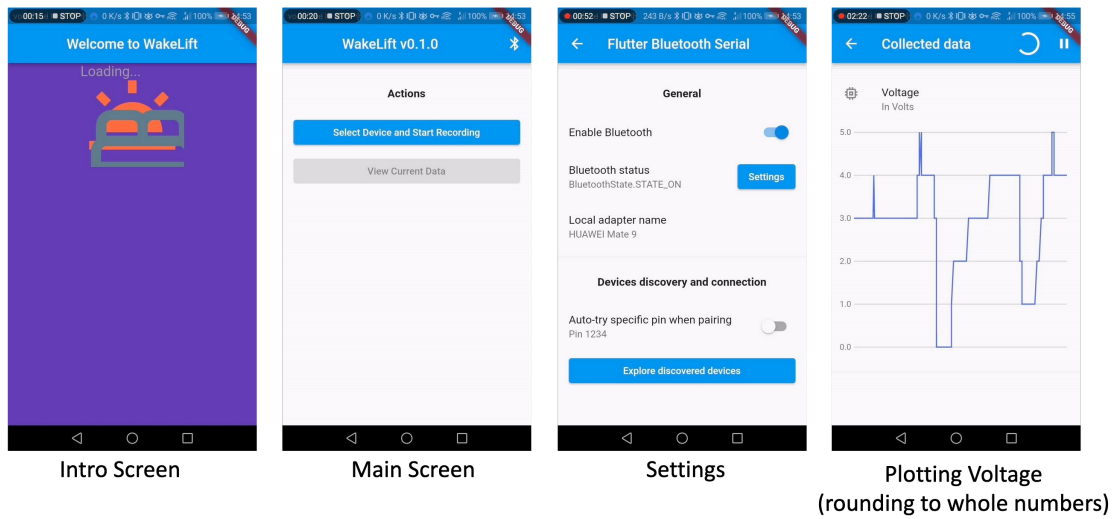


Figure B.2: Various screens in the early Flutter Android app, built for Android 9 and using Bluetooth Classic.

B.2 Arduino Nano 33 IoT — Full Exploration of the Bluetooth Classic Limitation

This section provides extended detail on the Bluetooth Classic (BC) limitation of the Arduino Nano 33 IoT introduced in Section 3.4.4.

While Arduino’s documentation indicates the Nano 33 IoT is capable of BC, in practice the BC capabilities of the board are severely limited as to make it impractical for use. Furthermore, the BC problems of the Nano 33 IoT were so poorly understood at the time that when Arduino Support finally responded to the author’s request for assistance, they sent links to the author’s own workaround and experimental results which had been posted on the Arduino forums.

The reason BC does not realistically work in the Nano 33 IoT is that the data rate is severely restricted, and the functioning workaround for this restricts access to analog inputs. Explicitly, the Nano 33 IoT board has two processing units: the main SAMD21G, and the NINA-W102 (ESP32), which is responsible for wireless communication. For reasons that still remain unclear, the serial communication between the SAMD21G and the NINA-W102 is bottlenecked. Any information that is first processed by the SAMD21G then sent to the NINA-W102 for BC transmission will not exceed 112 Bytes/second, though data rates much higher than this are normally supported by BC supported modules.

One workaround requires hacking the firmware on the board making the NINA-W102, and by extension its incorporated ESP32, becomes the main board processor rather than just the communication chip. This means the first and only processing

on the board is completed by the ESP32. This alleviates the bottlenecking with BC and the higher data rates (e.g., 3,125 Bytes/second) can indeed be transmitted without problem. However, as shown in Figure B.3, the NINA-W102 module can only provide a single analog-to-digital converter (ADC) input (PIN A6) (Guadalupi, 2022).

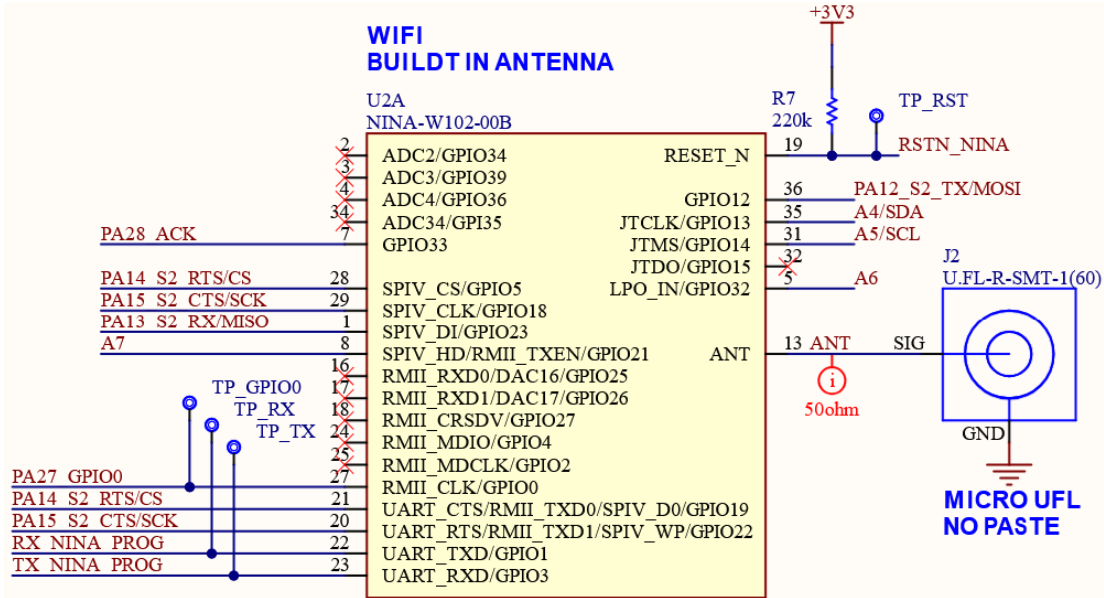


Figure B.3: NINA-W102 Schematic on Arduino Nano 33 IoT. Sourced from (Guadalupi, 2022).

To confirm this was a hardware limitation, as indicated by the schematic, rather than a software implementation issue, the author conducted a verification test using a Lolin Lite ESP32 development board. Since both the NINA-W102 and Lolin Lite ESP32 utilize the ESP32 microcontroller as their main processing unit, the code was directly transferable between MCUs. This verification test confirmed that the intended code functioned correctly and achieved sufficient Bluetooth performance with multiple ADC inputs, demonstrating that the single ADC pin limitation was indeed a hardware constraint of the NINA-W102 module on the Nano 33 IoT platform, not a software problem.

B.3 Bluetooth Classic and Bluetooth Low Energy — Technical Description

This section provides the conceptual technical background for BC and BLE referenced in Section 3.4.4.4. The specific OSSMM BLE implementation, including the GATT service definition, packet structure, and Maximum Transmission Unit settings, is described in Chapter 4.

B.3.1 Conceptual Models

A simple model for BC is to consider it like a telephone call, with the line always active and carrying information. The downside is this constant communication link similarly continually consumes power.

In contrast, BLE communication can be modeled as a public bulletin board, like that shown in Figure B.4. On this bulletin board there are “posters” called *services*, and under each service is at least one “Post-it note” or *characteristic*. Technically, this protocol is referred to as Generic ATtribute Profile, or GATT. In this system, there are two types of devices: peripherals and central devices. Each communication is always between a peripheral and a central device. The peripheral device “hosts” the bulletin board by initiating and defining the available services and characteristics that structure the communication. Both the central device and the peripheral device communicate by updating the values on the “Post-it note,” that is the *characteristic value*. These updates happen in determined intervals, not continuously. Because of this, BLE communication is more energy efficient because the line of communication is not continuous, and the energy consuming hardware can enter low-power modes (i.e., “sleep”) between transmission updates.

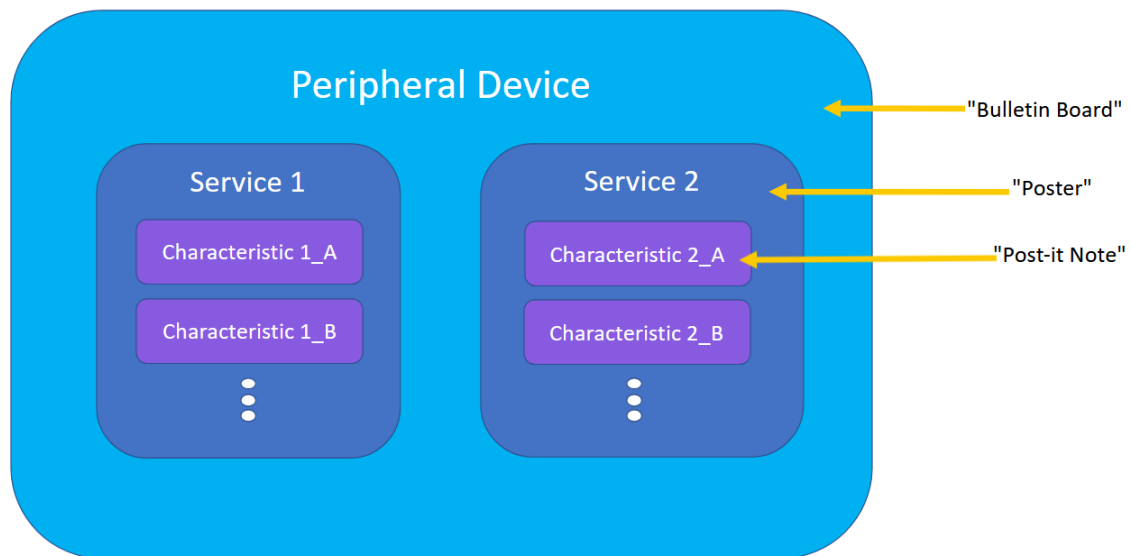


Figure B.4: BLE Bulletin Board Model illustrating the GATT structure of services and characteristics.

B.3.2 Interoperability, Hardware, and Libraries

Though BC and BLE use the same 2.4 GHz ISM band, their modes of operation are significantly different. They are not interchangeable on any level. From a hardware perspective, BC and BLE transceivers are different, and from a software perspective the necessary code and respective libraries to use them, both on the microcontroller

and in the smartphone application, are also different. Although this technology switch in the OSSMM project was not straightforward, the future impact on energy savings was essential for the project’s success.

Regarding the smartphone application, the Flutter application switched to the `Flutter_Blue_Plus` package for handling BLE ([Weinberger and contributors, 2024](#)). The investigated microcontrollers had their own BLE-specific libraries including `ArduinoBLE`, `BLEDevice` (for ESP32 architecture), `NimBLEDevice` (for ESP32 architecture), and `BlueFruit` ([Adafruit Industries, 2017](#); [Arduino SA, 2019](#); [Espressif Systems, 2019](#); [h2zero and contributors, 2020](#)).

B.3.3 Achieving Sufficient BLE Throughput

Achieving sufficient throughput to meet the design goal of each sensor sampling occurring at 250 Hz was not straightforward with BLE. Unlike BC with its continual stream of data, BLE updates occur at predetermined intervals. This interval length depends on the device and the supported Bluetooth version. Default intervals are 30 ms, but devices supporting Bluetooth versions after 4.0 may be able to achieve intervals as low as 7.5 ms ([Bluetooth SIG, 2025b](#)). However, even at the fastest interval BLE permits, if each sensor sampling was updated at this rate, the highest frequency for data collection would be just 133.3 Hz, well shy of the 250 Hz design target:

$$f = \frac{1}{T} \quad \Rightarrow \quad 133.3 \text{ Hz} = \frac{1}{7.5 \text{ ms}} \quad (\text{B.1})$$

Fortunately, this bottleneck can be surmounted because BLE permits changing the size of the “Post-it notes” in the “Bulletin board” model. The larger the note, the more information a characteristic can contain. This information limit is referred to as the Maximum Transmission Unit (MTU). By default the MTU is 23 bytes, with 3 bytes always serving as overhead regardless of MTU size ([Bluetooth SIG, 2025a](#)).

Adjusting the MTU is important for increasing throughput because it permits systems to overcome the 133.3 Hz update frequency bottleneck by incorporating more information in each update. The microcontroller can collect multiple samples and append these for a single future BLE update. Each BLE update therefore can carry several sensor samplings which occurred during the BLE transceiver sleep period. This avoids each BLE update being limited to a single sensor sampling.

In the case of OSSMM, the firmware buffers 10 samples and transmits them together in a single enlarged packet at an effective notification rate of 25 Hz. Because each packet carries 10 samples, the effective sampling rate becomes $10 \text{ samples} \times 25 \text{ Hz} =$

250 *Samples per second*, meeting the design target. The specific MTU value (184 bytes) and full packetisation scheme used in the OSSMM Recorder are described in Chapter 4. In practice, the author found that an appropriate connection interval and MTU must be requested on both the peripheral (headband) and central (Android app) devices to sustain this throughput.

B.4 HRMCS Modification Investigations

This section documents the investigations into modifying the HRMCS electrode configuration that were considered during MVP development but ultimately not adopted.

B.4.1 Snap-Electrode Relocation

A central practical issue with the HRMCS electrodes was how to form an electrical connection with the CTPU electrode in order to bring the signals to the AD8232 for amplification and filtering. The obvious solution would be to use the snap electrode for a connection with the use of 12 mm male snap-fasteners somehow in contact with wiring. However, the difficulty in this is due to the dimensions of the electrodes, each of which measures approximately 120 mm from snap electrode to the end of the exposed CTPU, and that the exposed portion of the snap electrode is on the opposite face of the strap. This length makes it difficult for placement on the face, without the excess non-conductive portion running into the nose or the other electrode. The exposed CTPU portion could be shortened, but the non-conductive portion could not be.

The ideal solution would have been to cut out the exposed CTPU and make new electrode snap connections. This would have resulted in only the necessary material of the CTPU being included in the prototype, without any excess. It also meant that a single CTPU electrode could be converted into two or even three individual electrodes if divided as such. If possible, this would have lowered the cost of each prototype as a single heart rate chest strap could provide multiple electrodes per unit, while potentially increasing the number of sensors used. The challenge remained in integrating a new electrode snap into the CTPU material in a manner that maintained electrical continuity and was mechanically stable.

Initial attempts to achieve this failed. The most straightforward method was using new female snap-fasteners pressed straight through the material. However, this left the exposed CTPU face, that which made contact with the body, partly covered by the flat bottom of the snap-fastener. From an electrical point of view, this was viable. From a user point of view, it was not. In addition to the issues with using

metal directly as a conductor on the skin (described in Section 3.4.2), because the snap-fastener would be on top of the CTPU it would be liable to leave markings on the user, especially after several hours of use. Therefore, from a user perspective this was not permissible, as a wearable sleep monitor will only be as effective as often as people are willing to wear it.

B.4.2 Direct CTPU Material and Conductive 3D-Printing Filament

Based on the analysis of the electrode and the understanding of this layered system, the author was unable to develop any practical methods for attaching a secure connection to the CTPU material from outside the existing snap-fastener arrangement. Since the only real material of interest here was the CTPU itself, it seemed feasible that this material should just be bought outside, instead of within the inconvenient system of the heart rate monitor chest strap. In this way, the electrodes themselves could be custom built, both to size and presumably to a pre-determined degree of conductance with appropriate carbon black doping. Unfortunately, the author was unable to find any suppliers outside of wholesale or commercial vendors.

Another solution considered was the idea that it would be possible to print this conductive TPU material directly. Indeed, there were some existing market solutions of conductive 3D filaments in 2022–2023, however their published results did not indicate they were conductive enough for biosignals detection, as published applications showed only basic DC circuit functionality. In the future it would be unsurprising if a TPU filament does not reach the market with enough carbon black doping to permit biosignal collection, including for EEG signals.

With both snap-electrode relocation and direct CTPU sourcing unavailable, MVP development proceeded with the only remaining option: preserving the original snap-electrode connection on each strap and modifying only the shape and size of the CTPU cut-outs as required. The MVP-stage assembly process consequent to this decision is described in Section 3.5.5.

B.5 Prototype #2 — Additional Trial Observations

An NREM-like observation is shown in Figure B.5. In comparison to figures with observable eye movements, little activity is shown in the EOG plots and the accelerometer plots also show virtually no activity.

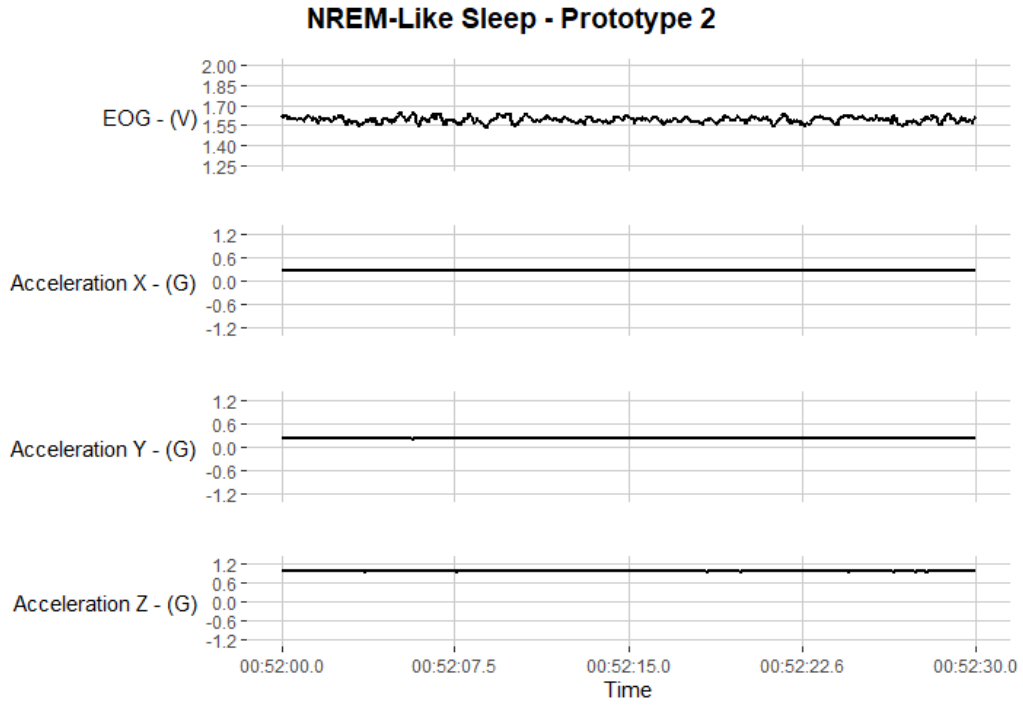


Figure B.5: NREM-Like sleep period captured with Prototype #2. Little activity is detected by the EOG and IMU sensors.

Figure B.6 shows the effects of movement on the CTPU electrodes caused by changes in pressure, producing artifacts in the EOG channel.

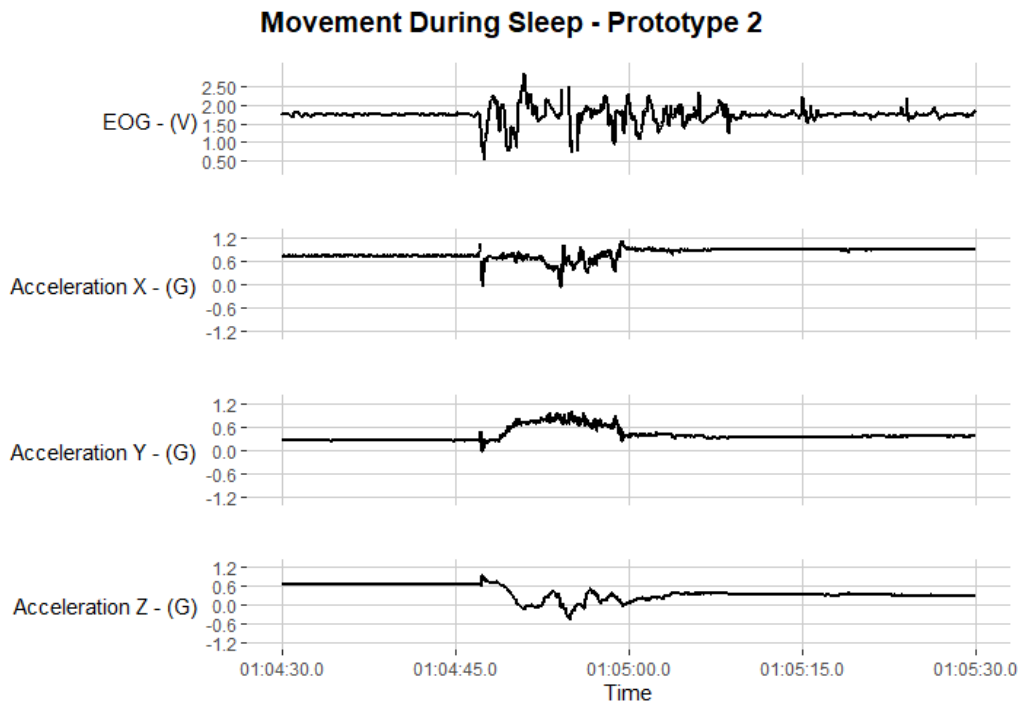


Figure B.6: Movement artifact during sleep showing characteristic accelerometer changes and corresponding EOG distortion from electrode displacement.

B.6 Component Sources and Pricing Receipts

This section provides the supplier details and purchase receipts that support the cost figures presented in Tables 3.9, 3.11, and 3.12 of Chapter 3. Unit prices may have changed since these purchases were made; the receipts are provided as a record of the sources used during this thesis.



Order information

Buyer jo293009
Seller powercells-shop
Placed on Nov 4, 2024
Payment method Credit Card
Paid on Nov 4, 2024

Shipping address



Order total

2 items	EUR 15.50
Shipping	EUR 2.80
Order total	EUR 18.30

Items bought from powercells-shop

Order number: 19-12268-06472

Quantity	Item name	Shipping service	Item price
2	LiPo 3.7V Li Ion Lithium Akku Batterie 1S JST-PH 2.0 Stecker Tablet GPS mit BMS (255510046348)	DHL Paket International	EUR 15.50



Order information

Buyer jo293009
Seller powercells-shop
Placed on Nov 1, 2022
Payment method Credit Card
Paid on Nov 1, 2022

Shipping address



Order total

2 items	EUR 15.50
Shipping	EUR 2.95
Order total	EUR 18.45

Items bought from powercells-shop

Order number: 19-09285-40669

Quantity	Item name	Shipping service	Item price
1	LiPo 3.7V Li Ion Lithium Akku Batterie 1S JST-PH 2.0 Stecker Tablet GPS mit BMS (255510046348)	DHL Paket International	EUR 7.75
1	LiPo 3.7V Li Ion Lithium Akku Batterie 1S JST-PH 2.0 Stecker Tablet GPS mit BMS (255510046348)	DHL Paket International	EUR 7.75

Sales Invoice

Farnell Components (Ireland) Ltd

Riverside One
Sir John Rogerson's Quay
Dublin 2, Dublin

[Contact-us](#)

Please email your remittance advice to:
irelandcashteam@farnell.com



www.farnell.com

Invoice Number	2449852
Invoice Date	05 NOV 2024
Order Date	05 NOV 2024
Despatch Date	05 NOV 2024
Account No	850609
Despatch No	001215644
Page No	1/1
Tracking No	

Invoice Address



Delivery Address



Customer VAT Number:

Customer Order No: 05/11/24 10.12 AM	Our Order Ref:
--------------------------------------	----------------

Line	Order Code/Description	Unit	Quantity	List Price	Discount Rate %	Net Price	Vat Rate	Amount
1	4060352 102010469 XIAO BOARD, ARM, ARDUINO BOARD Tariff Code: 84733020 ORDER PLACED BY JONATHAN GIORDANO *** PAID BY CREDIT/DEBIT CARD *** Authorisation Code: 465063 ***** Please note our new bank details: Bank of America N.A. Dublin Account Number: 55196012 Sort Code: 99-00-61 IBAN: IE86 BOFA 9900 6155 1960 12 SWIFT/BIC: BOFAIE3X ***** Do you have a new postcode? Please email iepostcode@farnell.com with the following details so we can update your account details: Account number, company name, new postcode for invoice address and any delivery address(es), if different. *****	EA	6	14.7600	0.00	14.7600	23.00	88.56

Very Important Title to the goods specific to this invoice remains with Farnell Components (Ireland) Ltd until such time as full settlement is received. Please refer to https://ie.farnell.com/terms-of-purchase for our full terms of purchase.	Vat Rate	Goods	Vat	P&P Charge	
	23.00	88.56	20.37	Invoice Subtotal	88.56
				Vat	20.37
				Invoice Total	EUR 108.93

Payment Due By: 05 NOV 2024	Payment Terms: Paid with thanks	Please Quote With Payment: 8506092449852
--------------------------------	------------------------------------	---

Farnell Components (Ireland) Limited
Company No. 113903
Registered Office: C/O HMP Secreterial Ltd,
Riverside One, Sir John Robinsons Quay,
Dublin 2, Ireland
Vat Reg No: IE4808209P
WEEE Reg. No. IE00283

Bank information
Farnell Components Ireland Ltd
ACCOUNT NO: 55196012
SORT CODE: 99-00-61
IBAN : IE86BOFA99006155196012
BICSWIFTCode : BOFAIE3X

Order Summary

Order placed 27 October 2024

Dispatch to

[Redacted]

Payment method

[Redacted]

Order Summary

Item(s) Subtotal:	EUR 30.40
Postage & Packing:	EUR 0.00
Total before VAT:	EUR 30.40
Estimated VAT:	EUR 6.99
Total:	EUR 37.39
Exchange rate guarantee fee: ▼	EUR 0.65
Grand Total:	EUR 38.04

[Exchange rate ▼](#)

1 GBP = 1.2217021 EUR

Includes the exchange rate guarantee fee



Siraya Tech Flex TPU Filament - 85A Shore, ISO 10993, High Flexibility & Toughness, High Flow & Speed, Easy to Print, Moisture-Resistant Aluminum Bag - Ideal for Footwear, Protective Gear (Black, 1kg)

Sold by: Siraya Tech Resin

£31.14

[Back to top](#)



[Help](#)

Order Summary

Order placed 18 August 2024

Dispatch to

Payment method

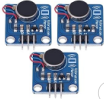
Order Summary

Item(s) Subtotal:	EUR 35.95
Postage & Packing:	EUR 0.00
Total before VAT:	EUR 35.95
Estimated VAT:	EUR 8.27
Total:	EUR 44.22
Promotion Applied:	-EUR 3.52
Exchange rate	EUR 0.71
guarantee fee: ▼	
Grand Total:	EUR 41.41

Exchange rate ▼

1 GBP = 1.1951453 EUR

Includes the exchange rate guarantee fee



3 Pcs Motor Vibration Motor Module DC5V 9000RPM Replacement for Arduino Module DIY Projects

Sold by: Amog313

£6.53

Sales Invoice

Farnell Components (Ireland) Ltd

Riverside One
Sir John Rogerson's Quay
Dublin 2, Dublin

[Contact-us](#)

Please email your remittance advice to:
Irelandcashteam@farnell.com



www.farnell.com

Invoice Number	2449852
Invoice Date	05 NOV 2024
Order Date	05 NOV 2024
Despatch Date	05 NOV 2024
Account No	850609
Despatch No	001215644
Page No	1/1
Tracking No	1Z60699Y0488781307

Invoice Address

JONATHAN GIORDANO
19 CLUAIN AOIBHINN
NEWTOWN
MAYNOOTH KILDARE
Ireland
IE

Delivery Address

JONATHAN GIORDANO
19 CLUAIN AOIBHINN
NEWTOWN
MAYNOOTH KILDARE
Ireland
IE

Customer VAT Number:

Customer Order No: 05/11/24 10.12 AM	Our Order Ref: 1230-7549/01
--------------------------------------	-----------------------------

Line	Order Code/Description	Unit	Quantity	List Price	Discount Rate %	Net Price	Vat Rate	Amount
1	4060352 102010469 XIAO BOARD, ARM, ARDUINO BOARD Tariff Code: 84733020 ORDER PLACED BY JONATHAN GIORDANO *** PAID BY CREDIT/DEBIT CARD *** Authorisation Code: 465063 ***** Please note our new bank details: Bank of America N.A. Dublin Account Number: 55196012 Sort Code: 99-00-61 IBAN: IE86 BOFA 9900 6155 1960 12 SWIFT/BIC: BOFAIE3X ***** Do you have a new postcode? Please email iepostcode@farnell.com with the following details so we can update your account details: Account number, company name, new postcode for invoice address and any delivery address(es), if different. *****	EA	6	14.7600	0.00	14.7600	23.00	88.56

Very Important Title to the goods specific to this invoice remains with Farnell Components (Ireland) Ltd until such time as full settlement is received. Please refer to https://ie.farnell.com/terms-of-purchase for our full terms of purchase.	Vat Rate	Goods	Vat	P&P Charge	
	23.00	88.56	20.37	Invoice Subtotal	88.56
				Vat	20.37
				Invoice Total	EUR 108.93

Payment Due By: 05 NOV 2024	Payment Terms: Paid with thanks	Please Quote With Payment: 8506092449852
--------------------------------	------------------------------------	---

Farnell Components (Ireland) Limited
Company No. 113903
Registered Office: C/O HMP Secreterial Ltd,
Riverside One, Sir John Robinsons Quay,
Dublin 2, Ireland
Vat Reg No: IE4808209P
WEEE Reg. No. IE00283

Bank information

Farnell Components Ireland Ltd
ACCOUNT NO: 55196012
SORT CODE: 99-00-61
IBAN : IE86BOFA99006155196012
BICSWIFTCode : BOFAIE3X

← **Order P1017**

Confirmed on 19. Oct. 2022

[Buy again](#)**Link expired**

To see your order details, log in or send a new link to the email address or phone number you used at checkout.

[Send new link](#)[Sign in](#)**On the road**

19. Oct. 2022

- Confirmed
19. Oct. 2022



Adafruit Feather nRF52840 Sense, 4516 €41.34

Interim total 41,34 €

Shipping 6,50 €

Total EUR **€47.84**

Incl. 8,95 € Tax

Contact information

[Redacted]

Delivery address

[Redacted]

Shipping method

[Redacted]

Payment

[Redacted]

Billing address

[Redacted]



Summary

Order ID:

Order time:

Mar 27, 2024

Subtotal:

US \$23.70

Shipping fee:

US \$0.00

Total:

US \$23.70

US \$4.43 VAT included

Shipping address

Payment method

USD 23.70

Item detail



Pulsesensor pulse heart rate senso

US \$2.37x10

GOHJMY Electronics Store



Summary

Order ID:

Order time:

Mar 27, 2024

Subtotal:

US \$46.26

Shipping fee:

US \$0.00

Total:

US \$46.26

US \$8.65 VAT included

Shipping address

Payment method

USD 46.26

Item detail



AD8232 ECG Physiological Measure
Module

US \$2.57x18

YueLung Store



Summary

Order ID:

Order time:

Mar 27, 2024

Subtotal:

US \$55.61

Shipping fee:

US \$2.78

Total:

US \$58.39

US \$10.92 VAT included

Shipping address

Payment method

USD 58.39

Item detail



20Pcs Heart Rate Chest Strap Adjus
Black

US \$55.61x1

FUN DAY Store

Appendix — Technical Specification of the Open-Source Release

Appendix C provides supplementary information for Chapter 4

C.1 OSSMM Recorder (headband) Arduino Code

The complete source code used during the technical evaluation is reproduced below for the permanent record.

The link to a digital copy on Github can be found here: <https://github.com/jvgiordano/OSSMM/tree/main/OSSMM%20-%20V1.0.4%20System%20Code/Microcontroller%20Code>

```

1
2  /*
3  ****
4
5  Notes:
6  Program: OSSMM V1.0.4
7  By: Jonny Giordano
8
9  Created: Friday July 15th, 2022
10 Switch to M5StickC-Plus: Tuesday June 21st, 2022
11 Switch to Adafruit Feather Sense: Friday October 25th, 2022
12 Working BLE on Adafruit: Monday November 14th, 2022
13 Change to Prototype 2L with one EOG sensor

```

```

13  Switch to Seeed Xiao Sense: Sometime April, 2024?
14  Added: April 21st, 2024
15  -"Xiao_Sense_LSM6DS3.h" and associated files were taken from the
16  Arduino LSM6DS3 library and modified by Github user 'aovestdipaperino'
17
18  Update: November 2nd, 2024
19  - Merged Code version with hardware version numbering, to keep things
    simple
20
21  CRITICAL UPDATE: December 2024
22  - Added automatic reconnection support
23  - Implemented graceful disconnect handling with timeout
24  - Added connection state management
25  - Device stays awake for reconnection attempts after disconnect
26
27  Update: April 25th, 2025
28  - Added easily adjustable sampling frequency
29
30  Final fixes: April30-May 2nd, 2025
31  - Fixed BLE loop error (last 18 bytes were not being sent)
32  - Implemented "Just Works" pair bonding
33
34  More fixes: August 24th, 2025
35  - Enabled the POW characteristic to shut-off the MCU
36  seperately from losing BLE connection. So, now reconnection
37  attempts are back.
38  - BLE is now AES encrypted,
39
40
41  ****
42  */
43
44  /* -----
45     Device Name and Version
46     - OSSMM VX.X.X
47     -----m-----
48  */
49  char DeviceName[] = "OSSMM V1.0.4 - Power Testing"; // Version
50
51  /* -----
52     Set the Sampling Frequency
53     -----
54  */
55  unsigned int SamplingFrequency = 250; // Set Sampling Frequency (Hz)
56  unsigned int sampling_interval = 0.9 * 1000000 / SamplingFrequency; //

```

```

        Calculate sampling interval in microseconds
57 unsigned long lastSampleTime = 0;
58
59 /* -----
60     Connection Management Variables
61     -----
62 */
63 unsigned long disconnectTime = 0;           // Time when disconnect
        occurred
64 const unsigned long RECONNECT_TIMEOUT = 600000; // 10 minutes timeout
        for reconnection (600000 ms)
65 const unsigned long RECONNECT_ADVERTISE_INTERVAL = 5000; // Re-advertise
        every 5 seconds
66 unsigned long lastAdvertiseTime = 0;        // Last time we
        refreshed advertising
67 bool waitingForReconnect = false;          // Flag for reconnection
        state
68 bool intentionalShutdown = false;          // Flag for intentional
        shutdown via power command
69 unsigned int reconnectAttempts = 0;         // Counter for
        reconnection attempts
70 const unsigned int MAX_RECONNECT_ATTEMPTS = 120; // Max attempts (120 * 5
        sec = 10 minutes)
71
72 /* -----
73     Load Libraries
74     -----
75 */
76 #include <bluefruit.h>           // Adafruit Bluetooth Library
77 #include "Xiao_Sense_LSM6DS3.h" // Modified LSM6DS3 IMU Library (
        accelerometer, gyroscope)
78 // #include <PDM.h>               // microphone library (not yet used)
79
80 /* -----
81     Initialize Variables for Sensors (Accelerometer, Gyroscope, EOG), Timers,
        BLE Transmission, and Power/Serial Connections
82     -----
83 */
84 unsigned long Start; // timer variable for measuring code
85 unsigned long End;   // timer variable for measuring code
86 unsigned long TransmitStart; // timer variable for measuring measurement+
        transmission time
87 unsigned long Start2; // timer variable for measuring code
88 unsigned long End2;   // timer variable for measuring code
89
90 bool isConnected = false; // If device is connected through BLE

```



```

91  bool initial = false; // If this is the 1st sampling loop on connection/
    disconnection
92
93  uint16_t update_num = 0;      // Transmission number
94  uint16_t loop_count = 0; // Number of messages added to a BLE packet, resets
    at = X
95  char mySleepDataPacket[181]; // Create 'c string' for one transmission, 180
    byte + null terminator
96
97  float gyroData[3] = { 0.0, 0.0, 0.0 }; // Gyroscope data, [x,y,z] format,
    range = [-2000,+2000]
98  float accData[3] = { 0.0, 0.0, 0.0 }; // Accelerometer, [x,y,z] format, range
    = [-8.0,+8.0]
99
100 uint16_t gyroVals[3];          // Formated gyroscope data to [0,+4000], [x,y,z
    ] format
101 uint16_t accVals[3]; // Formated accelerometer data to [0, 1600], [x,y,z]
    format
102
103 uint16_t eog = 0;              // EOG measurement
104 uint16_t hr = 0; // Heart Rate (pulse) measurement
105
106 /* -----
107  USB Power and Serial Connection
108  -----
109  */
110 bool powerUSB = (NRF_POWER->USBREGSTATUS) & POWER_USBREGSTATUS_VBUSDETECT_Msk
    ; // Determine if receiving power via USB
111 bool resetButtonLaunch = (NRF_POWER->RESETREAS) &
    POWER_RESETREAS_RESETPIN_Msk; // Determine if Start-up was due to Reset
    button ("ON" Button)
112
113 /* -----
114  Initialize BLE Variables, Service, and Characteristics
115  -----
116  */
117 #define OSSMM_SER_UUID "5aee1a8a-08de-11ed-861d-0242ac120002" // Unique
    Service ID (randomly generated)
118 #define OSSMM_CHAR_UUID "405992d6-0cf2-11ed-861d-0242ac120002" // Unique
    Characteristic ID (randomly generated)
119 #define OSSMM_MOD_UUID "1aa00c0d-469a-426b-985c-8299084aed72" // Unique
    Characteristic ID (randomly generated)
120 #define OSSMM_POW_UUID "018ec2b5-7c82-7773-95e2-a5f374275f0b" // Unique
    Characteristic ID (randomly generated)
121
122 BLEService BLEservice = BLEService(OSSMM_SER_UUID); // Service UUID for BLE,

```

```

    generated with random UUID generator
123 BLECharacteristic SleepData = BLECharacteristic(OSSMM_CHAR_UUID); //
    Characteristic ID for Sleep Data (IMU, EOG, HR)
124 BLECharacteristic SleepModulator = BLECharacteristic(OSSMM_MOD_UUID); //
    Characteristic ID for Sleep Modulator
125 BLECharacteristic PowerControl = BLECharacteristic(OSSMM_POW_UUID); //
    Characteristic ID for Power Control
126
127 // Forward declarations for callbacks
128 void sleep_modulator_callback(uint16_t conn_hdl, BLECharacteristic* chr,
    uint8_t* data, uint16_t len);
129 void power_control_callback(uint16_t conn_hdl, BLECharacteristic* chr,
    uint8_t* data, uint16_t len);
130 void restartAdvertising();
131 void handleReconnectionTimeout();
132
133 /* -----
134    Security and Connection Callbacks (BLE Bonding)
135    -----
136 */
137 // Callback when a connection is secured with encryption
138 void connection_secured_callback(uint16_t conn_handle) {
139     BLEConnection* connection = Bluefruit.Connection(conn_handle);
140     if (connection->secured()) {
141         Serial.println("Connection secured: Role = Peripheral");
142         // Get information about the bonded peer
143         ble_gap_addr_t peer_addr = connection->getPeerAddr();
144         // Print bonded peer address
145         /*
146         Serial.print("Peer address: ");
147         Serial.print(peer_addr.addr[5], HEX); Serial.print(":");
148         Serial.print(peer_addr.addr[4], HEX); Serial.print(":");
149         Serial.print(peer_addr.addr[3], HEX); Serial.print(":");
150         Serial.print(peer_addr.addr[2], HEX); Serial.print(":");
151         Serial.print(peer_addr.addr[1], HEX); Serial.print(":");
152         Serial.print(peer_addr.addr[0], HEX);
153         Serial.println();
154         */
155
156         // Request pairing if not already bonded
157         if (!connection->bonded()) {
158             Serial.println("Not bonded yet, requesting pairing");
159             connection->requestPairing();
160         }
161     } else {
162         Serial.println("Connection NOT secured");
    }
}

```

```

163   }
164 }
165
166 // Callback when pairing process completes
167 void pair_complete_callback(uint16_t conn_handle, uint8_t auth_status) {
168     if (auth_status == BLE_GAP_SEC_STATUS_SUCCESS) {
169         Serial.println("Pairing successful!");
170         // Flash blue LED to indicate successful pairing
171         for (int i = 0; i < 5; i++) {
172             analogWrite(LED_BLUE, 0); // Turn ON Blue LED
173             delay(100);
174             analogWrite(LED_BLUE, 255); // Turn OFF Blue LED
175             delay(100);
176         }
177     } else {
178         Serial.print("Pairing failed with status: ");
179         Serial.println(auth_status);
180     }
181 }
182
183 /* -----
184    BLE ServerCallBack Functions
185    -----
186 */
187 void connect_callback(uint16_t conn_handle) {
188     BLEConnection* connection = Bluefruit.Connection(conn_handle);
189
190     char central_name[32] = { 0 };
191     connection->getPeerName(central_name, sizeof(central_name));
192
193     Serial.print("Connected to ");
194     Serial.println(central_name);
195
196     // Reset reconnection state on successful connection
197     waitingForReconnect = false;
198     disconnectTime = 0;
199     reconnectAttempts = 0;
200     intentionalShutdown = false;
201
202     SleepModulator.write8(0x00); // Confirm GATT information of modulation
203                                // shows it is off
204     PowerControl.write8(0x00); // Confirm GATT information of power control
205                                // shows it is off
206
207     delay(1000);
208     isConnected = true;

```

```

207   initial = true;
208 }
209
210 void disconnect_callback(uint16_t conn_handle, uint8_t reason) {
211   (void)conn_handle;
212   (void)reason;
213
214   Serial.println();
215   Serial.print("Disconnected, reason = 0x");
216   Serial.println(reason, HEX);
217
218   SleepModulator.write8(0x00);
219   PowerControl.write8(0x00);
220
221   isConnected = false;
222   initial = true;
223
224   // Check if this was an intentional shutdown
225   if (intentionalShutdown) {
226     Serial.println("Intentional shutdown - going to deep sleep immediately");
227     // Turn off power pins
228     digitalWrite(A4, LOW);
229     digitalWrite(A5, LOW);
230     // Enter deep sleep
231     NRF_POWER->SYSTEMOFF = 1;
232   } else {
233     // Unintentional disconnect - enter reconnection mode
234     Serial.println("Unexpected disconnect - entering reconnection mode");
235     waitingForReconnect = true;
236     disconnectTime = millis();
237     reconnectAttempts = 0;
238
239     // Keep sensors powered for quick reconnection
240     digitalWrite(A5, HIGH);
241
242     // Ensure advertising is enabled
243     restartAdvertising();
244
245     // Flash yellow LED pattern to indicate reconnection mode
246     for (int i = 0; i < 3; i++) {
247       analogWrite(LED_RED, 128); // Half brightness red
248       analogWrite(LED_GREEN, 128); // Half brightness green = yellow
249       delay(200);
250       analogWrite(LED_RED, 255); // Turn OFF
251       analogWrite(LED_GREEN, 255); // Turn OFF
252       delay(200);

```

```

253     }
254 }
255 }
256
257 /* -----
258  Restart Advertising Function
259  -----
260 */
261 void restartAdvertising() {
262     // Stop advertising if currently active
263     if (Bluefruit.Advertising.isRunning()) {
264         Bluefruit.Advertising.stop();
265         delay(100);
266     }
267
268     // Clear any existing advertising data
269     Bluefruit.Advertising.clearData();
270     Bluefruit.ScanResponse.clearData();
271
272     // Re-configure advertising
273     Bluefruit.Advertising.addFlags(BLE_GAP_ADV_FLAGS_LE_ONLY_GENERAL_DISC_MODE)
274         ;
275     Bluefruit.Advertising.addTxPower();
276     Bluefruit.Advertising.addService(BLEservice);
277     Bluefruit.Advertising.addAppearance(BLE_APPEARANCE_GENERIC_WATCH);
278     Bluefruit.ScanResponse.addName();
279
280     // Set aggressive advertising intervals for reconnection
281     Bluefruit.Advertising.setInterval(32, 100); // Fast advertising: 20ms to
282         62.5ms
283     Bluefruit.Advertising.setFastTimeout(0); // Stay in fast mode
284
285     // Start advertising indefinitely
286     Bluefruit.Advertising.start(0);
287
288     Serial.println("Advertising restarted for reconnection");
289 }
290
291 /* -----
292  Handle Reconnection Timeout
293  -----
294 */
295 void handleReconnectionTimeout() {
296     if (!waitingForReconnect) return;
297
298     unsigned long currentTime = millis();

```

```

297   unsigned long timeSinceDisconnect = currentTime - disconnectTime;
298
299   // Check if we should refresh advertising
300   if (currentTime - lastAdvertiseTime >= RECONNECT_ADVERTISE_INTERVAL) {
301       lastAdvertiseTime = currentTime;
302       reconnectAttempts++;
303
304       Serial.print("Reconnection attempt #");
305       Serial.print(reconnectAttempts);
306       Serial.print(" of ");
307       Serial.println(MAX_RECONNECT_ATTEMPTS);
308
309       // Refresh advertising to ensure visibility
310       restartAdvertising();
311
312       // Quick LED flash to show activity
313       analogWrite(LED_BLUE, 0);    // Turn ON Blue LED
314       delay(50);
315       analogWrite(LED_BLUE, 255);  // Turn OFF Blue LED
316   }
317
318   // Check if timeout has been reached
319   if (timeSinceDisconnect >= RECONNECT_TIMEOUT || reconnectAttempts >=
320       MAX_RECONNECT_ATTEMPTS) {
321       Serial.println("Reconnection timeout reached - going to deep sleep");
322
323       waitingForReconnect = false;
324
325       // Flash red LED to indicate giving up
326       for (int i = 0; i < 5; i++) {
327           digitalWrite(LED_RED, LOW);    // Turn ON Red LED
328           delay(500);
329           digitalWrite(LED_RED, HIGH);   // Turn OFF Red LED
330           delay(250);
331       }
332
333       // Turn off power pins
334       digitalWrite(A4, LOW);
335       digitalWrite(A5, LOW);
336
337       // Enter deep sleep
338       NRF_POWER->SYSTEMOFF = 1;
339   }
340
341   /* -----

```

```

342  Connect and Disconnect Functions
343  - Blinks LED upon connection/disconnection
344  -----
345  */
346  void nowConnected() {
347      for (int i = 0; i < 8; i++) {
348          digitalWrite(LED_GREEN, LOW); // Turn the GREEN LED on
349          delay(250);
350          digitalWrite(LED_GREEN, HIGH); // Turn the GREEN LED off
351          delay(250);
352      }
353
354      // TURN ON PULSE SENSOR and AD8232
355      digitalWrite(A5, HIGH);
356      digitalWrite(LED_BUILTIN, HIGH); // Ensure LED is off to save power
357  }
358
359  void nowDisconnected() {
360      // Don't flash LEDs here - handled in disconnect_callback
361      digitalWrite(LED_BUILTIN, HIGH); // Ensure LED is off to save power
362  }
363
364  void setup() {
365      /* -----
366         Initialize Adafruit Feather Sense, IMU
367         -----
368      */
369      Serial.begin(115200);
370      delay(500);
371      Serial.println("Xiao Sense nRF52840 - OSSMM Application V1.0.4");
372      delay(500);
373
374      if (!IMU.begin()) {
375          Serial.println("Failed to initialize IMU!");
376      }
377      Serial.println("IMU initialized!");
378      Wire.setClock(400000); // Increase I2C speed to 400 Khz
379
380      /* -----
381         Initialize INPUT, CHARGING, POWER, and LED Pins
382         -----
383      */
384      // Battery Charging
385      pinMode(22, OUTPUT);
386      digitalWrite(22, LOW); // Set LOW -> 100 mA charge
387

```

```

388 // AD8232 PINS
389 pinMode(A0, INPUT); // EOG/EEG analog signal
390
391 // Heartrate Sensor
392 pinMode(A1, INPUT); // Heartrate monitor analog signal
393
394 // Power pin (Delivers power to AD8232 and HR Sensor)
395 pinMode(A5, OUTPUT);
396 digitalWrite(A5, LOW);
397
398 // Vibration Disc
399 pinMode(A4, OUTPUT);
400 digitalWrite(A4, LOW);
401
402 /* -----
403      Detect USB Charging vs Serial Connection - Detect Start-Up Reason
404      -----
405 */
406 // Check if USB Power is for Charging or for Serial Connection
407 if (powerUSB) {
408     unsigned long serialTimeout = millis() + 10000; // Wait 10 Seconds for
409     Serial Connection
410
411     while (!Serial && (millis() < serialTimeout)) {
412         delay(150);
413         Serial.println("OSSMM - Serial Connection Attempt...");
414         digitalWrite(LED_RED, !digitalRead(LED_RED));
415     }
416
417     if (!Serial) {
418         for (int i = 0; i < 8; i++) {
419             digitalWrite(LED_RED, LOW);
420             delay(500);
421             digitalWrite(LED_RED, HIGH);
422             delay(250);
423         }
424         NRF_POWER->SYSTEMOFF = 1; // USB Power only, go to sleep!
425     }
426     else {
427         Serial.println("OSSMM - Serial Connection Detected");
428         digitalWrite(LED_RED, HIGH);
429         digitalWrite(LED_GREEN, LOW);
430         digitalWrite(LED_BLUE, LOW);
431     }
432 }

```



```

433 // Check if Start-up was Intentional
434 if (!resetButtonLaunch){
435     for (int i = 0; i < 8; i++) {
436         digitalWrite(LED_RED, LOW);
437         delay(500);
438         digitalWrite(LED_RED, HIGH);
439         delay(250);
440     }
441     NRF_POWER->SYSTEMOFF = 1;
442 }
443
444 /* -----
445     Initialize Bluetooth Low Energy (BLE) with Security
446     -----
447 */
448 Bluefruit.configPrphBandwidth(BANDWIDTH_MAX);
449 Bluefruit.configUuid128Count(15);
450
451 Bluefruit.begin();
452 Bluefruit.setTxPower(0); // Set 0dB for better battery life
453 Bluefruit.autoConnLed(false);
454 Bluefruit.setName(DeviceName);
455
456 // Configure security for "Just Works" pairing
457 Bluefruit.Security.setIOCaps(false, false, false);
458 Bluefruit.Security.setPairPasskeyCallback(NULL);
459 Bluefruit.Security.setPairCompleteCallback(pair_complete_callback);
460 Bluefruit.Security.setSecuredCallback(connection_secured_callback);
461
462 Bluefruit.Periph.setConnectCallback(connect_callback);
463 Bluefruit.Periph.setDisconnectCallback(disconnect_callback);
464 Bluefruit.Periph.setConnInterval(6, 8); // 7.5 - 15 ms
465
466 BLEService.begin();
467
468 // Configure SLEEP DATA Characteristic
469 SleepData.setProperties(CHR_PROPS_NOTIFY);
470 SleepData.setPermission(SECMODE_ENC_NO_MITM, SECMODE_NO_ACCESS);
471 SleepData.setFixedLen(180);
472 SleepData.begin();
473
474 // Configure POWER CONTROL Characteristic
475 PowerControl.setProperties(CHR_PROPS_READ | CHR_PROPS_WRITE);
476 PowerControl.setPermission(SECMODE_ENC_NO_MITM, SECMODE_ENC_NO_MITM);
477 PowerControl.setFixedLen(1);
478 PowerControl.begin();

```

```

479   PowerControl.write8(0x00);
480   PowerControl.setWriteCallback(power_control_callback);
481
482   // Configure SLEEP MODULATOR Characteristic
483   SleepModulator.setProperties(CHR_PROPS_READ | CHR_PROPS_WRITE);
484   SleepModulator.setPermission(SECMODE_ENC_NO_MITM, SECMODE_ENC_NO_MITM);
485   SleepModulator.setFixedLen(1);
486   SleepModulator.begin();
487   SleepModulator.write8(0x00);
488   SleepModulator.setWriteCallback(sleep_modulator_callback);
489
490   // Advertising configuration
491   Bluefruit.Advertising.addFlags(BLE_GAP_ADV_FLAGS_LE_ONLY_GENERAL_DISC_MODE)
492   ;
493   Bluefruit.Advertising.addTxPower();
494   Bluefruit.Advertising.addService(BLEservice);
495   Bluefruit.Advertising.addAppearance(BLE_APPEARANCE_GENERIC_WATCH);
496   Bluefruit.ScanResponse.addName();
497
498   Bluefruit.Advertising.restartOnDisconnect(true);
499   Bluefruit.Advertising.setInterval(100, 244); // Normal advertising
500   Bluefruit.Advertising.setFastTimeout(30);
501   Bluefruit.Advertising.start(0); // Don't stop advertising
502
503   // Startup complete LED sequence
504   for (int j = 0; j <= 5; j++) {
505     for (int i = 255; i >= 0; i = i - 17) {
506       analogWrite(LED_BLUE, i);
507       delay(50);
508     }
509     analogWrite(LED_BLUE, 255);
510     delay(100);
511
512     Serial.println("Device ready - advertising started");
513   }
514
515   void loop() {
516     /*-----
517     Handle reconnection timeout if waiting for reconnect
518     -----
519     */
520     if (waitingForReconnect) {
521       handleReconnectionTimeout();
522       delay(100); // Small delay while waiting for reconnection
523       return; // Skip normal loop processing

```

```

$24   }
$25
$26   /*-----
$27       Check if BLE is connected, and if initial loop
$28   -----
$29   */
$30   if (isConnected == true) {
$31       if (initial == true) {
$32           nowConnected();
$33           initial = false;
$34           lastSampleTime = micros();
$35       }
$36
$37       while (isConnected == true) {
$38           /*-----
$39               Collect Measurements using precise timing at set frequency
$40           -----
$41           */
$42           unsigned long currentTime = micros();
$43           if (currentTime - lastSampleTime >= sampling_interval) {
$44
$45               if (currentTime - lastSampleTime >= 2 * sampling_interval) {
$46                   lastSampleTime = currentTime;
$47               }
$48               else {
$49                   lastSampleTime = lastSampleTime + sampling_interval;
$50               }
$51
$52               // Get IMU Data
$53               IMU.readAcceleration(accData[0], accData[1], accData[2]);
$54               IMU.readGyroscope(gyroData[0], gyroData[1], gyroData[2]);
$55
$56               // Convert IMU Data
$57               for (int i = 0; i < 3; i++) {
$58                   gyroVals[i] = (uint16_t)(gyroData[i] + 2000);
$59                   accVals[i] = (uint16_t)((accData[i] + 8.0) * 100);
$60               }
$61
$62               // Get ADC data
$63               eog = analogRead(A0);
$64               hr = analogRead(A1);
$65
$66           /* -----
$67               BLE Update
$68           -----
$69           */

```

```

$70     updateBLE();
$71     }
$72
$73     yield(); // Give other system tasks a chance to run
$74     }
$75 } else {
$76     if (initial == true) {
$77         nowDisconnected();
$78         initial = false;
$79     }
$80     delay(100);
$81 }
$82 }
$83
$84 /* -----
$85 BLE Update Function ( Optimized transmission is inspired by Souichirou
$86 Kikuchi M5StickC BLE project)
$87 -----
$88 */
$89 void updateBLE() {
$90     if (loop_count >= 10) {
$91         SleepData.notify(mySleepDataPacket, 180);
$92         loop_count = 0;
$93     } else {
$94         mySleepDataPacket[(18 * loop_count) + 0] = (char)(update_num & 0xff);
$95         mySleepDataPacket[(18 * loop_count) + 1] = (char)((update_num >> 8) & 0
$96         xff);
$97         mySleepDataPacket[(18 * loop_count) + 2] = (char)(eog & 0xff);
$98         mySleepDataPacket[(18 * loop_count) + 3] = (char)((eog >> 8) & 0xff);
$99         mySleepDataPacket[(18 * loop_count) + 4] = (char)(hr & 0xff);
$100        mySleepDataPacket[(18 * loop_count) + 5] = (char)((hr >> 8) & 0xff);
$101        mySleepDataPacket[(18 * loop_count) + 6] = (char)(accVals[0] & 0xff);
$102        mySleepDataPacket[(18 * loop_count) + 7] = ((char)((accVals[0] >> 8) & 0
$103        xff));
$104        mySleepDataPacket[(18 * loop_count) + 8] = ((char)(accVals[1] & 0xff));
$105        mySleepDataPacket[(18 * loop_count) + 9] = ((char)((accVals[1] >> 8) & 0
$106        xff));
$107        mySleepDataPacket[(18 * loop_count) + 10] = ((char)(accVals[2] & 0xff));
$108        mySleepDataPacket[(18 * loop_count) + 11] = ((char)((accVals[2] >> 8) & 0
$109        xff));
$110        mySleepDataPacket[(18 * loop_count) + 12] = ((char)(gyroVals[0] & 0xff));
$111        mySleepDataPacket[(18 * loop_count) + 13] = ((char)((gyroVals[0] >> 8) &
$112        0xff));
$113        mySleepDataPacket[(18 * loop_count) + 14] = ((char)(gyroVals[1] & 0xff));
$114        mySleepDataPacket[(18 * loop_count) + 15] = ((char)((gyroVals[1] >> 8) &
$115        0xff));

```

```

609     mySleepDataPacket[(18 * loop_count) + 16] = ((char)(gyroVals[2] & 0xff));
610     mySleepDataPacket[(18 * loop_count) + 17] = ((char)((gyroVals[2] >> 8) &
        0xff));
611
612     update_num++;
613     loop_count++;
614 }
615 }
616
617 /* -----
618     Sleep Modulator Callback
619     -----
620 */
621 void sleep_modulator_callback(uint16_t conn_hdl, BLECharacteristic* chr,
        uint8_t* data, uint16_t len) {
622     Serial.println("Arrived at Sleep Modulator Callback");
623     (void)conn_hdl;
624     (void)chr;
625     (void)len;
626
627     if (data[0]) {
628         for (int i = 0; i < 2; i++) {
629             for (int j = 0; j < 2; j++) {
630                 digitalWrite(A4, HIGH);
631                 Serial.println("Vibration Motor ON");
632                 delay(400);
633                 digitalWrite(A4, LOW);
634                 Serial.println("Vibration Motor OFF");
635                 delay(60);
636             }
637             delay(125);
638         }
639     }
640 }
641
642 /* -----
643     Power Control Callback
644     -----
645 */
646 void power_control_callback(uint16_t conn_hdl, BLECharacteristic* chr,
        uint8_t* data, uint16_t len) {
647     Serial.println("Arrived at Power Control Callback");
648     (void)conn_hdl;
649     (void)chr;
650     (void)len;
651

```

```
652 // data = 2 -> Turn Off Device
653 if (data[0] == 2) {
654     Serial.println("Received Turn Off command. Shutting down.");
655
656     // Set flag for intentional shutdown
657     intentionalShutdown = true;
658
659     // Turn off power to external sensors and vibration motor
660     digitalWrite(A5, LOW);
661     digitalWrite(A4, LOW);
662
663     // Disconnect if connected
664     if (isConnected) {
665         // Find the connection handle and disconnect
666         for (uint16_t conn_hdl = 0; conn_hdl < BLE_MAX_CONNECTION; conn_hdl++)
667         {
668             BLEConnection* connection = Bluefruit.Connection(conn_hdl);
669             if (connection && connection->connected()) {
670                 connection->disconnect();
671                 delay(100); // Give time for disconnect to process
672                 break;
673             }
674         }
675
676         // Go to deep sleep
677         NRF_POWER->SYSTEMOFF = 1;
678     }
679 }
680
681
```

Listing C.1: OSSMM V1.0.4 Firmware — Main Implementation

Appendix — Proof-of-Concept Evaluation: Classical Machine Learning

Appendix D provides supplementary information for [Chapter 5](#)

D.1 Feature Selections

Table D.1: PPG-derived heart-rate and HRV features using FFT, including LF/HF spectral measures.

Feature name	Units	Type	Description / Notes
hr_mean_fft	bpm	HR	Heart rate from dominant spectral peak in 0.6–3.5 Hz (36–210 bpm), with harmonic ($2\times/3\times$) scoring to favor physiologic peaks.
hrv_sdnfft	ms	HRV	SDNN proxy from spectral spread around the dominant HR frequency using a default $\pm 10\%$ window.
hrv_rmssd_fft	ms	HRV	RMSSD proxy derived from relative high-frequency power near $\sim 2\times$ HR.
hrv_sdnfft_5pct	ms	HRV	SDNN proxy with a narrower $\pm 5\%$ window centered on the dominant HR frequency.
hrv_sdnfft_15pct	ms	HRV	SDNN proxy with a wider $\pm 15\%$ window centered on the dominant HR frequency.
pulse_lf_power	a.u.	Spectral	Power in 0.04–0.15 Hz band of the normalized pulse spectrum.
pulse_hf_power	a.u.	Spectral	Power in 0.15–0.40 Hz band of the normalized pulse spectrum.
pulse_lf_hf_ratio	—	Spectral	Ratio of LF to HF power in the normalized spectrum.

Table D.2: IMU-derived features from acceleration and gyroscope signals.

Feature	Units	Description
Gyroscope		
magnitude_mean	deg/s	Mean rotational magnitude $\sqrt{\omega_x^2 + \omega_y^2 + \omega_z^2}$ across the window.
magnitude_std	deg/s	Std. dev. of rotational magnitude (rotation variability).
bursts	count	Number of samples where rotational magnitude $>$ mean+2·SD.
burst_percentage	% (0–100)	Percentage of samples where rotational magnitude $>$ mean+2·SD.
Accelerometer		
magnitude_mean	g	Mean acceleration magnitude $\sqrt{a_x^2 + a_y^2 + a_z^2}$ across the window.
magnitude_std	g	Std. dev. of acceleration magnitude (movement variability).
magnitude_max	g	Maximum acceleration magnitude observed in the window.
bursts	count	Number of samples where acceleration magnitude $>$ mean+2·SD.
burst_percentage	% (0–100)	Percentage of samples where acceleration magnitude $>$ mean+2·SD.
x_std	g	Std. dev. of a_x (lateral axis variability).
y_std	g	Std. dev. of a_y (longitudinal axis variability).
z_std	g	Std. dev. of a_z (vertical axis variability).

Notes. Table shows feature names without prefixes for readability. In the dataset, gyroscope features are prefixed `imu_gyro_` and accelerometer features `imu_acc_` (e.g., `magnitude_std` $\rightarrow$ `imu_gyro_magnitude_std` or `imu_acc_magnitude_std`). Burst thresholds are computed within-window as mean+2·SD for the respective magnitude.

Table D.3: Raw EOG burst features capturing eye-movement activity. Names shown without prefixes; see notes for how groups map to full column names.

Feature	Units	Description
Global movement (raw)		
bursts_high	count	Samples above the global high movement threshold.
bursts_low	count	Samples below the global low movement threshold.
percentage	% (0–100)	Fraction of samples outside global movement bounds (high or low).
Global signal (raw)		
bursts_high	count	Samples above the global high signal threshold but below the movement-high threshold.
bursts_low	count	Samples below the global low signal threshold but above the movement-low threshold.
percentage	% (0–100)	Fraction of samples within global signal bands (excluding movement outliers).
Local movement (raw)		
bursts_high	count	Samples above the local high movement threshold (mean + 4·SD).
bursts_low	count	Samples below the local low movement threshold (mean − 4·SD).
percentage	% (0–100)	Fraction of samples outside local movement bounds (high or low).
Local signal (raw)		
bursts_high	count	Samples above the local high signal threshold (mean + 3·SD) but below local movement-high.
bursts_low	count	Samples below the local low signal threshold (mean − 3·SD) but above local movement-low.
percentage	% (0–100)	Fraction of samples within local signal bands (excluding movement outliers).

Notes. (1) *Global vs Local thresholds:* Global thresholds use dataset-wide min/max to set bounds (movement $\approx$ top/bottom 10%, signal $\approx$ middle 30–70% range in the code), whereas Local thresholds use per-epoch mean/SD (movement ± 4 SD; signal ± 3 SD). (2) *Movement vs Signal:* “Movement” captures extreme excursions (likely large eye movements/blinks); “Signal” captures substantial but non-extreme deflections. Signal percentages exclude movement outliers.

(3) *Prefix mapping (for raw features):* Global movement $\rightarrow$ eog_raw_global_movement_{bursts_high, bursts_low, percentage} Global signal $\rightarrow$ eog_raw_global_signal_{bursts_high, bursts_low, percentage} Local movement $\rightarrow$ eog_raw_local_movement_{bursts_high, bursts_low, percentage} Local signal $\rightarrow$ eog_raw_local_signal_{bursts_high, bursts_low, percentage}.

Table D.4: Raw EOG frequency-band features (prefix omitted). Band ranges follow the implementation in the code.

Feature	Band (Hz)	Units	Description
Frequency bands (raw)			
delta_power	≤ 4	a.u.	Sum of power in the ≤ 4 Hz band.
theta_power	$> 4 \ \& \ \leq 8$	a.u.	Sum of power in the 4–8 Hz band.
alpha_power	$> 8 \ \& \ \leq 13$	a.u.	Sum of power in the 8–13 Hz band.
beta_power	$> 13 \ \& \ \leq 30$	a.u.	Sum of power in the 13–30 Hz band.
gamma_power	$> 30 \ \& \ \leq 100$	a.u.	Sum of power in the 30–100 Hz band.

Notes. (1) “a.u.” = arbitrary units because powers are computed from a normalized signal and summed across periodogram bins (no physical calibration). (2) Full column names in the dataset are `eog_raw_{delta,theta,alpha,beta,gamma}_power`. (3) If desired, you can report each as a *fraction of total variance* by integrating with Δf and dividing by total power; the values then become unitless fractions (or %).

D.2 Hyperparameter Configurations test using Group K-Fold on Training Data

Table D.5: Hyperparameter Configurations for Support Vector Machine

Configuration	Hyperparameters	F1	Acc.
Baseline	C=1.0, gamma=scale	0.551	0.548
High Reg.	C=0.1, gamma=scale	0.532	0.533
Low Reg.	C=10.0, gamma=scale	0.538	0.546
Wide Infl.	C=1.0, gamma=0.01	0.550	0.547
Simple Model	C=0.01, gamma=0.01	0.479	0.484

Note: Bold row indicates the selected configuration. All configurations use kernel=rbf and class_weight=balanced. F1-scores represent macro F1.

Table D.6: Hyperparameter Configurations for Random Forest

Configuration	Hyperparameters	F1	Acc.
Baseline	n_estimators=100 max_depth=None	0.762	0.777
Shallow Forest	n_estimators=100 max_depth=10 min_samples_leaf=5	0.763	0.774
Balanced Forest	n_estimators=150 max_depth=15 min_samples_split=3 min_samples_leaf=2	0.766	0.779
Large Forest	n_estimators=500 max_depth=None min_samples_split=10	0.764	0.779
Small Forest	n_estimators=50 max_depth=6	0.738	0.750

Note: Bold row indicates the selected configuration. F1-scores represent macro F1.

Table D.7: Hyperparameter Configurations for XGBoost

Configuration	Hyperparameters	F1	Acc.
Baseline	max_depth=6 learning_rate=0.1 n_estimators=100 subsample=0.8 colsample_bytree=0.8	0.762	0.776
Shallow Boost	max_depth=4 learning_rate=0.1 n_estimators=100 subsample=0.8 colsample_bytree=0.9 min_child_weight=5	0.758	0.770
Balanced Boost	max_depth=8 learning_rate=0.08 n_estimators=150 subsample=0.75 colsample_bytree=0.75 min_child_weight=4	0.760	0.775
Large Boost	max_depth=10 learning_rate=0.05 n_estimators=500 subsample=0.7 colsample_bytree=0.7 min_child_weight=6	0.761	0.778
Small Boost	max_depth=4 learning_rate=0.1 n_estimators=50 subsample=0.9 colsample_bytree=0.9	0.748	0.760

Note: Bold row indicates the selected configuration. F1-scores represent macro F1.

D.3 Confusion Matrices for Selected SVM, RF, and XGB Configurations on Test Data

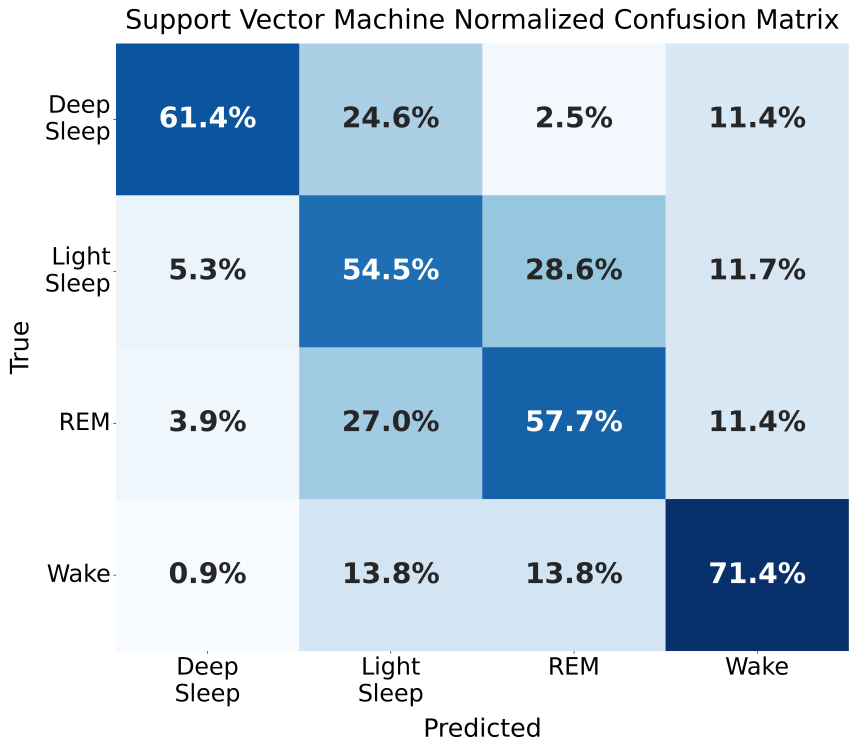


Figure D.1: Normalized Confusion Matrix for SVM Classifier on Test Data.

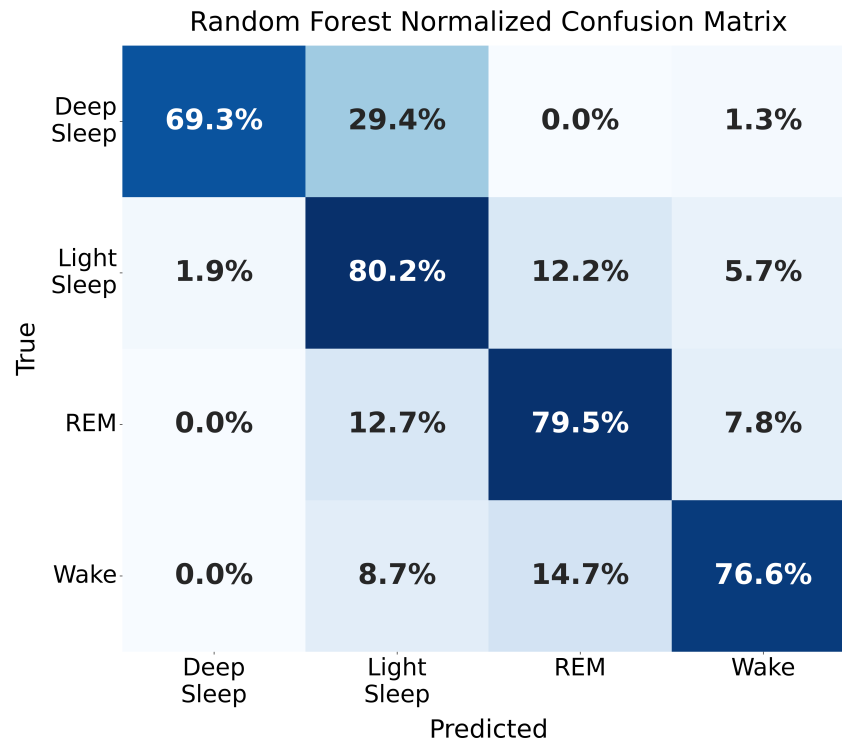


Figure D.2: Normalized Confusion Matrix for Random Forest Classifier on Test Data.

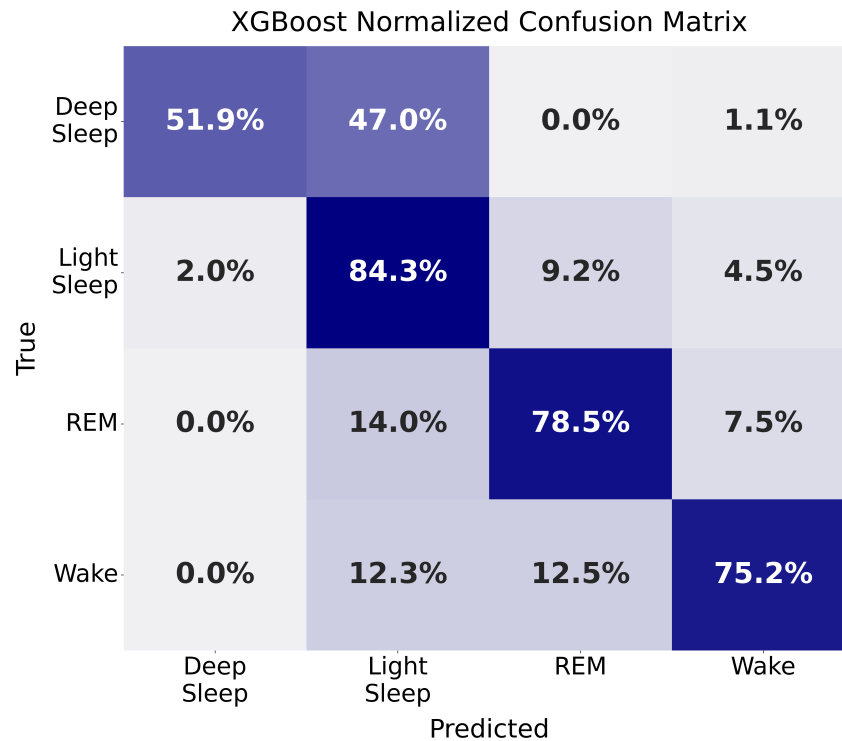


Figure D.3: Normalized Confusion Matrix for XGBoost Classifier on Test Data.

D.4 ROC Plots for SVM, RF, and XGB on Test Set

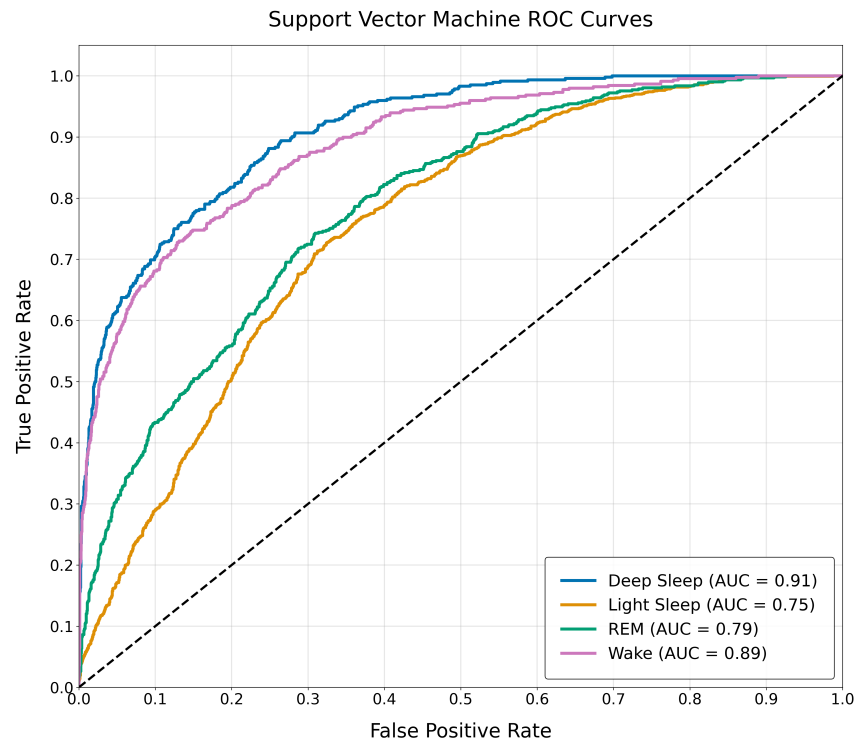


Figure D.4: SVM ROC Plot

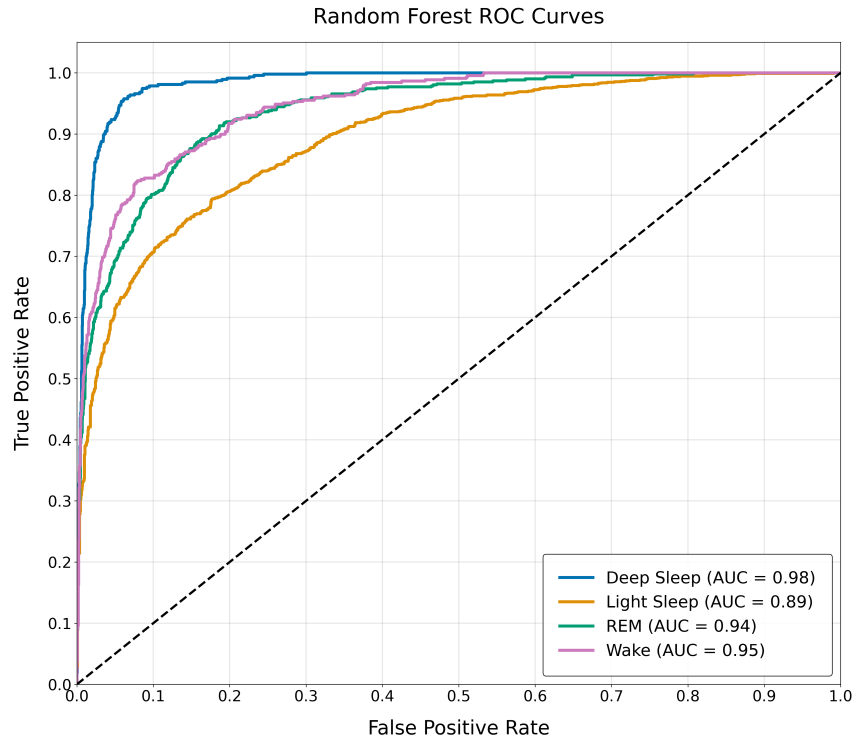


Figure D.5: Random Forest ROC Plot

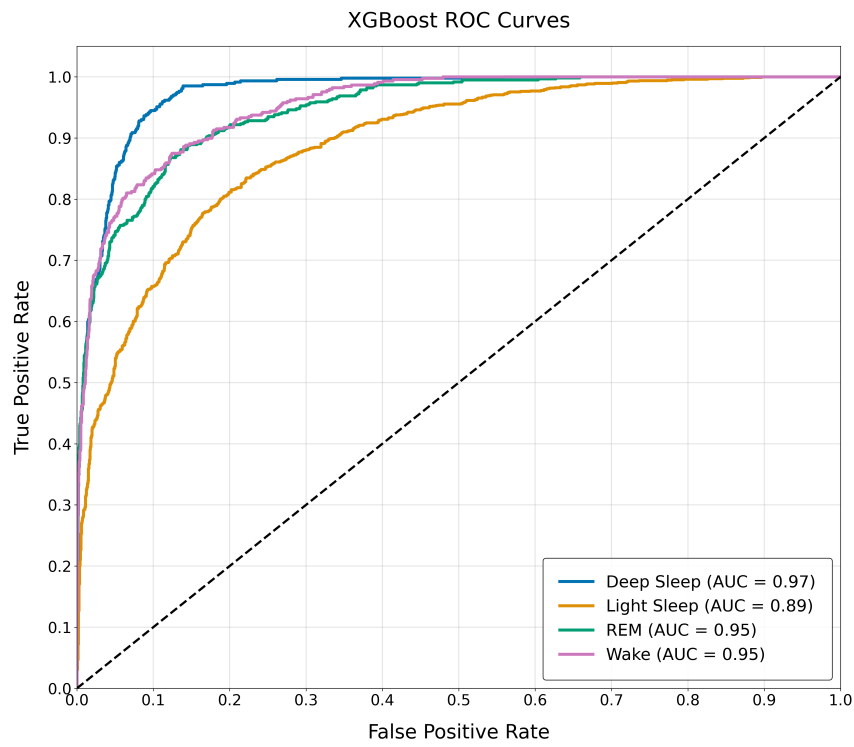


Figure D.6: XGBoost ROC Plot

D.5 Per-Class Metrics

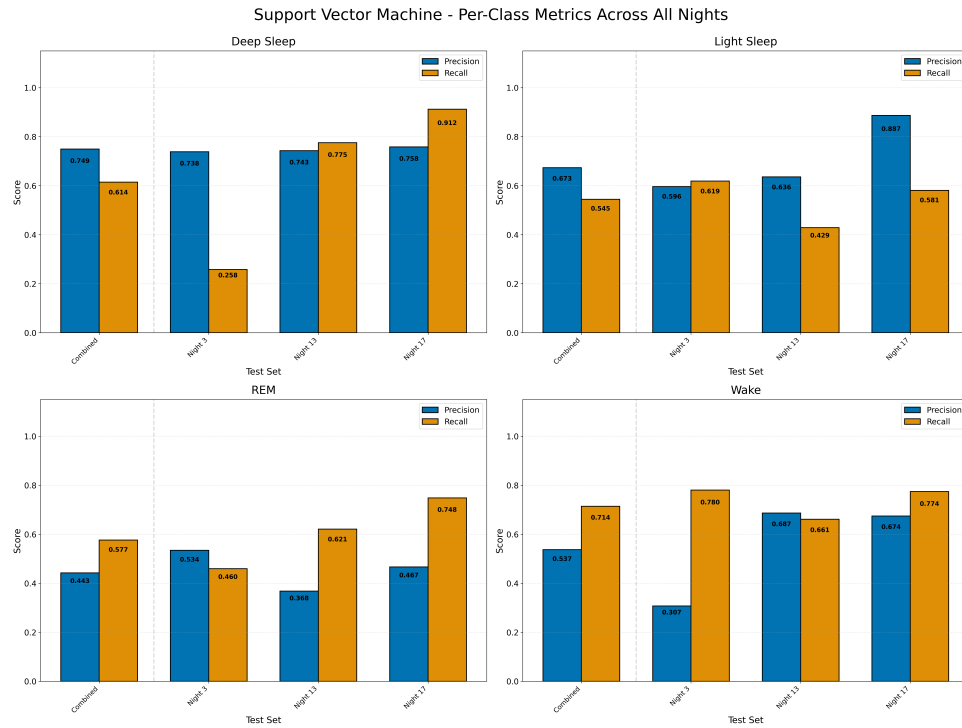


Figure D.7: SVM Per-Class Metrics

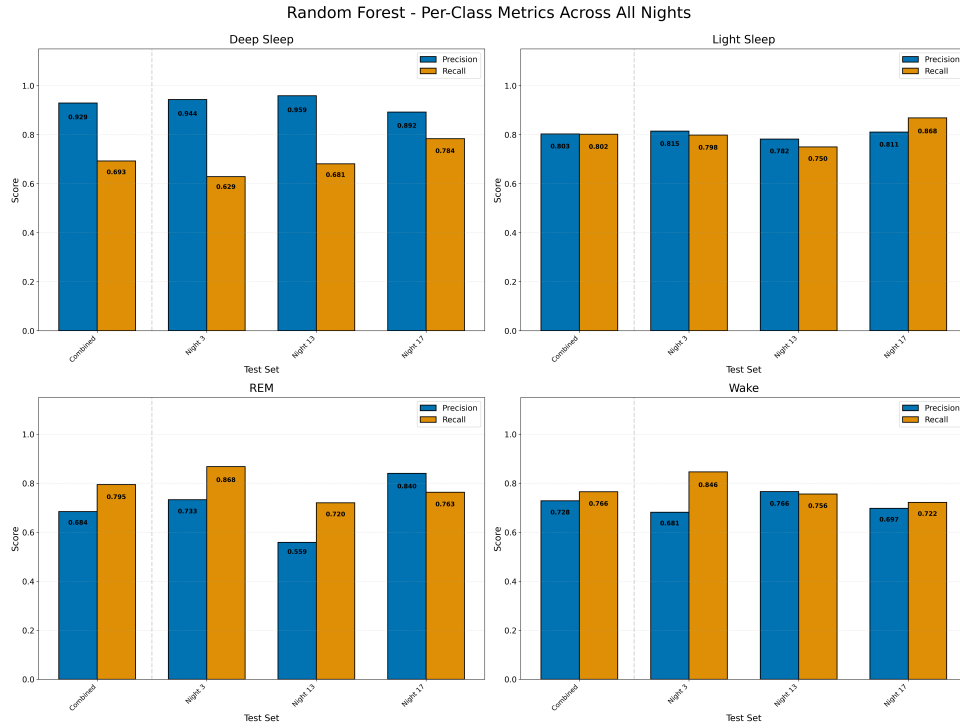


Figure D.8: Random Forest Per-Class Metrics

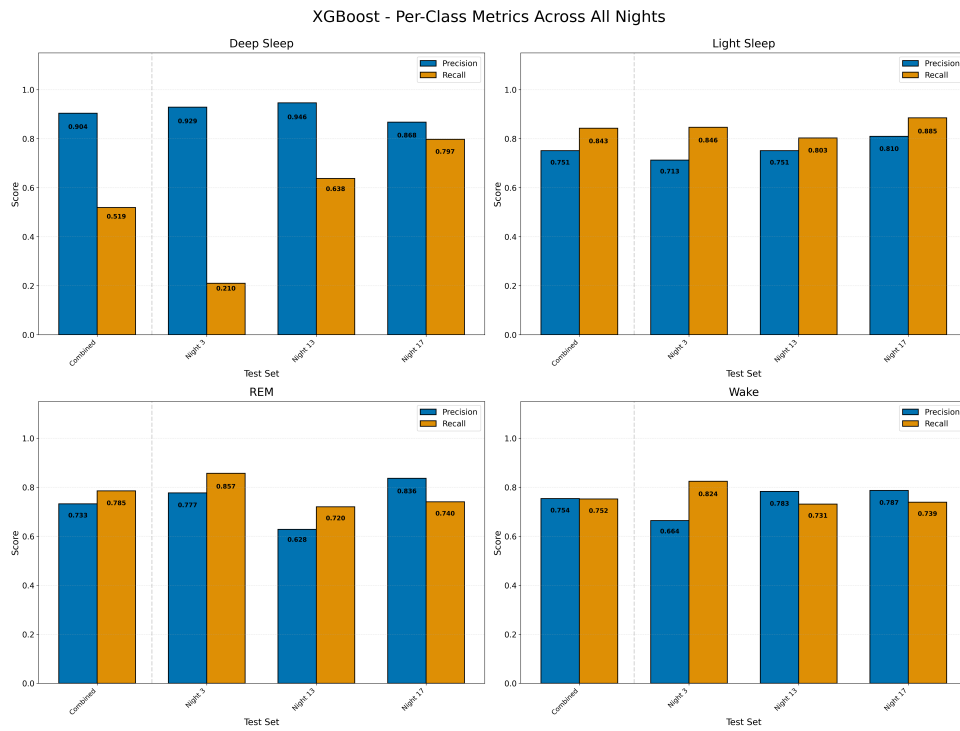


Figure D.9: XGBoost Per-Class Metrics

D.6 Feature Importance of Random Forest and XGBoost on Test Data

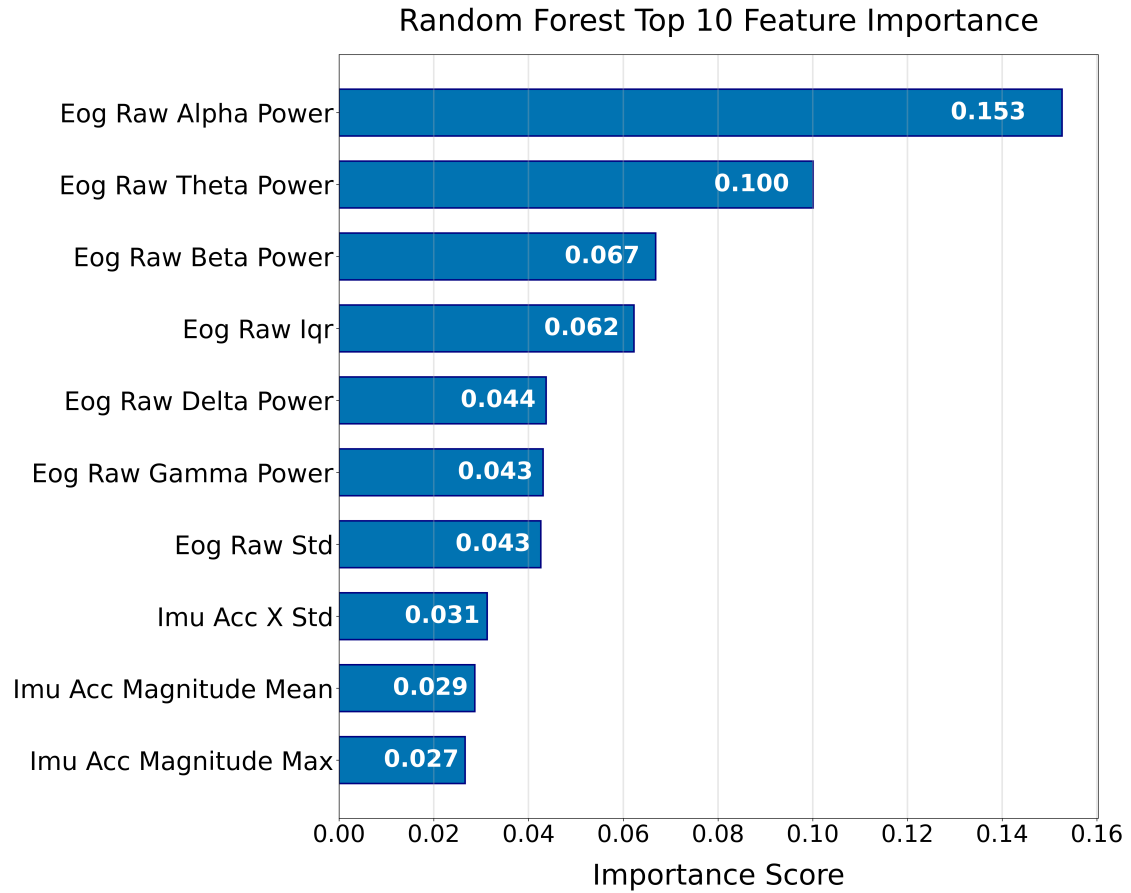


Figure D.10: Top 10 Feature Importance Rankings for Random Forest Classifier using Mean Decrease in Impurity.

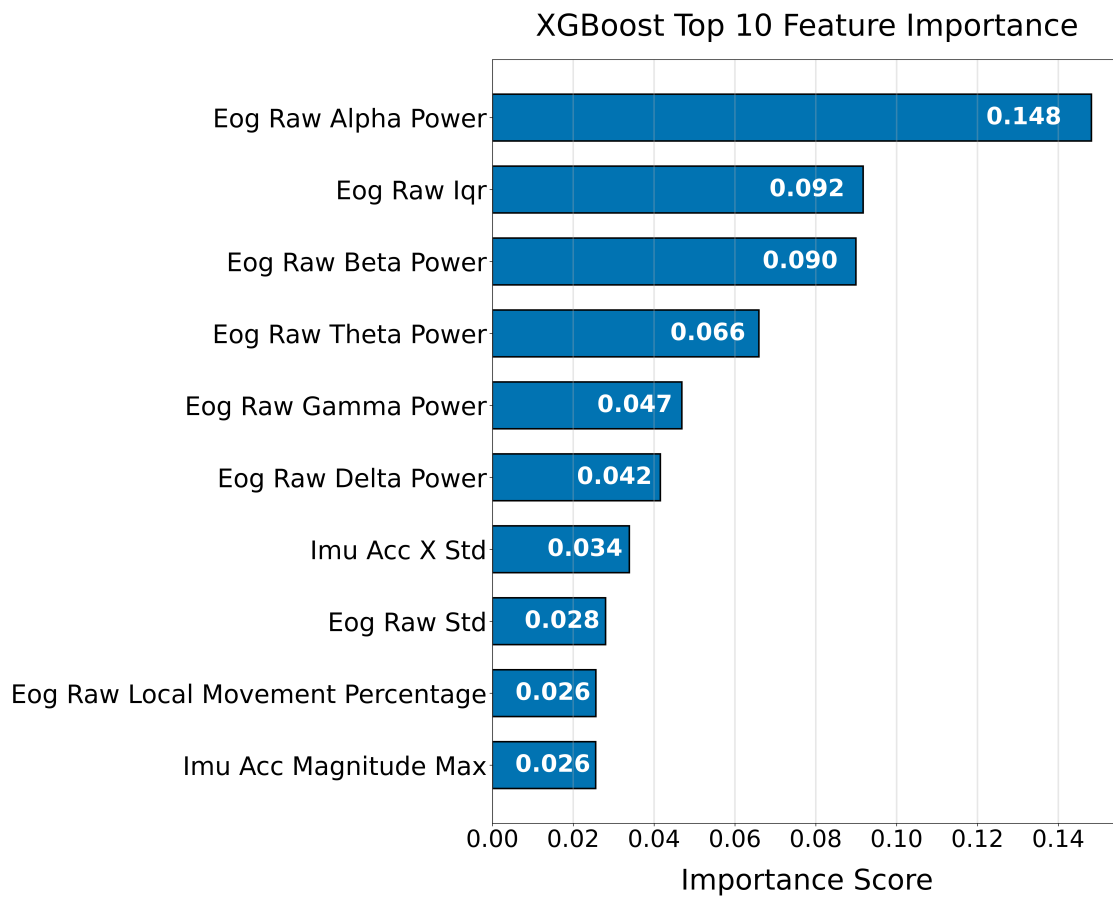


Figure D.11: Top 10 Feature Importance Rankings for XGBoost Classifier using Mean Decrease in Impurity.

D.7 Random Forest — Comprehensive Ablation Study Results

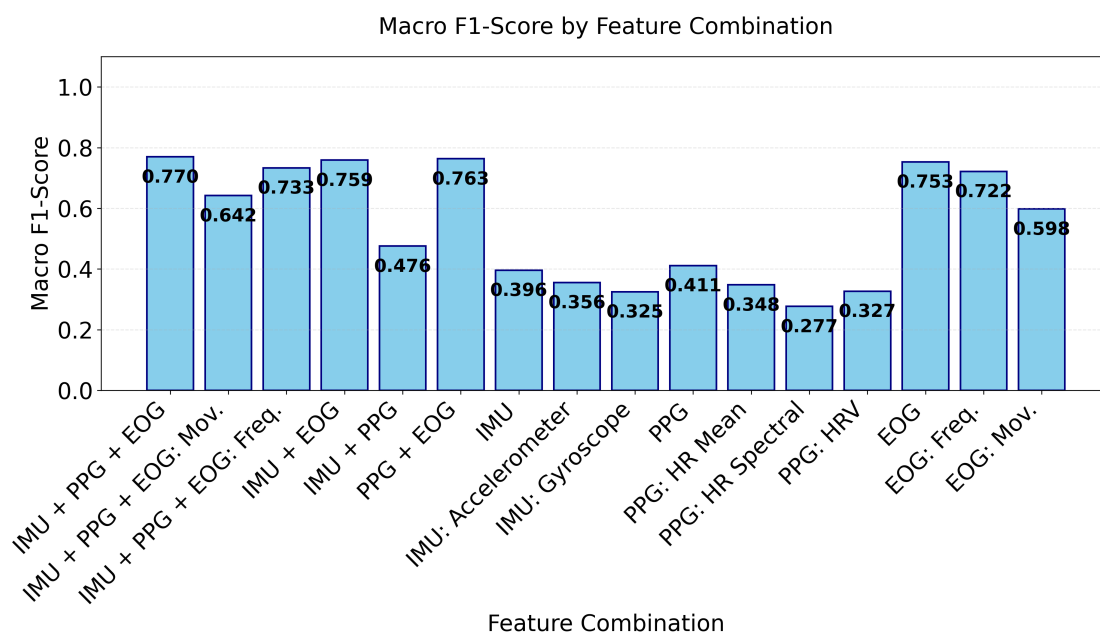


Figure D.12: F1 Scores of the Random Forest Ablation Study

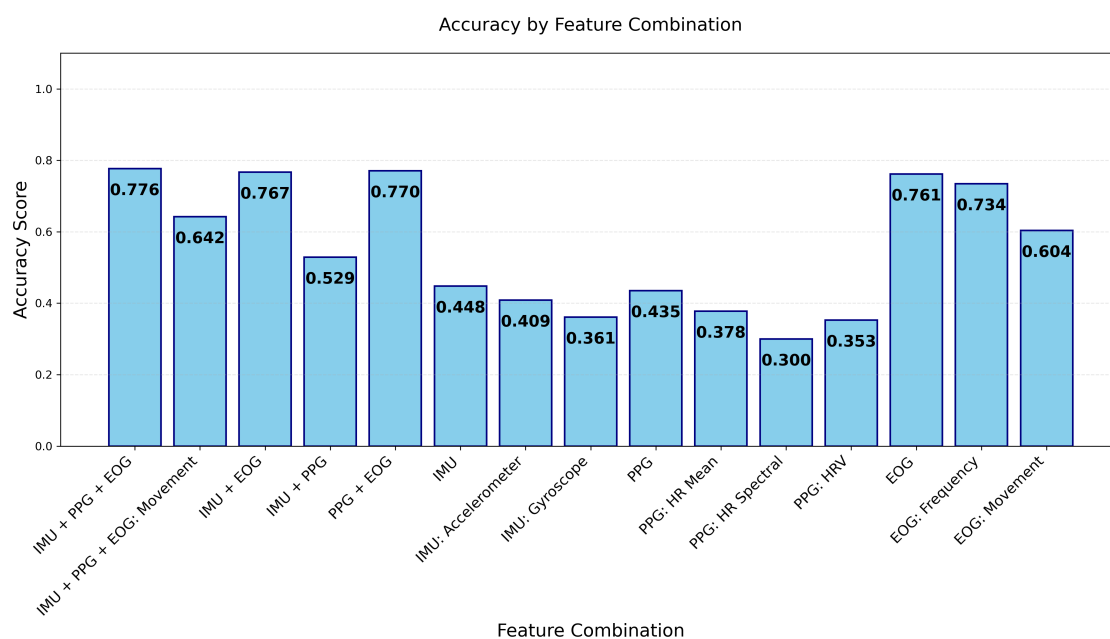


Figure D.13: Accuracy Scores of the Random Forest Ablation Study

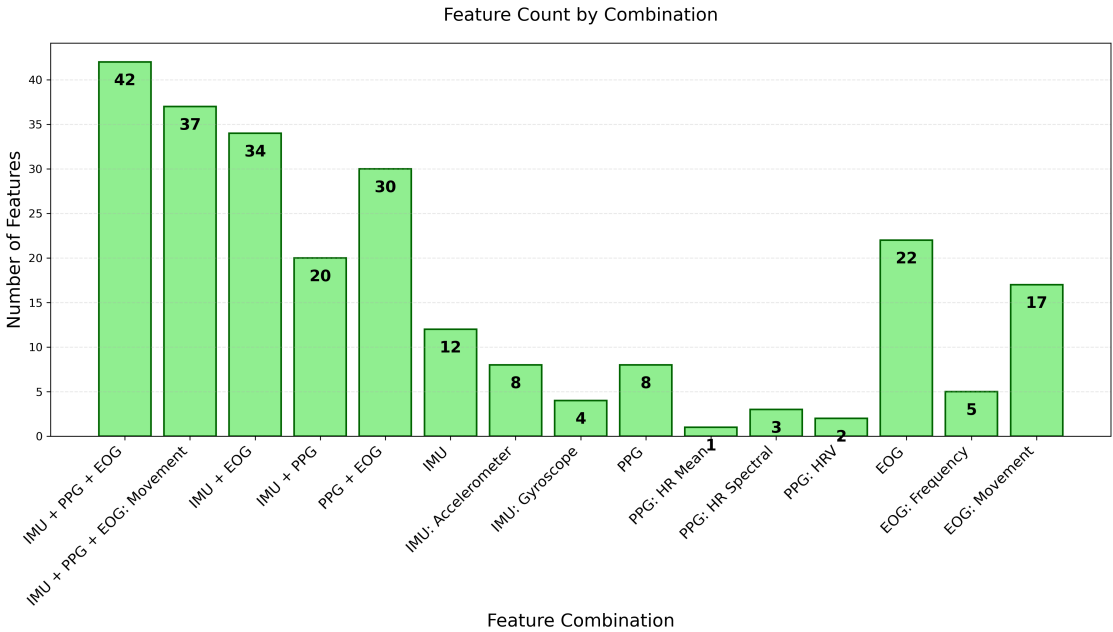


Figure D.14: Feature Counts of the Random Forest Ablation Study

D.8 Complete Ablation Investigation Metrics

Table D.8: Performance metrics for Random Forest classification across multiple sensor combinations.

Feature Combination	Features	Accuracy	Kappa	Precision	Recall	Macro F1-Score	Macro AUC	Weighted AUC
<i>Multi-sensor combinations</i>								
IMU + PPG + EOG	42	0.776	0.675	0.786	0.764	0.770	0.941	0.927
IMU + PPG + EOG: Mov.	37	0.642	0.476	0.663	0.630	0.642	0.858	0.834
IMU + PPG + EOG: Freq.	25	0.748	0.632	0.763	0.726	0.733	0.927	0.911
<i>Two-sensor combinations</i>								
IMU + EOG	34	0.767	0.661	0.775	0.753	0.759	0.937	0.923
IMU + PPG	20	0.529	0.307	0.491	0.491	0.476	0.732	0.716
PPG + EOG	30	0.770	0.670	0.770	0.766	0.763	0.939	0.926

Table D.9: Performance metrics for Random Forest classification across individual sensor feature combinations.

Feature Combination	Features	Accuracy	Kappa	Precision	Recall	Macro F1-Score	Macro AUC	Weighted AUC
<i>IMU features</i>								
IMU (all)	12	0.448	0.206	0.402	0.424	0.396	0.681	0.663
IMU: Accelerometer	8	0.409	0.144	0.365	0.369	0.356	0.651	0.640
IMU: Gyroscope	4	0.361	0.082	0.323	0.328	0.325	0.583	0.571
<i>PPG features</i>								
PPG (all)	8	0.435	0.196	0.408	0.420	0.411	0.673	0.649
PPG: HR Mean	1	0.378	0.126	0.438	0.354	0.348	0.589	0.601
PPG: HR Spectral	3	0.300	0.053	0.285	0.291	0.277	0.548	0.544
PPG: HRV	2	0.353	0.106	0.328	0.339	0.327	0.600	0.589
<i>EOG features</i>								
EOG (all)	22	0.761	0.658	0.754	0.758	0.753	0.934	0.923
EOG: Freq.	5	0.734	0.620	0.721	0.729	0.722	0.914	0.899
EOG: Mov.	17	0.604	0.422	0.605	0.593	0.598	0.799	0.776

Appendix — Proof-of-Concept Evaluation: Deep Learning Approaches

Appendix provides supplementary information for Chapter 6

E.1 Hyperparameter Exploration

Table E.1: EEG-Only CNN Performance Across Configurations and Cross-Validation Folds

	Fold 1	Fold 2	Fold 3	Fold 4	Mean
Medium Configuration					
Training Accuracy	0.824	0.748	0.839	0.812	0.806
Validation Accuracy	0.785	0.777	0.797	0.767	0.782
Validation F1	0.784	0.773	0.782	0.752	0.773
Validation κ	0.699	0.676	0.702	0.659	0.684
Overfitting Gap	—	—	—	—	0.024
Low Regularization					
Training Accuracy	0.953	0.787	0.838	0.818	0.849
Validation Accuracy	0.796	0.786	0.801	0.769	0.788
Validation F1	0.796	0.781	0.784	0.755	0.779
Validation κ	0.708	0.688	0.706	0.661	0.691
Overfitting Gap	—	—	—	—	0.061
Heavy Regularization					
Training Accuracy	0.791	0.759	0.804	0.788	0.786
Validation Accuracy	0.786	0.775	0.783	0.763	0.777
Validation F1	0.783	0.772	0.770	0.749	0.768
Validation κ	0.699	0.673	0.686	0.655	0.678
Overfitting Gap	—	—	—	—	0.009

Table E.2: PPG-Only CNN Performance Across Configurations and Cross-Validation Folds

	Fold 1	Fold 2	Fold 3	Fold 4	Mean
Medium Configuration					
Training Accuracy	0.689	0.672	0.743	0.567	0.668
Validation Accuracy	0.441	0.583	0.602	0.516	0.536
Validation F1	0.437	0.568	0.597	0.526	0.532
Validation κ	0.256	0.402	0.438	0.343	0.360
Overfitting Gap	—	—	—	—	0.132
Low Regularization					
Training Accuracy	0.719	0.735	0.754	0.863	0.768
Validation Accuracy	0.468	0.586	0.608	0.613	0.569
Validation F1	0.468	0.562	0.591	0.571	0.548
Validation κ	0.281	0.392	0.425	0.416	0.379
Overfitting Gap	—	—	—	—	0.199
Heavy Regularization					
Training Accuracy	0.435	0.625	0.555	0.647	0.565
Validation Accuracy	0.347	0.583	0.520	0.507	0.489
Validation F1	0.324	0.566	0.520	0.518	0.482
Validation κ	0.170	0.395	0.351	0.341	0.314
Overfitting Gap	—	—	—	—	0.076

E.2 Per Class Metrics on Withheld Data

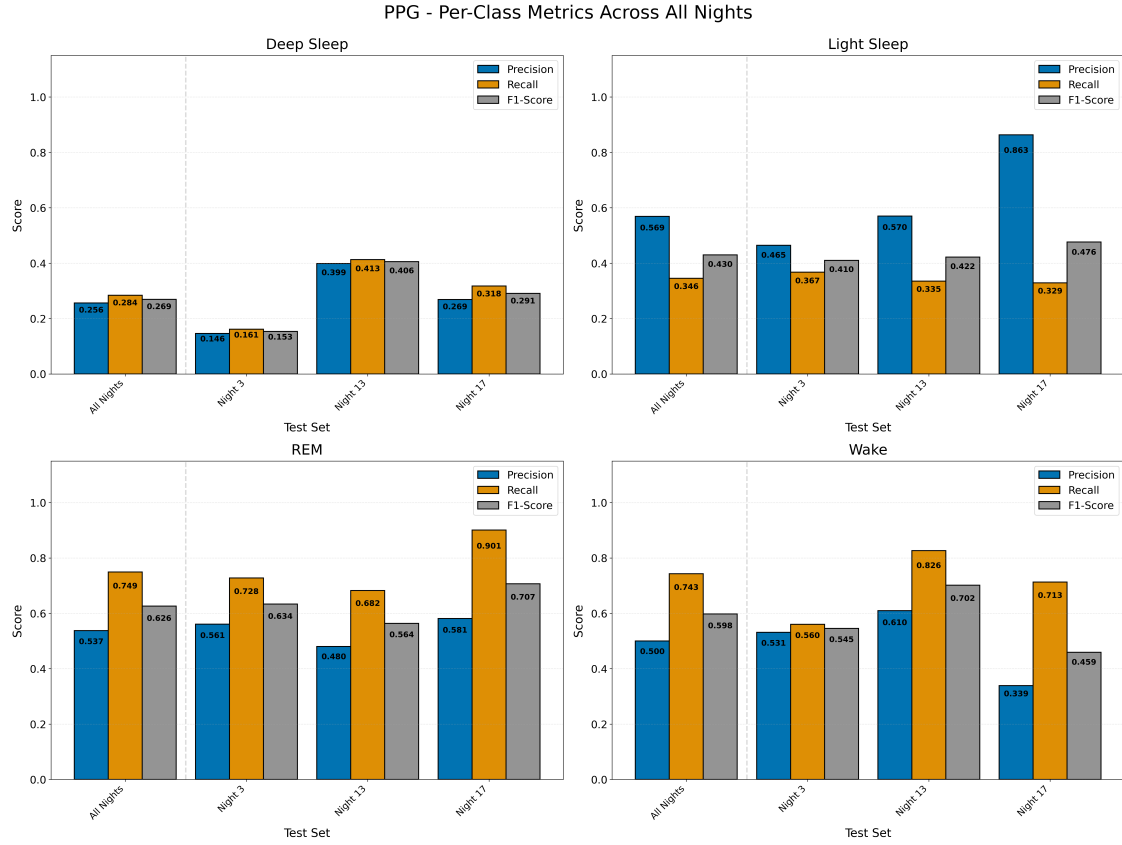


Figure E.1: Precision, Recall, and F1 for the PPG CNN Model

E.3 ROC Plots on Withheld Data

E.4 Normalized Confusion Matrices

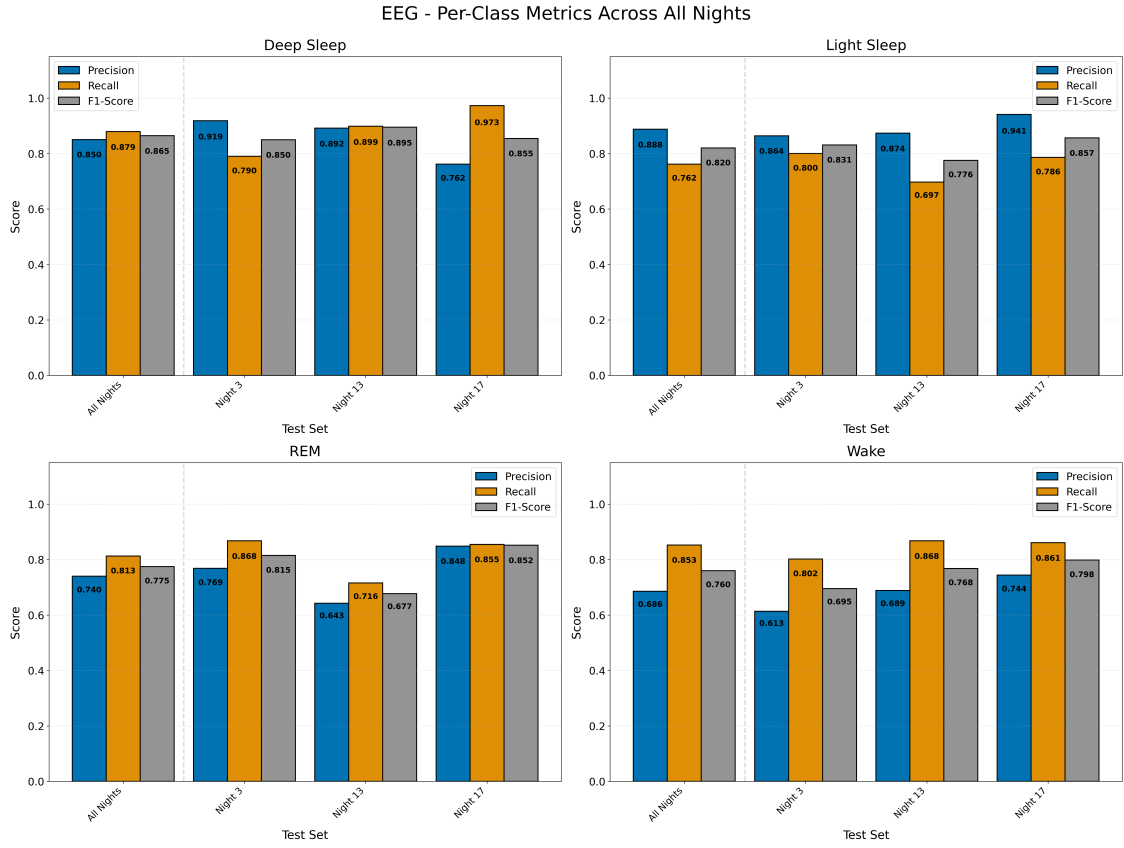


Figure E.2: EEG

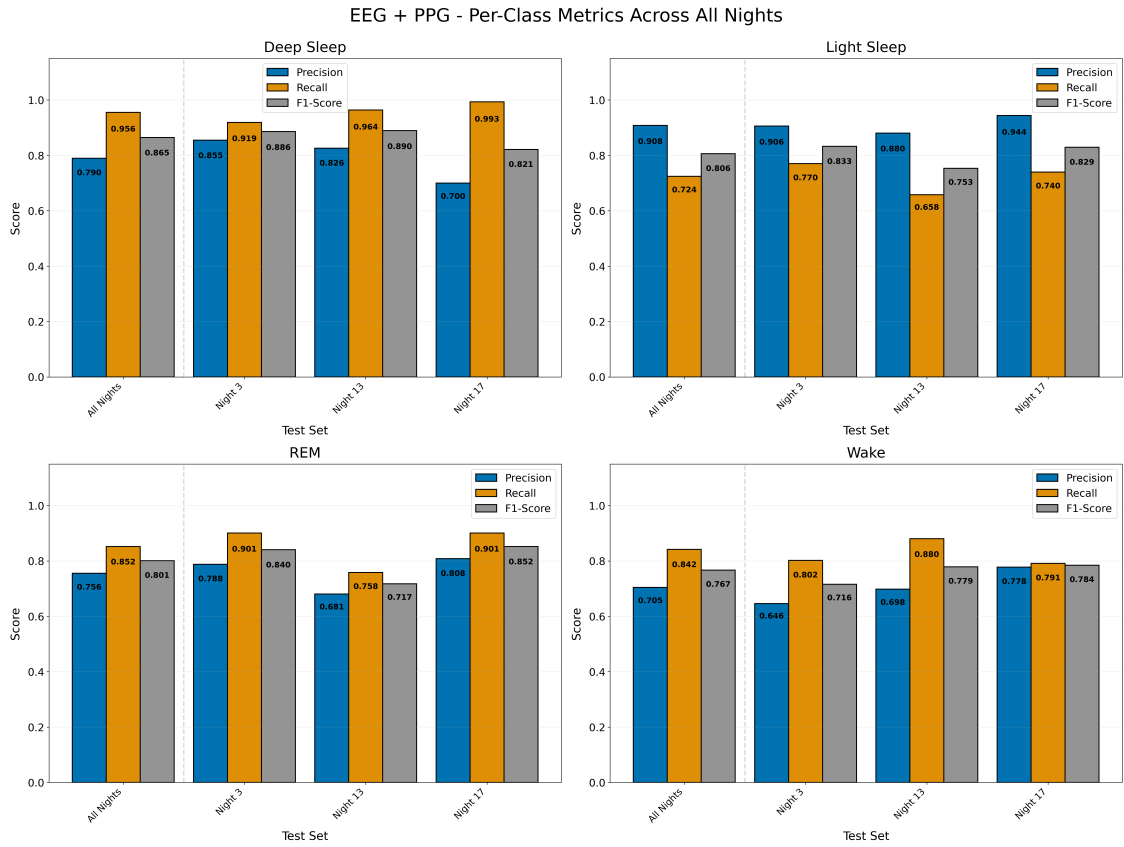


Figure E.3: EEG+PPG

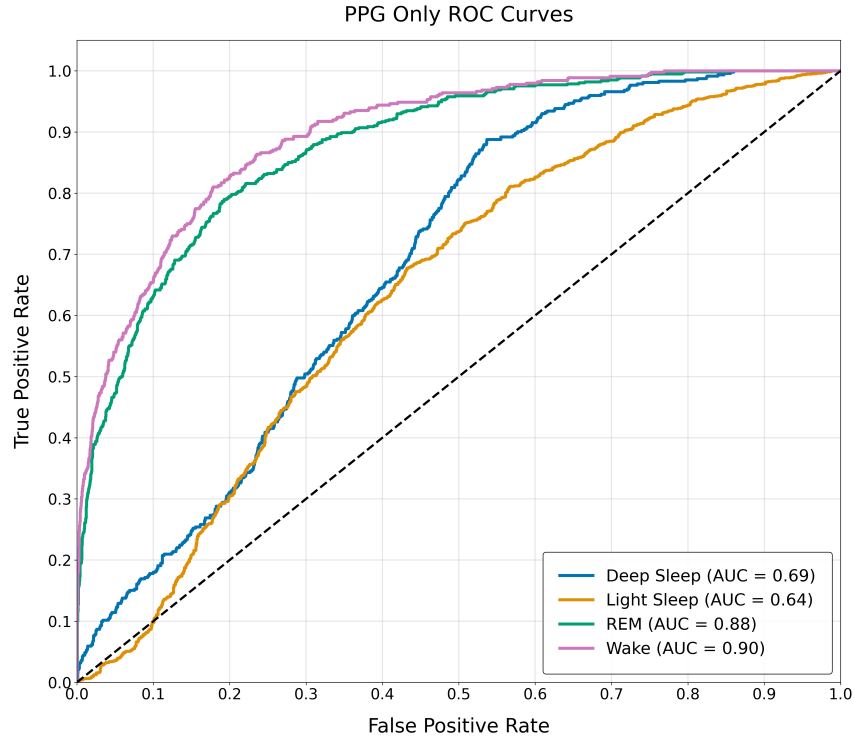


Figure E.4: ROC Plot for the PPG CNN Model

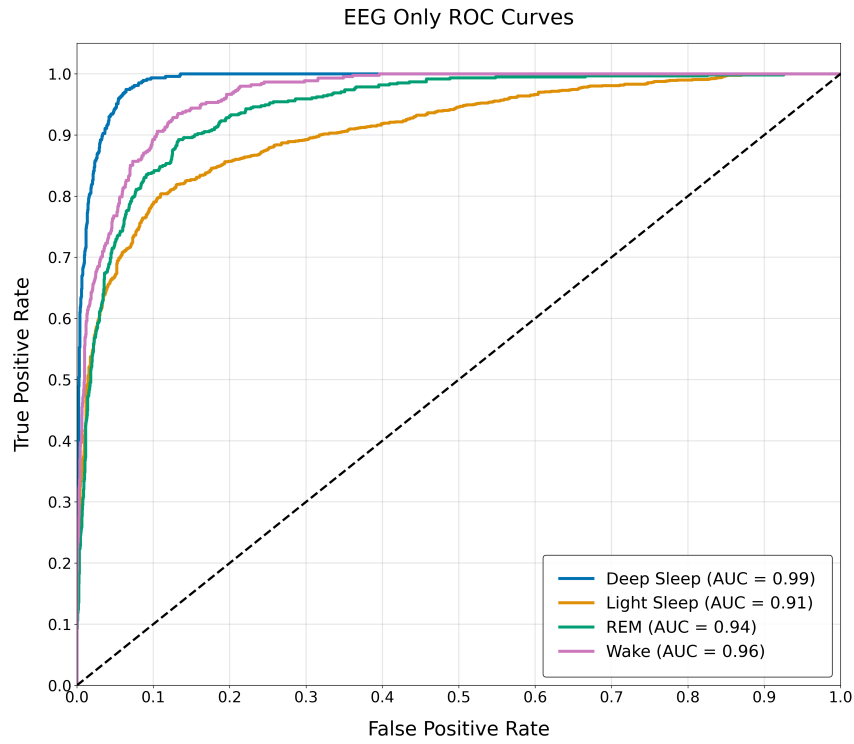


Figure E.5: ROC Plot for the EEG CNN Model

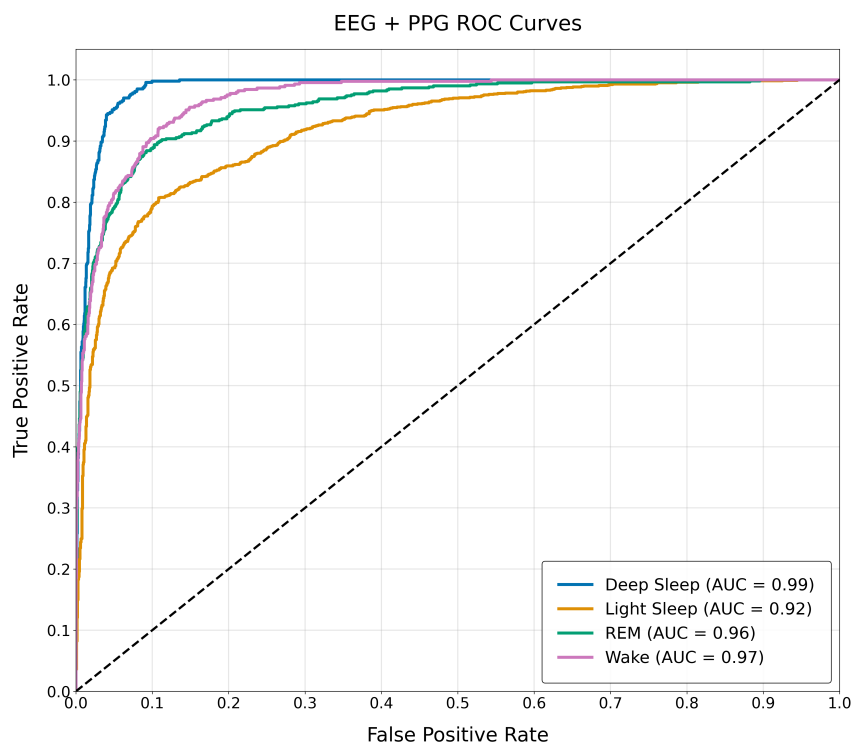


Figure E.6: ROC Plot for the EEG+PPG CNN Model

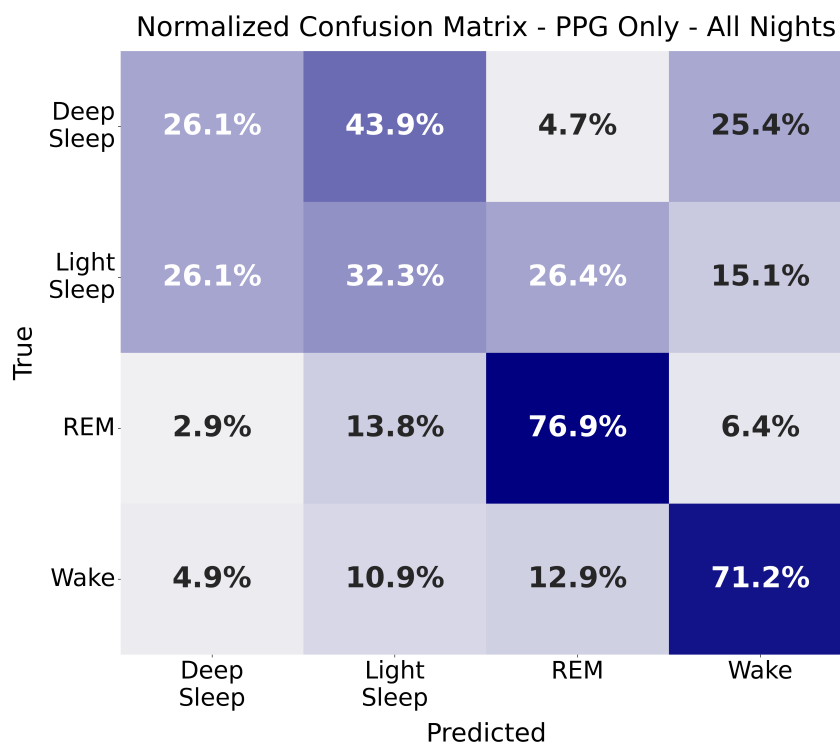


Figure E.7: Normalized Confusion Matric for PPG CNN Model

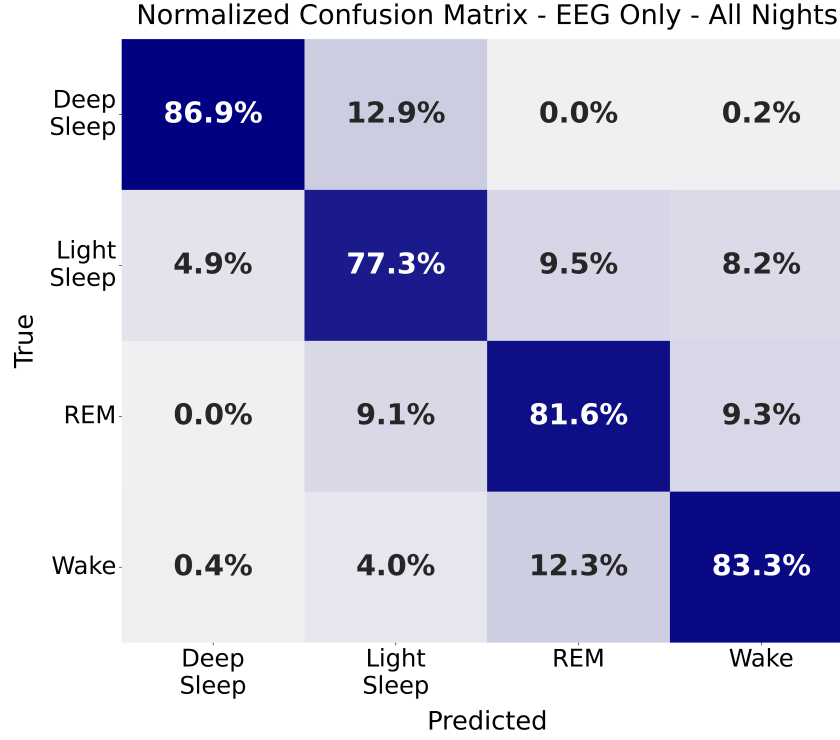


Figure E.8: Normalized Confusion Matric for EEG CNN Model

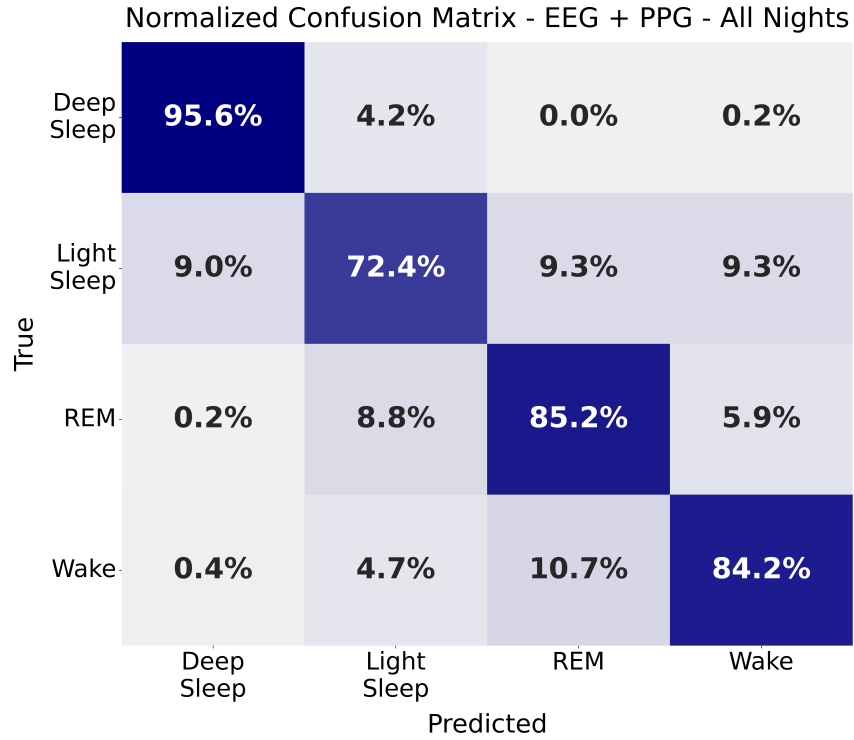


Figure E.9: Normalized Confusion Matric for EEG+PPG CNN Model

Appendix — Ethics

Appendix [F](#) provides supplementary information for Ethics Approval

F.1 Approval for OSSMM Evaluation on Awake Participants

MAYNOOTH UNIVERSITY RESEARCH ETHICS COMMITTEE

MAYNOOTH UNIVERSITY,
MAYNOOTH, CO. KILDARE, IRELAND



Dr Carol Barrett
Secretary to Maynooth University Research Ethics Committee

17 December 2024
Jonathan Vittorio Giordano

Maynooth University

Re: Application for ethical approval for a Project entitled: Open-Source Sleep Monitor and Modulator – Awake Device Assessment

Dear Jonathan,

The above project has been evaluated under Tier 2 process, expedited review and we would like to inform you that ethical approval has been granted.

Any deviations from the project details submitted to the ethics committee will require further evaluation. This ethical approval will expire on 30/10/2025.

Please note: all projects now require an end of project report which is attached. Please complete and upload the end of project report to your RIS ethics record after the project end date.

Kind Regards,

A handwritten signature in blue ink, appearing to read "Carol Barrett".

Dr Carol Barrett
Secretary,
Maynooth University Research Ethics Committee

Reference Number BRESC-2024-39757

F.2 Approval for OSSMM Technical Evaluation

MAYNOOTH UNIVERSITY RESEARCH ETHICS COMMITTEE
MAYNOOTH UNIVERSITY,
MAYNOOTH, CO. KILDARE, IRELAND



Dr Rafael De Andrade Moral
Chair of the Biomedical & Life Sciences Research Ethics Subcommittee
Maynooth University Research Ethics Committee

23 July 2025
Jonathan Giordano
Department of Computer Science
Maynooth University

Re: Application for ethical approval for a Project entitled: Open-Source Sleep Monitor and Modulator – Capability Assessment.

Dear Jonathan,

The Biomedical & Life Sciences Research Ethics Subcommittee has reviewed the ethical protocol for your project: Open-Source Sleep Monitor and Modulator – Capability Assessment, and we would like to inform you that ethical approval has been granted. Any deviations from the project details submitted to the ethics committee will require further evaluation. This ethical approval will expire on 30/12/2025. Please note: all projects now require an end of project report which is attached. Please complete and upload the end of project report to your RIS ethics record after the project end date.

Kind Regards,

A handwritten signature in black ink, appearing to read "Aileen O'Keefe".

On behalf of
Dr Rafael De Andrade Moral
Chair of the Biomedical & Life Sciences Research Ethics Subcommittee
Maynooth University Research Ethics Committee

Reference Number BSRESC-2025-40694

Bibliography

- AASM. *AASM Manual for the Scoring of Sleep and Associated Events*. American Academy of Sleep Medicine, Darien, IL, version 3 edition, 2023. (Cited on pages 4, 5, 6, 7, 77, 162, 182, and 184.)
- Adafruit Industries. Adafruit nRF52 arduino, 2017. URL https://github.com/adafruit/Adafruit_nRF52_Arduino. Arduino BSP and library for Nordic nRF52 BLE SoC. Last accessed: 2026-07-01. (Cited on page 237.)
- N. Ahmadi, P. Saleh, and C. M. Shapiro. The association between sleep disorders and depression: implications for treatment. In S. R. Pandi-Perumal and M. Kramer, editors, *Sleep and Mental Illness*, chapter 15, pages 165–170. Cambridge University Press, Cambridge, 2010. Section 3: Clinical science. (Cited on pages 11 and 58.)
- American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-5*. American Psychiatric Association, Washington, DC, 2013. (Cited on pages 56, 57, and 230.)
- Analog Devices. AD8232 (rev. d) single-lead, heart rate monitor front end data sheet, 2012. (Cited on page 73.)
- Arduino SA. Arduinoble, 2019. URL <https://github.com/arduino-libraries/ArduinoBLE>. Arduino library for Bluetooth Low Energy connectivity. Last accessed: 2026-07-01. (Cited on page 237.)
- P. J. Arnal, V. Thorey, E. Debellemanniere, M. E. Ballard, A. Bou Hernandez, A. Guillot, H. Jourde, M. Harris, M. Guillard, P. Van Beers, et al. The dreem headband compared to polysomnography for electroencephalographic signal acquisition and sleep staging. *Sleep*, 43(11):zsaa097, 2020. (Cited on pages 8, 9, and 162.)
- E. Aserinsky and N. Kleitman. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science*, 118(3062):273–274, 1953. (Cited on pages 4, 17, and 162.)

- A. T. Beck, R. A. Steer, and G. K. Brown. *Manual for the Beck Depression Inventory-II*. Psychological Corporation, San Antonio, TX, 1996. (Cited on pages 35 and 57.)
- M. E. Beitinger and S. Fulda. Long-term effects of antidepressants on sleep. In S. R. Pandi-Perumal and M. Kramer, editors, *Sleep and Mental Illness*, chapter 17, pages 183–201. Cambridge University Press, Cambridge, 2010. (Cited on pages 2, 4, 20, 46, 48, 49, 50, and 51.)
- P. M. Bet, J. G. Hugtenburg, B. W. Penninx, and W. J. Hoogendijk. Side effects of antidepressants during long-term use in a naturalistic setting. *European Neuropsychopharmacology*, 23(11):1443–1451, 2013. (Cited on page 2.)
- A. Bhandari, V. Khare, M. Trikha, and S. Anand. Wavelet based novel technique for signal conditioning of electro-oculogram signals. In *2006 Annual IEEE India Conference*, pages 1–6. IEEE, 2006. (Cited on page 73.)
- Bluetooth SIG. *Bluetooth Core Specification Version 6.1*. Bluetooth Special Interest Group, 2025a. Vol. 3, Part A: Logical Link Control and Adaptation Protocol Specification, p. 1134. (Cited on page 237.)
- Bluetooth SIG. *Bluetooth Core Specification Version 6.1*. Bluetooth Special Interest Group, 2025b. Vol. 6, Part B: Link Layer Specification, p. 3098. (Cited on page 237.)
- Bluetooth Special Interest Group. Bluetooth technology overview. URL <https://www.bluetooth.com/learn-about-bluetooth/tech-overview/>. Last accessed: 2024-03-22. (Cited on page 82.)
- A. Borbély and A. Wirz-Justice. Sleep, sleep deprivation and depression. *Hum Neurobiol*, 1(205):10, 1982. (Cited on pages 2 and 18.)
- A. A. Borbély et al. A two process model of sleep regulation. *Hum neurobiol*, 1(3): 195–204, 1982. (Cited on pages 3 and 6.)
- R. Boyce, S. D. Glasgow, S. Williams, and A. Adamantidis. Causal evidence for the role of REM sleep theta rhythm in contextual memory consolidation. *Science*, 352(6287):812–816, 2016. (Cited on page 17.)
- C. E. Carney, D. J. Buysse, S. Ancoli-Israel, J. D. Edinger, A. D. Krystal, K. L. Lichstein, and C. M. Morin. The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep*, 35(2):287–302, 2012. (Cited on pages 124 and 165.)

- R. Cartwright, E. Baehr, J. Kirkby, S. Pandi-Perumal, and J. Kabat. REM sleep reduction, mood regulation and remission in untreated depression. *Psychiatry Research*, 121(2):159–167, 2003. (Cited on pages 3, 11, 18, 19, 20, 34, 51, 56, 163, and 222.)
- N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16: 321–357, 2002. (Cited on page 168.)
- T. Chen and C. Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016. (Cited on page 168.)
- H. B. Cohen and W. C. Dement. Sleep: changes in threshold to electroconvulsive shock in rats after deprivation of "paradoxical" phase. *Science*, 150(3701):1318–1319, 1965. (Cited on page 3.)
- P. A. Constable, M. Bach, L. J. Frishman, B. G. Jeffrey, A. G. Robson, and I. S. for Clinical Electrophysiology of Vision. Iscev standard for clinical electro-oculography (2017 update). *Documenta Ophthalmologica*, 134(1):1–9, 2017. (Cited on page 73.)
- Core Electronics. Adafruit feather nRF52840 sense. <https://core-electronics.com.au/adafruit-feather-nrf52840-sense.html>, 2026. SKU: ADA4516. Last accessed: 2026-07-01. (Cited on page 89.)
- A. Coskun, S. Ozsen, S. Yucelbas, C. Yucelbas, G. Tezel, S. Kuccukturk, and S. Yosunkaya. Detection of REM in sleep EOG signals. *Indian Journal of Science and Technology*, 9(25):1–8, 2016. (Cited on pages 72 and 77.)
- CPAP Europa. Nox t3 portable sleep monitor, 2024a. URL <https://www.cpap.europa.com/product/nox-t3-portable-sleep-monitor/>. Last accessed: 2024-12-03. Price: €5,145.00. (Cited on pages 8 and 162.)
- CPAP Europa. Somnoscreen plus psg+aasm, 2024b. URL <https://www.cpap.europa.com/product/somnoscreen-plus-psg-aasm-demo/>. Last accessed: 2024-12-03. Price: €8,424.00. (Cited on pages 8 and 162.)
- C. A. Crişan, Z. Milhem, R. Stretea, I.-M. Țața, R. M. Cherecheș, and I. V. Micluția. A narrative review on REM sleep deprivation: A promising non-pharmaceutical alternative for treating endogenous depression. *Journal of Personalized Medicine*, 13(2):306, 2023. (Cited on page 19.)

- M. de Manaceine. Quelques observations expérimentales sur l'influence de l'insomnie absolue. *Arch Ital. Biol.*, 21:322–325, 1894. In French. (Cited on page 3.)
- W. Dement. The occurrence of low voltage, fast, electroencephalogram patterns during behavioral sleep in the cat. *Electroencephalography and clinical Neurophysiology*, 10(2):291–296, 1958. (Cited on page 17.)
- W. Dement. The effect of dream deprivation: The need for a certain amount of dreaming each night is suggested by recent experiments. *Science*, 131(3415):1705–1707, 1960. (Cited on page 17.)
- W. Dement. Studies on the function of rapid eye movement (paradoxical) sleep in human subjects. *Aspects Anatomo-Fonctionels de la Physiologie du Sommeil*, 1965a. (Cited on page 17.)
- W. Dement and C. Fisher. Iii. experimental interference with the sleep cycle. *Canadian Psychiatric Association Journal*, 8(6):400–405, 1963. (Cited on page 17.)
- W. C. Dement. Recent studies on the biological role of rapid eye movement sleep. *American Journal of Psychiatry*, 122(4):404–408, 1965b. (Cited on pages 3 and 18.)
- A. J. Derbyshire, B. Rempel, A. Forbes, and E. Lambert. The effects of anesthetics on action potentials in the cerebral cortex of the cat. *American Journal of Physiology-Legacy Content*, 116(3):577–596, 1936. (Cited on page 17.)
- Dreem / Business Wire. Dreem evolves its strategy, names quentin soulet de brugière as ceo and anne reiser to its board of directors. Business Wire (press release), Mar. 08 2021. URL <https://www.businesswire.com/news/home/20210308005681/en/Dreem-Evolves-its-Strategy-Names-Quentin-Soulet-de-Brugiere-as-CEO-and-Anne-Reiser-to-its-Board-of-Directors>. Last accessed: 2025-10-12. (Cited on page 9.)
- W. C. Duncan Jr, E. D. Ballard, and C. A. Zarate. Ketamine-induced glutamatergic mechanisms of sleep and wakefulness: insights for developing novel treatments for disturbed sleep and mood. In *Sleep-Wake Neurobiology and Pharmacology*, pages 337–358. Springer, 2017. (Cited on page 49.)
- Espressif Systems. ESP32 BLE library for arduino, 2019. URL <https://github.com/espressif/arduino-esp32/tree/master/libraries/BLE>. Bluetooth Low Energy library for the ESP32 Arduino core; originally developed by Neil Kolban. Last accessed: 2026-07-01. (Cited on page 237.)

- C. A. Everson, B. M. Bergmann, and A. Rechtschaffen. Sleep deprivation in the rat: Iii. total sleep deprivation. *Sleep*, 12(1):13–21, 1989. (Cited on page 3.)
- P. Feuillade, D. Pringuey, J. Belugou, P. Robert, and G. Darcourt. Trimipramine: acute and lasting effects on sleep in healthy and major depressive subjects. *Journal of affective disorders*, 24(3):135–145, 1992. (Cited on pages 32 and 52.)
- C. Fisher. Psychoanalytic implications of recent research on sleep and dreaming: Part i: Empirical findings. *Journal of the American Psychoanalytic Association*, 13(2):197–270, 1965. (Cited on page 17.)
- J. A. Floyd, J. J. Janisse, E. S. Jenuwine, and J. W. Ager. Changes in rem-sleep percentage over the adult lifespan. *Sleep*, 30(7):829–836, 2007. (Cited on pages 31 and 40.)
- B. Frauscher, N. von Ellenrieder, I. Dolezalova, S. Bouhadoun, J. Gotman, and L. Peter-Derex. Rapid eye movement sleep sawtooth waves are associated with widespread cortical activations. *Journal of Neuroscience*, 40(46):8900–8912, 2020. (Cited on page 124.)
- E. I. Fried and R. M. Nesse. Depression sum-scores don’t add up: why analyzing specific depression symptoms is essential. *BMC medicine*, 13(1):72, 2015a. (Cited on pages 38, 57, and 230.)
- E. I. Fried and R. M. Nesse. Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR* D study. *Journal of affective disorders*, 172:96–102, 2015b. (Cited on page 57.)
- S. Gibney and T. McCarthy. Profile of smartphone ownership and use in Ireland, May 2020. URL <https://assets.gov.ie/81401/e6c10ac5-e6b4-438e-b31d-30cd2b241f7c.pdf>. Research Services & Policy Unit, Research & Development and Health Analytics Division, Department of Health, Government of Ireland. Last accessed: 2026-07-01. (Cited on page 66.)
- H. Giedke and F. Schwärzler. Therapeutic use of sleep deprivation in depression. *Sleep medicine reviews*, 6(5):361–377, 2002. (Cited on pages 2, 18, 19, 24, 32, and 41.)
- C. M. Gillan, M. Kosinski, R. Whelan, E. A. Phelps, and N. D. Daw. Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *elife*, 5:e11305, 2016. (Cited on pages 57 and 67.)

- R. W. Glosemeyer, S. Diekelmann, W. Cassel, K. Kesper, U. Koehler, S. Westermann, A. Steffen, S. Borgwardt, I. Wilhelm, L. Müller-Pinzler, et al. Selective suppression of rapid eye movement sleep increases next-day negative affect and amygdala responses to social exclusion. *Scientific reports*, 10(1):17325, 2020. (Cited on page 55.)
- C. Gottesmann. The golden age of rapid eye movement sleep discoveries. 1. lucretius—1964. *Progress in neurobiology*, 65(3):211–287, 2001. (Cited on page 17.)
- M. Grözinger. Personal communication regarding omission of an early rem reduction result from the later publication. Personal communication, 09 2023. Dr.Grözinger stated that it is not known, even to himself, why the NREM vs REM result published in an earlier abstract was not included in the later main paper. (Cited on page 31.)
- M. Grözinger, C. Wolf, T. Uhl, C. Schäffner, and J. Röscke. Online detection of rem sleep based on the comprehensive evaluation of short adjacent eeg segments by artificial neural networks. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 21(6):951–963, 1997. (Cited on pages 30, 33, and 221.)
- M. Grözinger, P. Kögel, T. Uhl, and J. Röscke. Does an intracyclic REM pressure induce antidepressive effects? *Sleep Res Online*, 2(Suppl 1):532, 1999. (Cited on pages 8 and 31.)
- M. Grözinger, P. Kögel, and J. Röscke. Effects of REM sleep awakenings and related waking paradigms on the ultradian sleep cycle and the symptoms in depression. *Journal of psychiatric research*, 36(5):299–308, 2002. (Cited on pages 3, 8, 10, 11, 12, 18, 19, 20, 29, 33, 34, 41, 51, 56, 162, 163, and 222.)
- A. Guadalupi. Arduino Nano 33 IoT Schematics. Technical schematic, Arduino SA, September 2022. URL <https://docs.arduino.cc/resources/schematics/ABX00027-schematics.pdf>. Reference Design Schematic, Version 5.3. Last accessed: 2025-06-25. (Cited on pages 83 and 235.)
- h2zero and contributors. Nimble-arduino, 2020. URL <https://h2zero.github.io/NimBLE-Arduino/>. Bluetooth Low Energy library for Arduino based on Apache NimBLE stack. Last accessed: 2026-07-01. (Cited on pages 85 and 237.)
- R. Haro and R. Drucker-Colín. Effects of long-term administration of nicotine and fluoxetine on sleep in depressed patients. *Archives of Medical Research*, 35(6):499–506, 2004. (Cited on page 49.)

- M. P. Hengartner. Methodological flaws, conflicts of interest, and scientific fallacies: Implications for the evaluation of antidepressants' efficacy and harm. *Frontiers in psychiatry*, 8:275, 2017. (Cited on page 2.)
- E. Holsboer-Trachsler, U. Hemmeter, M. Hatzinger, E. Seifritz, U. Gerhard, and V. Hobi. Sleep deprivation and bright light as potential augmenters of antidepressant drug treatment—neurobiological and psychometric assessment of course. *Journal of psychiatric research*, 28(4):381–399, 1994. (Cited on page 32.)
- B. Hu, C. Liu, T. Mou, F. Luo, T. Lv, C. Qian, J. Zhang, M. Ye, and Z. Liu. Meta-analysis of sleep deprivation effects on patients with depression. *Frontiers in Psychiatry*, 12:783091, 2021. (Cited on pages 2 and 18.)
- Hypnodyne Corp. ZMax – EEG sleep headband, 2024. URL <https://hypnodynecorp.com/>. Last accessed: 2025-08-30. (Cited on page 11.)
- IEEE. IEEE Code of Ethics, June 2020. URL <https://www.ieee.org/content/dam/ieee-org/ieee/web/org/about/corporate/ieee-code-of-ethics.pdf>. Adopted by the IEEE Board of Directors, June 2020. Last accessed: 2025-10-02. (Cited on page 165.)
- T. Insel, B. Cuthbert, M. Garvey, R. Heinssen, D. S. Pine, K. Quinn, C. Sanislow, and P. Wang. Research domain criteria (rdoc): toward a new classification framework for research on mental disorders. *American Journal of psychiatry*, 167(7):748–751, 2010. (Cited on page 58.)
- S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr, 2015. (Cited on page 195.)
- Jiangsu Kerbio. In vitro cytotoxicity test. Technical Report SSMT-R-2024-02665-01A, Jiangsu Kerbio Medical Technology Group Co., Ltd., Changzhou, Jiangsu, China, June 2024a. Biocompatibility test report for Siraya Tech Flex TPU filament, conducted according to ISO 10993-5:2009 standard. (Cited on pages 121, 142, and 166.)
- Jiangsu Kerbio. Skin irritation test. Technical Report SSMT-R-2024-02665-02A, Jiangsu Kerbio Medical Technology Group Co., Ltd., Changzhou, Jiangsu, China, June 2024b. Biocompatibility test report for Siraya Tech Flex TPU filament, conducted according to ISO 10993-23:2021 standard. (Cited on pages 121, 142, and 166.)

- Jiangsu Kerbio. Skin sensitization test. Technical Report SSMT-R-2024-02665-03A, Jiangsu Kerbio Medical Technology Group Co., Ltd., Changzhou, Jiangsu, China, June 2024c. Biocompatibility test report for Siraya Tech Flex TPU filament , conducted according to ISO 10993-10:2021 standard. (Cited on pages [121](#), [142](#), and [166](#).)
- V. K. Kapur, D. H. Auckley, S. Chowdhuri, D. C. Kuhlmann, R. Mehra, K. Ramar, and C. G. Harrod. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an american academy of sleep medicine clinical practice guideline. *Journal of clinical sleep medicine*, 13(3):479–504, 2017. (Cited on page [7](#).)
- N. Ketz, A. P. Jones, N. B. Bryant, V. P. Clark, and P. K. Pilly. Closed-loop slow-wave tacs improves sleep-dependent long-term memory generalization by modulating endogenous oscillations. *Journal of Neuroscience*, 38(33):7314–7326, 2018. (Cited on pages [10](#) and [163](#).)
- M. Kheirkhah, W. C. Duncan Jr, Q. Yuan, P. R. Wang, H. Jamalabadi, L. Leistritz, M. Walter, D. Goldman, C. A. Zarate Jr, and N. S. Hejazi. Rem density predicts rapid antidepressant response to ketamine in individuals with treatment-resistant depression. *Neuropsychopharmacology*, pages 1–6, 2025. (Cited on page [49](#).)
- R. D. Kim, V. K. Kapur, J. Redline-Bruch, M. Rueschman, D. H. Auckley, R. M. Benca, N. R. Foldvary-Schafer, C. Iber, P. C. Zee, C. L. Rosen, et al. An economic evaluation of home versus laboratory-based diagnosis of obstructive sleep apnea. *Sleep*, 38(7):1027–1037, 2015. (Cited on pages [8](#) and [162](#).)
- J. D. Knoke. Nonparametric analysis of covariance for comparing change in randomized studies with baseline values subject to error. *Biometrics*, pages 523–533, 1991. (Cited on page [25](#).)
- A. Koushik, J. Amores, and P. Maes. Real-time smartphone-based sleep staging using 1-channel eeg. In *2019 IEEE 16th international conference on wearable and implantable body sensor networks (BSN)*, pages 1–4. IEEE, 2019. (Cited on page [8](#).)
- A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012. (Cited on page [189](#).)
- M. H. Kryger, T. Roth, and W. C. Dement. *Principles and Practice of Sleep Medicine E-Book: Expert Consult-Online and Print*. Elsevier Health Sciences, 2010. (Cited on page [82](#).)

- M. H. Kryger, T. Roth, and W. C. Dement. *Principles and practice of sleep medicine*. Elsevier, Philadelphia, PA, sixth edition. edition, 2017. ISBN 9780323242882. (Cited on pages 3, 4, 5, 7, and 73.)
- J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174, 1977. (Cited on page 169.)
- E. C. Landsness, M. R. Goldstein, M. J. Peterson, G. Tononi, and R. M. Benca. Antidepressant effects of selective slow wave sleep deprivation in major depression: a high-density eeg investigation. *Journal of psychiatric research*, 45(8):1019–1026, 2011. (Cited on pages 3, 10, 12, 18, 20, 42, 52, and 163.)
- C. J. Lauer, W. Schreiber, F. Holsboer, and J.-C. Krieg. In quest of identifying vulnerability markers for psychiatric disorders by all-night polysomnography. *Archives of General Psychiatry*, 52(2):145–153, 1995. (Cited on page 31.)
- E. Leibenluft, D. E. Moul, P. J. Schwartz, P. A. Madden, and T. A. Wehr. A clinical trial of sleep deprivation in combination with antidepressant medication. *Psychiatry Research*, 46(3):213–227, 1993. (Cited on page 32.)
- N. Letov and Y. F. Zhao. A geometric modelling framework to support the design of heterogeneous lattice structures with non-linearly varying geometry. *Journal of Computational Design and Engineering*, 9(5):1565–1584, 2022. (Cited on page 143.)
- M. Levenson and C. Holland. Antidepressants and suicidality in adults: Statistical evaluation. *Presentation at Psychopharmacologic Drugs Advisory Committee*, 2006. (Cited on page 2.)
- C. Li, Y. Qi, X. Ding, J. Zhao, T. Sang, and M. Lee. A deep learning method approach for sleep stage classification with eeg spectrogram. *International journal of environmental research and public health*, 19(10):6322, 2022. (Cited on pages xviii, 5, 8, 189, 202, and 203.)
- M. M. Lia. The advantages of substitution of REM sleep stages with waking episodes to perform REM sleep reduction. *EC Neurology*, 11:400–407, 2019. (Cited on page 19.)
- A. L. Loomis, E. N. Harvey, and G. A. Hobart. Cerebral states during sleep, as studied by human brain potentials. *Journal of experimental psychology*, 21(2):127, 1937. (Cited on pages 4 and 162.)
- M5Stack. M5StickC PLUS ESP32-PICO Mini IoT Development Kit, 2025. URL <https://shop.m5stack.com/products/m5stickc-plus-esp32-pico-mini-i>

- [ot-development-kit](#). ESP32-based IoT development board with Wi-Fi, 1.14-inch LCD, 6-axis IMU, and 120mAh battery. Last accessed: 2025-06-25. (Cited on pages [xvi](#), [83](#), and [84](#).)
- R. Maker. Garmin forerunner 245 music gps watch in-depth review, 04 2019. URL <https://www.dcrainmaker.com/2019/04/garmin-forerunner-music-review.html>. Launch price: Forerunner 245 Music \$349; non-Music \$299. Last accessed: 2025-10-12. (Cited on page [9](#).)
- M. Martin et al. Computer and Internet use in the United States: 2018. *American Community Survey Reports*, 2021. (Cited on page [66](#).)
- P. McNamara. *The Neuroscience of Sleep and Dreams*. Cambridge Fundamentals of Neuroscience in Psychology. Cambridge University Press, 2019. ISBN 9781107171107; 1107171105. (Cited on pages [3](#), [4](#), [5](#), and [6](#).)
- J. J. A. Mendes Junior, D. P. Campos, L. C. d. A. V. D. Biassio, P. C. Passos, P. B. Júnior, A. E. Lazzaletti, and E. Krueger. Ad8232 to biopotentials sensors: Open source project and benchmark. *Electronics*, 12(4):833, 2023. (Cited on page [114](#).)
- N. Meyer, R. Lok, C. Schmidt, S. D. Kyle, C. A. McClung, C. Cajochen, F. A. Scheer, M. W. Jones, and S. L. Chellappa. The sleep–circadian interface: A window into mental disorders. *Proceedings of the National Academy of Sciences*, 121(9):e2214756121, 2024. (Cited on page [58](#).)
- D. J. Miller, C. Sargent, and G. D. Roach. A validation of six wearable devices for estimating sleep, heart rate and heart rate variability in healthy adults. *Sensors*, 22(16):6317, 2022. (Cited on pages [8](#) and [9](#).)
- J. D. Minkel, A. D. Krystal, and R. M. Benca. Unipolar major depression. In *Principles and Practice of Sleep Medicine*, chapter 137, pages 1352–1362. Elsevier, 6 edition, 2017. Part II, Section 17: Psychiatric Disorders. (Cited on pages [2](#), [11](#), and [18](#).)
- J. Moncrieff, R. E. Cooper, T. Stockmann, S. Amendola, M. P. Hengartner, and M. A. Horowitz. The serotonin theory of depression: a systematic umbrella review of the evidence. *Molecular psychiatry*, pages 1–14, 2022. (Cited on page [51](#).)
- U. Moretto and L. Palagini. Sleep in major depression. In *Handbook of Sleep Research*, volume 30 of *Handbook of Behavioral Neuroscience*, chapter 46, pages 693–706. Elsevier B.V., Amsterdam, 2019. doi: 10.1016/B978-0-12-813743-7.00046-3. Part H: Sleep in Aging and Disease. (Cited on pages [2](#), [11](#), [18](#), and [58](#).)

- S. A. Mota-Rolim, A. Pavlou, G. C. Nascimento, J. Fontenele-Araujo, and S. Ribeiro. Portable devices to induce lucid dreams—are they reliable? *Frontiers in neuroscience*, 13:428, 2019. (Cited on page 11.)
- S. Mousavi, F. Afghah, and U. R. Acharya. Sleepegnet: Automated sleep stage scoring with sequence to sequence deep learning approach. *PloS one*, 14(5):e0216456, 2019. (Cited on pages 8 and 189.)
- M. Murphy, B. A. Riedner, R. Huber, M. Massimini, F. Ferrarelli, and G. Tononi. Source modeling sleep slow waves. *Proceedings of the National Academy of Sciences*, 106(5):1608–1613, 2009. (Cited on page 123.)
- National Institute for Health and Care Excellence. Depression in adults: treatment and management. Technical Report NG222, National Institute for Health and Care Excellence (NICE), June 2022. URL <https://www.nice.org.uk/guidance/ng222>. Last accessed: 2026-07-01. (Cited on page 2.)
- C. S. Nayak and A. C. Anilkumar. Eeg normal sleep. In *StatPearls [Internet]*. StatPearls Publishing, Treasure Island (FL), May 2023. URL <https://www.ncbi.nlm.nih.gov/books/NBK537023/>. Last updated May 23, 2023. Bookshelf ID: NBK537023. Last accessed: 2025-10-10. (Cited on pages 3, 4, 5, 6, and 202.)
- J. W. Nielsen, M. W. Witek, and S. Hurwitz. Optimal antidepressant dosing: Practical framework for selection, titration, and duration of therapy. *Postgraduate Medicine*, 108(5):111–115, 2000. (Cited on page 2.)
- S. Nikolin, A. Rodgers, A. Schwaab, A. Bahji, C. Zarate, G. Vazquez, and C. Loo. Ketamine for the treatment of major depression: a systematic review and meta-analysis. *EClinicalMedicine*, 62, 2023. (Cited on pages 49 and 50.)
- Nordic Semiconductor. Power profiler kit II (PPK2), 2020. URL <https://www.nordicsemi.com/Products/Development-hardware/Power-Profiler-Kit-2>. Product Code: nRF-PPK2. Last accessed: 2026-07-01. (Cited on pages 81, 106, and 145.)
- B. H. Olstad and C. Zinner. Validation of the polar oh1 and m600 optical heart rate sensors during front crawl swim training. *PLoS One*, 15(4):e0231522, 2020. (Cited on pages 74 and 121.)
- C. C. Paglinawan, G. M. Cruz, and K. R. M. Del Villar. Design of an arduino-powered sleep monitoring system based on electrooculography (eog) with temperature control applications. In *2021 IEEE 11th International Conference on System Engineering and Technology (ICSET)*, pages 297–302. IEEE, 2021. (Cited on pages 73, 74, and 114.)

- L. Palagini, C. Baglioni, A. Ciapparelli, A. Gemignani, and D. Riemann. REM sleep dysregulation in depression: state of the art. *Sleep medicine reviews*, 17(5): 377–390, 2013. (Cited on pages 4, 19, 42, 58, and 162.)
- E. W. Patton, M. Tissenbaum, and F. Harunani. Mit app inventor: Objectives, design, and development. *Computational thinking education*, pages 31–49, 2019. (Cited on pages 69 and 232.)
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011. (Cited on page 168.)
- F. Placidi, S. Zannino, M. Albanese, A. Romigi, F. Izzi, M. G. Marciani, and M. G. Palmieri. Increased cortical excitability after selective rem sleep deprivation in healthy humans: a transcranial magnetic stimulation study. *Sleep medicine*, 14(3):288–292, 2013. (Cited on page 55.)
- Polar Customer Care and Maria. Personal correspondence: Polar Pro Strap material specifications and safety compliance. Email correspondence, April 2025. Personal communication from Polar Electro Customer Care regarding skin contact materials, conductivity characteristics, and regulatory compliance of Polar Pro Strap. (Cited on page 121.)
- PTC Inc. Onshape, 2024. URL <https://www.onshape.com/>. Cloud-based computer-aided design software. Last accessed: 2026-07-01. (Cited on page 93.)
- J. M. Pyne, E. J. Kuc, P. J. Schroeder, J. C. Fortney, M. Edlund, and G. Sullivan. Relationship between perceived stigma and depression severity. *The Journal of nervous and mental disease*, 192(4):278–283, 2004. (Cited on page 2.)
- A. Rechtschaffen, M. A. Gilliland, B. M. Bergmann, and J. B. Winter. Physiological correlates of prolonged sleep deprivation in rats. *Science*, 221(4606):182–184, 1983. (Cited on page 3.)
- Remserve Medical. Compumedics somfit. RemServe Medical website, 2025. URL <https://remservemedical.com/product/compumedics-somfit>. £1,554.00 (listed price), device for home sleep testing; features include EEG, EMG, EOG, SpO₂, etc. Last accessed: 2025-10-12. (Cited on page 9.)
- R. Robbins, M. D. Weaver, J. P. Sullivan, S. F. Quan, K. Gilmore, S. Shaw, A. Benz, S. Qadri, L. K. Barger, C. A. Czeisler, et al. Accuracy of three commercial wearable devices for sleep tracking in healthy adults. *Sensors*, 24(20):6532, 2024. (Cited on pages 8, 163, 184, and 209.)

- A. Rosales-Lagarde, J. L. Armony, Y. del Río-Portilla, D. Trejo-Martínez, R. Conde, and M. Corsi-Cabrera. Enhanced emotional reactivity after selective REM sleep deprivation in humans: an fmri study. *Frontiers in behavioral neuroscience*, 6:25, 2012. (Cited on pages 12, 54, and 163.)
- A. M. Ruscio, D. J. Stein, W. T. Chiu, and R. C. Kessler. The epidemiology of obsessive-compulsive disorder in the national comorbidity survey replication. *Molecular psychiatry*, 15(1):53–63, 2010. (Cited on page 57.)
- P. Sawers. Dreem machine: Rythm raises \$11 million and launches \$500 sleep-monitoring headband, 06 2017. URL <https://venturebeat.com/ai/dreem-machine-rythm-raises-11-million-and-launches-its-500-sleep-tracking-headband>. Original Dreem (first-generation) launch coverage indicating a \$500 list price. Last accessed: 2025-10-12. (Cited on page 9.)
- M. Schaffarczyk, B. Rogers, R. Reer, and T. Gronwald. Validity of the polar h10 sensor for heart rate variability analysis during resting state and incremental exercise in recreational men and women. *Sensors*, 22(17):6536, 2022. (Cited on pages 74 and 121.)
- A.-M. Schyvens, B. Peters, N. C. Van Oost, J.-M. Aerts, F. Masci, A. Neven, H. Dirix, G. Wets, V. Ross, and J. Verbraecken. A performance validation of six commercial wrist-worn wearable sleep-tracking devices for sleep stage scoring compared to polysomnography. *Sleep Advances*, 6(2):zpaf021, 2025. (Cited on pages 8, 163, 184, and 209.)
- J. W. Shepard, D. J. Buysse, A. L. Chesson, W. C. Dement, R. Goldberg, C. Guilleminault, C. D. Harris, C. Iber, E. Mignot, M. M. Mitler, et al. History of the development of sleep medicine in the united states. *Journal of clinical sleep medicine*, 1(01):61–82, 2005. (Cited on pages 4 and 162.)
- J. Siegel and T. P. Gordon. Paradoxical sleep: deprivation in the cat. *Science*, 148(3672):978–980, 1965. (Cited on page 17.)
- P. Sikka, A. Revonsuo, V. Noreika, and K. Valli. Eeg frontal alpha asymmetry and dream affect: Alpha oscillations over the right frontal cortex during rem sleep and presleep wakefulness predict anger in REM sleep dreams. *Journal of Neuroscience*, 39(24):4775–4784, 2019. (Cited on pages 4 and 162.)
- K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. (Cited on pages 189 and 195.)

- J. A. Sirey, M. L. Bruce, G. S. Alexopoulos, D. A. Perlick, S. J. Friedman, and B. S. Meyers. Stigma as a barrier to recovery: Perceived stigma and patient-rated severity of illness as predictors of antidepressant drug adherence. *Psychiatric services*, 52(12):1615–1620, 2001. (Cited on page 2.)
- A. So. Whoop strap 3.0 review: For hardcore athletes only, 03 2021. URL <https://www.wired.com/review/whoop-strap-3/>. Hardware free; membership starts at \$30/month (or \$18/month with 18-month commitment). Last accessed: 2025-10-12. (Cited on page 9.)
- A. Sonntag, B. Rothe, J. Guldner, A. Yassouridis, F. Holsboer, and A. Steiger. Trimipramine and imipramine exert different effects on the sleep eeg and on nocturnal hormone secretion during treatment of major depression. *Depression*, 4(1):1–13, 1996. (Cited on page 32.)
- SparkFun Electronics. Sparkfun single lead heart rate monitor - ad8232, 2014. URL <https://learn.sparkfun.com/tutorials/ad8232-heart-rate-monitor-hookup-guide>. Last accessed: 2026-07-01. (Cited on page 73.)
- V. I. Spoormaker, M. S. Schröter, K. C. Andrade, M. Dresler, S. A. Kiem, R. Goya-Maldonado, T. C. Wetter, F. Holsboer, P. G. Sämann, and M. Czisch. Effects of rapid eye movement sleep deprivation on fear extinction recall and prediction error signaling. *Human brain mapping*, 33(10):2362–2376, 2012. (Cited on page 55.)
- L. Staner, C. Staner, and R. Luthringer. Antidepressant-induced alteration of sleep EEG. In S. R. Pandi-Perumal and M. Kramer, editors, *Sleep and Mental Illness*, chapter 18, pages 202–221. Cambridge University Press, Cambridge, 2010. (Cited on pages 20 and 46.)
- Statista. Global market share held by mobile operating systems since 2009, 2024. URL <https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/>. Last accessed: 2024-03-22. (Cited on page 66.)
- S. S. Steiner and S. J. Ellman. Relation between REM sleep and intracranial self-stimulation. *Science*, 177(4054):1122–1124, 1972. (Cited on pages 3 and 18.)
- J. W. Stewart, P. J. McGRATH, F. M. Quitkin, J. G. Rabkin, W. Harrison, S. Wager, E. Nunes, K. Ocepek-Welikson, and E. Tricamo. Chronic depression: response to placebo, imipramine, and phenelzine. *Journal of clinical psychopharmacology*, 13(6):391–396, 1993. (Cited on pages 24, 26, and 56.)

- A. Supratak, H. Dong, C. Wu, and Y. Guo. Deepsleepnet: A model for automatic sleep stage scoring based on raw single-channel eeg. *IEEE transactions on neural systems and rehabilitation engineering*, 25(11):1998–2008, 2017. (Cited on pages 8 and 189.)
- M. E. Thase, H. Murck, and A. Post. Clinical relevance of disturbances of sleep and vigilance in major depressive disorder: a review. *The Primary Care Companion for CNS Disorders*, 12(6):26104, 2010. (Cited on page 46.)
- O. Tikkanen. Guide to embletta mpr pricing, 2025a. URL <https://web.fibion.com/articles/embletta-mpr-pricing-accessories/>. Fibion Blog. Last accessed: 2025-10-02. (Cited on pages 8 and 162.)
- O. Tikkanen. Guide to sleep profiler & psg2 pricing, 2025b. URL <https://web.fibion.com/articles/sleep-profiler-psg2-pricing-accessories/>. Fibion Blog. Last accessed: 2025-10-02. (Cited on pages 8 and 162.)
- S. Toften, S. Pallesen, M. Hrozanova, F. Moen, and J. Grønli. Validation of sleep stage classification using non-contact radar technology and machine learning (somnofy®). *Sleep Medicine*, 75:54–61, 2020. (Cited on pages 9, 122, 162, 164, 181, 182, 203, and 211.)
- M. H. Trivedi, A. J. Rush, R. Armitage, C. M. Gullion, B. D. Grannemann, P. J. Orsulak, and H. P. Roffwarg. Effects of fluoxetine on the polysomnogram in outpatients with major depression. *Neuropsychopharmacology*, 20(5):447–459, 1999. (Cited on page 49.)
- E. H. Turner, A. M. Matthews, E. Linardatos, R. A. Tell, and R. Rosenthal. Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine*, 358(3):252–260, 2008. (Cited on pages 46, 48, 49, and 57.)
- R. Uher, A. Farmer, N. Henigsberg, M. Rietschel, O. Mors, W. Maier, D. Kozel, J. Hauser, D. Souery, A. Placentino, et al. Adverse reactions to antidepressants. *The British Journal of Psychiatry*, 195(3):202–210, 2009. (Cited on page 2.)
- D. S. Uygun and R. Basheer. Circuits and components of delta wave regulation. *Brain research bulletin*, 188:223–232, 2022. (Cited on page 123.)
- A. Vaccaro, Y. K. Dor, K. Nambara, E. A. Pollina, C. Lin, M. E. Greenberg, and D. Rogulja. Sleep loss can cause death through accumulation of reactive oxygen species in the gut. *Cell*, 181(6):1307–1328, 2020. (Cited on page 3.)

- A. J. Vickers. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. *BMC medical research methodology*, 1(1):1–4, 2001. (Cited on page 23.)
- J. Vincent. Polar’s new fitness watch will measure the wattage of your exercise, 09 2018. URL <https://www.theverge.com/2018/9/13/17854730/polar-fitness-watch-vantage-m-v-launch-buy>. Launch price: Vantage V \$499.90 (US). Last accessed: 2025-10-12. (Cited on page 9.)
- Vitalthings / Somnify. Somnify: Non-contact smart sleep monitor. URL <https://www.indiegogo.com/projects/somnify-non-contact-smart-sleep-monitor>. Indiegogo campaign page with pricing tiers and product details. Last accessed: 2025-10-12. (Cited on pages 8 and 9.)
- G. Vogel and A. Traub. Further studies on REM deprivation of depressed patients. *Psychophysiology*, 1968a. (Cited on page 18.)
- G. Vogel, J. Cohen, D. Mullis, T. Kensler, and S. Kaplita. Nefazodone and REM sleep: how do antidepressant drugs decrease REM sleep? *Sleep*, 21(1):70–77, 1998. (Cited on page 46.)
- G. W. Vogel. A review of REM sleep deprivation. *Archives of General Psychiatry*, 32(6):749–761, 1975. (Cited on pages 4, 21, and 45.)
- G. W. Vogel. Evidence for REM sleep deprivation as the mechanism of action of antidepressant drugs. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 7(2-3):343–349, 1983. (Cited on pages 45 and 56.)
- G. W. Vogel and A. C. Traub. REM deprivation: I. the effect on schizophrenic patients. *Archives of General Psychiatry*, 18(3):287–300, 1968b. (Cited on pages 18 and 45.)
- G. W. Vogel, A. C. Traub, P. Ben-Horin, and G. M. Meyers. REM deprivation: II. the effects on depressed patients. *Archives of General Psychiatry*, 18(3):301–311, 1968. (Cited on pages 11, 18, 51, and 56.)
- G. W. Vogel, A. Thurmond, P. Gibbons, K. Sloan, M. Boyd, and M. Walker. REM sleep reduction effects on depression syndromes. *Archives of General Psychiatry*, 32(6):765–777, 1975. (Cited on pages xviii, 3, 4, 10, 11, 18, 19, 20, 21, 56, 162, 163, and 221.)
- G. W. Vogel, R. McAbee, K. Barker, and A. Thurmond. Endogenous depression improvement and REM pressure. *Archives of General Psychiatry*, 34(1):96–97, 1977. (Cited on pages 26, 51, 56, and 221.)

- G. W. Vogel, F. Vogel, R. S. McAbee, and A. J. Thurmond. Improvement of depression by rem sleep deprivation: new findings and a theory. *Archives of General Psychiatry*, 37(3):247–253, 1980. (Cited on pages 26, 33, 51, and 221.)
- G. W. Vogel, A. Buffenstein, K. Minter, and A. Hennessey. Drug effects on REM sleep and on endogenous depression. *Neuroscience & Biobehavioral Reviews*, 14(1):49–63, 1990. (Cited on pages 20 and 45.)
- J. Volpi-Abadie, A. M. Kaye, and A. D. Kaye. Serotonin syndrome. *Ochsner Journal*, 13(4):533–540, 2013. (Cited on page 2.)
- U. Voss, R. Holzmann, A. Hobson, W. Paulus, J. Koppehele-Gossel, A. Klimke, and M. A. Nitsche. Induction of self awareness in dreams through frontal low current stimulation of gamma activity. *Nature neuroscience*, 17(6):810–812, 2014. (Cited on page 163.)
- T. Walczak and S. Chokroverty. 7 - electroencephalography, electromyography and electrooculography: General principles and basic technology. In S. Chokroverty, editor, *Sleep Disorders Medicine*, pages 95–117. Butterworth-Heinemann, 1994. ISBN 978-0-7506-9002-7. (Cited on page 73.)
- J. C. Ware, F. W. Brown, P. J. Moorad Jr, J. T. Pittard, and B. Cobert. Effects on sleep: a double-blind study comparing trimipramine to imipramine in depressed insomniac patients. *Sleep*, 12(6):537–549, 1989. (Cited on page 32.)
- J. D. Watson and F. H. Crick. The structure of dna. In *Cold Spring Harbor symposia on quantitative biology*, volume 18, pages 123–131. Cold Spring Harbor Laboratory Press, 1953. (Cited on pages 4 and 162.)
- C. Weinberger and contributors. flutter_blue_plus: Flutter plugin for connecting and communicating with bluetooth low energy devices, 2024. URL https://pub.dev/packages/flutter_blue_plus. Flutter package for Bluetooth Low Energy (BLE) communication. Last accessed: 2026-07-01. (Cited on pages 87 and 237.)
- A. Wirz-Justice and R. H. Van den Hoofdakker. Sleep deprivation in depression: what do we know, where do we go? *Biological psychiatry*, 46(4):445–453, 1999. (Cited on pages 2, 18, 42, and 230.)
- World Famous Electronics LLC. Pulsesensor datasheet: The easy to use heart rate sensor & kit. Datasheet, World Famous Electronics LLC, New York, USA, November 2024. URL https://cdn.shopify.com/s/files/1/0100/6632/files/PulseSensor_Datasheet_2024-Nov.pdf?v=1732032216. Revision B. Last accessed: 2025-09-18. (Cited on pages 116 and 145.)

World Health Organization. Depression and other common mental disorders: Global health estimates. Technical report, World Health Organization (WHO), 2017. (Cited on page [2](#).)