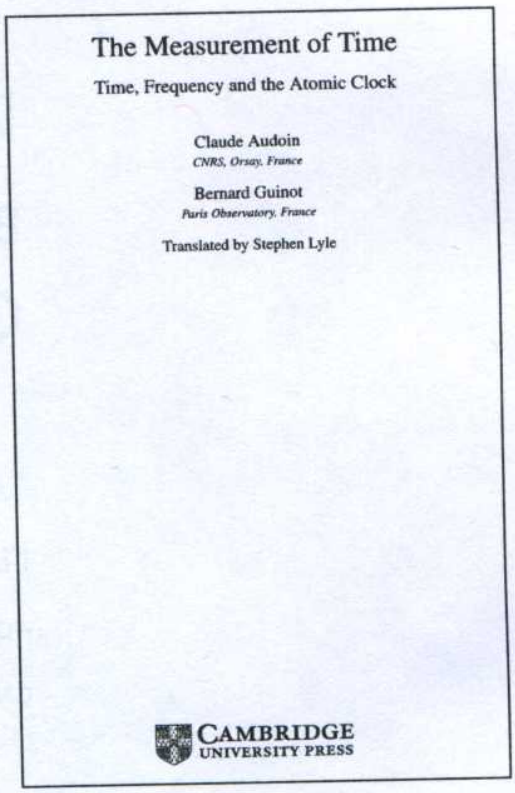
К. Одуан, Б. Гино
Измерение времени. Основы GPS
Москва:
Техносфера, 2002. – 400 с.

Первое за последние 50 лет всеобъемлющее издание по физике и метрологии времени. Точность атомных часов и атомной шкалы авторы рассматривают в контексте фундаментальных физических исследований, а также в применении к таким техническим системам, как глобальные системы навигации и позиционирования.

Книга станет настольной для метрологов, астрономов, инженеров-разработчиков высокоточных время-частотных приборов и систем, послужит учебным пособием и справочником для специалистов в таких областях, как геофизика, атомная физика, навигация и телекоммуникации. Русское издание дополнено обширным разделом по физико-техническим основам и алгоритмам глобальной навигационной системы GPS, что позволяет рассматривать эту часть книги как первое учебное пособие по данной тематике в стране.


The Measurement of Time

Time, Frequency and the Atomic Clock

Claude Audoin
CNRS, Orsay, France

Bernard Guinot
Paris Observatory, France

Translated by Stephen Lyle

CAMBRIDGE
UNIVERSITY PRESS

«Les fondements de la mesure du temps»
by Claude AUDOIN & Bernard GUINOT
© Masson, Paris 1998
© 2002, ЗАО «РИЦ «Техносфера»
перевод на русский язык, оригинал-макет,
оформление

ISBN 5-94836-006-7

Оглавление

Глава 1.
Введение 12

Глава 2.
Принципы измерения времени 18
2.1. Введение 18
2.2. Время и воспроизводимость: понятие длительности 19
2.2.1. Хаос 19
2.2.2. Измерение времени через измерение других величин 20
2.2.3. Локальная природа явлений 20
2.2.4. Разное старение? 21
2.2.5. Воспроизводимость и измерительные эталоны 22
2.3. Время, эволюция и шкалы времени 22
2.4. Два способа измерения времени 24
2.4.1. Использование воспроизводимости 25
2.4.2. Использование динамической модели 27

Глава 3.
Измерение времени и теоретические модели 30
3.1. Классическая модель и абсолютное время 31
3.2. Специальная теория относительности 33
3.2.1. Физика в инерциальных системах 33
3.2.2. Лоренцевское преобразование и инвариантность интервала 36
3.2.3. Время в специальной теории относительности 37
3.2.4. Область применения специальной теории относительности 38
3.3. Общая теория относительности 39
3.3.1. Обзор 39
3.3.2. Пост-ньютоновское приближение 42
3.4. Выводы и заключение 52

Глава 4.
Эволюция измерений времени 54
4.1. Дата, календарь и час 54
4.2. Измерение времени, основанное на смене дня и ночи 56
4.2.1. Среднее солнечное время 56
4.2.2. Шкалы всемирного времени и часовые пояса 57

4.2.3. К реализации Всемирного Времени 59
4.2.4. Определение секунды до 1960 г:
секунда среднего солнечного времени 62
4.3. Время, основанное на динамике Солнечной Системы:
эфемеридное время 63
4.3.1. Первые сомнения в равномерности вращения Земли 63
4.3.2. Принятие неравномерности вращения Земли 63
4.3.3. Определение секунды с 1960 по1967 гг.:
эфемеридная секунда 66
4.3.4. Эфемеридное время: шкала сохраненная для астрономов 68
4.4. Измерение атомного времени 68
4.4.1. Атомные стандарты частоты
и определение атомной секунды (1967 г.) 68
4.4.2. От Всемирного Времени
к Международному Атомному Времени 74
4.5. Всемирное координированное время 79
4.6. Финальные заметки 82

Глава 5.
Время и часы 84

5.1. Введение 84
5.2. Нестабильность частоты и времени 86
5.2.1. Определения 86
5.2.2. Условные обозначения 87
5.2.3. Измерения во временной области 88
5.2.4. Измерения в спектральной области 91
5.2.5. Моделирование частотных и временных флуктуаций 91
5.2.6. Характеристики нестабильности частоты 93
5.2.7. Характеристики нестабильности времени 101
5.2.8. Определение вклада от каждого генератора 102
5.3. Механические осцилляторы 103
5.4. Пьезоэлектрические осцилляторы 103
5.4.1. Резонаторы 104
5.4.2. Осцилляторы 106
5.4.3. Характерные величины 107
5.5. Атомные стандарты времени и частоты.
Представление о точности 107
5.6. Сличения времени и частоты 109
5.6.1. Общие наблюдения 109
5.6.2. Соотношение между частотой
и приращением собственного времени 110

5.6.3. Сличения времени 111
5.6.4. Сличение частоты 125

Глава 6.
Атомные стандарты частоты 132

6.1. Спектроскопические основы атомных стандартов частоты 132
6.1.1. Инвариантность и универсальность атомных свойств 132
6.1.2. Спектральные свойства щелочных атомов 133
6.1.3. Правила отбора 139
6.1.4. Ширина резонанса 141
6.2. Стадии работы атомных часов 142
6.2.1. Наблюдение в очень разреженных средах 142
6.2.2. Подготовка и детектирование атомных состояний 142
6.2.3. Устранение эффекта Доплера 146
6.2.4. Взаимодействие между атомами и электромагнитным полем 149
6.2.5. Соответствующие электронные системы 149
6.3. Реализация приближения изолированного атома в покое 150
6.4. Цезиевые пучковые часы 153
6.4.1. Цезиевые часы с магнитным отклонением пучка 154
6.4.2. Цезиевые пучковые часы с оптической накачкой 187
6.4.3. Цезиевые часы с лазерным охлаждением 195
6.4.4. Рубидиевые часы с лазерным охлаждением 209
6.5. Водородные мазеры 210
6.5.1. Основные принципы работы водородного мазера 211
6.5.2. Условия генерации 213
6.5.3. Активный водородный генератор 217
6.5.4. Пассивный водородный мазер 227
6.5.5. Метрологические свойства водородного мазера 232
6.5.6. Криогенный водородный мазер 235
6.6. Часы на рубидиевой ячейке 237
6.6.1. Описание 237
6.6.2. Принцип действия 239
6.6.3. Электроника 240
6.6.4. Метрологические свойства часов с рубидиевой ячейкой 241
6.6.5. Современные исследования и перспективы 243
6.7. Часы на удержанных ионах 243
6.7.1. Удержание ионов в радиочастотной ловушке 244
6.7.2. Удержание облака ионов 249
6.7.3. Описание стандарта частоты на облаке ионов ртути 251
6.7.4. Метрологические свойства стандартов частоты на облаке холодных ионов ртути 254

6.7.5. Стандарт частоты на ионах иттербия 256
6.7.6. Стандарт частоты на ионе ртути с лазерным охлаждением 256
6.8. Другие атомные стандарты 257
6.8.1. Стандарт частоты на пучке магния 257
6.8.2. Оптические стандарты частоты 258
6.8.3. Оптический стандарт частоты на одиночном ионе в покое 258
6.8.4. Оптические стандарты для измерения времени 260
6.9. Заключение 261

Глава 7.
Измерения атомного времени 262

7.1. Определения 262
7.1.1. Определение секунды 262
7.1.2. Международное атомное время и его связь с геоцентрическим и барицентрическим координатными временами 263
7.2. Установление международного атомного времени 266
7.2.1. Алгоритмы шкалы времени 266
7.2.2. Алгоритм для международного атомного времени 268
7.3. Свойства TAI и UTC 280
7.3.1. Надежность 280
7.3.2. Точность показаний 280
7.3.3. Стабильность частоты 280
7.3.4. Точность частоты 281
7.4. Мировая организация измерений времени 282
7.4.1. Всемирное координированное время 282
7.4.2. Локальное представление UTC и независимые локальные атомные времена 284
7.4.3. Другие формы атомного времени 286
7.5. Распространение времени и частоты 287
7.5.1. Частота 287
7.5.2. Время 288
7.6. Итоги и перспективы 289

Глава 8.
Астрономические времена 292

8.1. Всемирное время 292
8.1.1. Концептуальные определения 292
8.1.2. Практическое определение UT1 295
8.1.3. Другие виды всемирного времени 297

8.1.4. Техника измерений ... 298
8.1.5. Неравномерности вращения Земли ... 302
8.1.6. Применение всемирного времени ... 305
8.2. Эфемеридное время ... 307
8.2.1. Определение ... 307
8.2.2. Установление эфемеридного времени ... 308
8.2.3. Трудности и текущие решения ... 309
8.3. Пульсарное время ... 310

Глава 9.
Применение сверхточного времени и частоты ... 316

9.1. Фундаментальные исследования ... 317
9.1.1. Метрология ... 317
9.1.2. Измерение физических констант ... 319
9.1.3. Воздействие на атомную физику ... 320
9.1.4. Строение пространства-времени и гравитация ... 321
9.2. Определение положения, геодезия и навигация ... 326
9.2.1. Принципы время частотных методов ... 326
9.2.2. Глобальная система определения положения ... 328
9.2.3. Другие системы определения положения ... 332
9.3. Интерферометрия со сверхдлинной базой (РСДБ) ... 333
9.4. Программа TOPEX-POSEIDON ... 334
9.5. Телекоммуникация ... 338

Глава 10.
Навигационная система GPS – генератор координатно-временного поля ... 340

10.1. Физические и геометрические принципы действия GPS ... 341
10.1.1. Геодезические построения ... 341
10.1.2. Методы измерения расстояний ... 343
10.1.3. Орбиты спутников ... 345
10.2. Сигналы GPS ... 347
10.2.1. Техника формирования кодовых последовательностей ... 347
10.2.2. Структура и передача в эфир навигационного сигнала ... 349
10.2.3. Прием и обработка сигналов ... 351
10.3. Методы учета факторов, искажающих результаты измерений ... 354
10.3.1. Влияние ионосферы ... 354
10.3.2. Влияние тропосферы ... 356
10.3.3. Другие факторы ... 357

10.4. Алгоритмы и точность координатно-временных определений ... 359
10.4.1. Абсолютный и относительные методы измерений 359
10.4.2. Особенности обработки фазовых измерений 362
10.4.3. Точность различных методов GPS-измерений 363
10.5. Организационная структура GPS 366
10.5.1. Космический, управляющий и пользовательский сегменты 366
10.5.2. Функциональные дополнения к GPS 371
10.5.3. Международная служба IGS 372
10.6. Области применения GPS 375

Приложение 1.
Акронимы для лабораторий времени 378

Приложение 2.
Сокращения 380

Приложение 3.
Определение основных еденицСИ 382

Приложение 4.
Международные службы 384

Ссылки 386

Предметный указатель 398

Глава 1

Введение

Использование свойств атомов для применения времени зародилось в 1955 году, когда в Великобритании начал действовать на регулярной основе первый цезиевый стандарт частоты. Конечно к этому времени уже были испытаны другие атомные стандарты частоты. Однако они не были способны, как цезиевый стандарт частоты обеспечить более точную единицу времени и более равномерную шкалу времени по сравнению с этими величинами, основанными на движении небесных сил. Число цезиевых стандартов быстро возрастало, а их точность стремительно повышалась. Хорошее согласие между ними порождало уверенность и в 1967 году стало возможным принятие определения атомной секунды для замены прежнего астрономического определения. Несколько лет спустя существование адекватной атомной шкалы времени для применения в мировом масштабе было официально признано на 14 генеральной конференции по мерам и весам в 1971 году.

Несмотря на эти достижения более четверти века спустя измерения атомного времени все еще не стали привычными. Являлось ли это результатом глубоко укоренившейся привычки соизмерять нашу повседневную жизнь с движением небесных тел? Или это реакция на отсутствие романтичности в атомных часах и в их непостижимой точности? Может быть это и так. Но существуют и более объективные причины, которые делают трудным принятие атомного времени. Одна из этих причин заключается в необходимости формулирования измерений времени в рамках релятивистский теории относительности, а другая причина заключается в сложных соотношениях между атомным временем и временем, используемым астрономами в динамических теориях. Последняя проблема была предметом споров и дискуссий на протяжении более 10 лет и даже международный астрономический союз был не способен освободиться от призрачных надежд в своей резолюции, принятой в 1991 году. Недавно появились предположения, что открытие быстрых пульсаров, так называемых миллисекундных пульсаров, может вернуть измерения времени назад под юрисдикцию астрономии.

Тем временем в тиши своих лабораторий физики едва ли были восприимчивы к таким запросам. Последующий период относительного застоя, который продолжался около 20 лет, внезапно завершился резким повышением точности стандартов частоты в 1994 году. Новая техника, позволившая получить такой рывок в точности, вселяет также надежды на прогресс в ближайшем будущем. С другой стороны мировая индустрия также не оказывается в стороне от прогресса. Производимые кварцевые генераторы и атомные стандарты частоты демонстрируют постоянное совершенствование. Наиболее популярными инструментами для фундаментальных метрологических исследований всегда были цезиевые часы и тысячи таких часов были произведены промышленностью. В 1994 году была создана промышленная модификация цезиевых часов имеющих десятикратное улучшение по всем эксплуатационным характеристикам по сравнению с предыдущими версиями часов.

Благодаря промышленным разработкам оборудования для измерения времени появились на свет очень важные применения техники время-частотных измерений. В качестве примера приведем систему определения положения GPS, которая начала функционировать в полном объеме в 1995 году. Эта система может удовлетворить нужды и представлять интерес для множества потенциальных пользователей от геодезистов, желающих представить всю планету с точностью до нескольких сантиметров, до праздношатающихся бродяг. Не забудем также и нужды военных, для которых эта система и была первоначально разработана.

Рассмотрение этих недавних достижений в уже хорошо испытанной технике атомного измерения времени и рассмотрение новых проблем, порождаемых таким стремительным улучшением качества, кажется совершенно подходящим моментом для представления основ измерения времени, а именно: основных принципов, конструкций лучших стандартов, способов сличения времени и частоты с предельно высокой точностью, способов оценки первичных стандартов, т.е. лучшей реализации секунды и международного атомного времени на наивысшем уровне точности. Кажется также полезным напомнить значение астрономических шкал времени, которые по ряду причин по-прежнему имеют важное значение.

Как мы уже упоминали фундаментальные измерения времени уже не могут быть представлены вне рамок релятивистской теории относительности. Вплоть до настоящего времени эйнштейновская теория относительности, простейшая из теорий относительности, всегда находилась в согласии с экспериментальными результатами. По этой причине мы примем ее в качестве нашей рабочей гипотезы. Однако в этом отношении мы должны оставить также некоторые оговорки. Вначале мы надеялись трак-

товать измерения времени в рамках классической физики, а затем указывать маленьким шрифтом, что было бы необходимо учесть в релятивистском истолковании. Но затем нам показалось, что этот метод рискован в том смысле, что обширный обзор общей теории относительности может рассматриваться как дополнение к классической теории, т.е. что релятивистская теория может рассматриваться как набор релятивистских поправок. В действительности релятивистская теория вместе с гравитационными эффектами является завершенной моделью структуры пространства и времени. Поэтому мы покажем полезные разработки непосредственно в контексте этой книги. Читатель, едва знакомый с этими вопросами, может тем не менее не волноваться, потому что для наибольшей простоты мы будем просто напоминать постулаты этой теории и их математические последствия.

Неизбежная линейность численных расчетов плохо подходит для множества взаимодействий между различными аспектами измерения времени. Представленное здесь краткое описание в различных частях книги направлено на ориентацию читателя в этих проблемах.

В главе 2 напоминаются основные принципы измерения времени в рассматриваемых здесь макроскопических явлениях. Вообще говоря это внимательный обзор и анализ постулатов Пуанкаре, важность которых простирается и на релятивистские аспекты.

В главе 3 показано, как измерения времени связаны с общей теорией относительности. В расширенном пространстве время является одним из элементов четырехмерной координатной системы и не может быть отделено от пространственных координат. Мы определим используемые координатные системы и установим некоторые уравнения, связывающие локальную физику с пространственно-временными координатами. Мы обсудим также соглашения по синхронизации и сличению разнесенных часов.

В главе 4 представлены исторические аспекты. В ней показано развитие измерений времени, в частности, развитие в двадцатом столетии. Мы обсудим идеи, которые стимулировали развитие техники и решение проблем, появившихся в результате перехода от астрономического к атомному измерению времени.

В главе 5 описаны инструменты и методы, привлекаемые в атомных стандартах и, в частности, методы представления метрологических характеристик и их сличений. Представление стабильности частоты привело к новым разработкам в статистике. Оно подводит к понятием серий измерений во времени и к определению вариаций, что позволяет многим производителям стандартов частоты и их пользователям говорить на общем языке. Представление точности частоты является также специфи-

ческой проблемой в измерениях времени хотя оно может быть расширено и на любые физические величины, единицы измерения которых основаны непосредственно на естественных явлениях, которые рассматриваются в качестве воспроизводимых явлений. Касаясь вопросов сличения частот и шкал времени на расстояниях мы увидим важность этого вопроса в том смысле, что здесь мы часто приходим к бóльшим неопределенностям, чем неопределенности сличаемых величин.

В главе 6 объясняются принципы действия стандартов времени и частоты, называемых также атомными часами. Мы начнем с обзора на элементарном уровне основных идей атомной физики и спектроскопии, необходимых для понимания принципов работы этих стандартов и их основных особенностей. Существует несколько типов атомных часов, использующих некоторые переходы между атомными уровнями одного и того же типа, но в разных атомах, таких как цезий, рубидий или водород. Используются даже ионы, такие, например, как ион ртути. Именно разработанные цезиевые часы используются в метрологических лабораториях, в которых достигнута лучшая реализация определения секунды. Они являются первичными стандартами времени и частоты. Их главная особенность это точность. Промышленные цезиевые часы используются в относительно недорогом способе передачи и распространения времени, а следовательно, и самого атомного времени. Будут также представлены и другие типы атомных часов. Они были разработаны и созданы для применения в тех областях, где они подходят наилучшим образом. Некоторые из них громоздкие, но исключительно стабильные, в то время как другие менее стабильные, но очень компактные. В этой главе мы обратимся к принципам управления атомами с помощью лазерного излучения. Появившаяся недавно экспериментальная техника с помощью лазерного излучения позволяет замедлить и поймать атомы, прежде чем осуществить их запуск с очень низкой скоростью и очень хорошим управлением их траекторией. В 1996 году с помощью этой техники на новом типе стандарта была достигнута точность, превосходящая точность любых других первичных стандартов частоты и времени. Потенциал этого метода управления атомами в сочетании с исключительным международным состязанием, к которому он привел, является надежной гарантией будущего улучшения характеристик атомных часов.

Глава 7 посвящена главным образом созданию единственной атомной шкалы времени, принятой по соглашению в качестве международного стандарта, а именно международного атомного времени (обозначаемого на всех языках как TAI от французского названия Temps atomic international). Мы рассмотрим, как с помощью международной кооперации стало возможным установить и поддерживать такую шкалу времени и исследо-

вать ее качество и способы ее распространения. Но мы также увидим, насколько был необходим компромисс при принятии в качестве практического стандарта шкалы времени, основанной на TAI и шкалы, изобретательно приспособленной к поддержанию связи с всемирным временем UT, определяемым вращением Земли. Этот гибрид известен как всемирное координированное время (официально обозначаемое акронимом UTC). Мы покажем, как так называемую собственную секунду, необходимую для работы лабораторий, можно получить из TAI или UTC.

В главе 8 даны определения различных астрономических времен для общего применения. Эти шкалы времени были постепенно разработаны путем наблюдения небесных тел на протяжении долгого периода истории. Их обычные определения несут в себе отпечаток исторического развития и соответственно что-то непонятное для тех, кто не знаком с астрономией. Таким образом мы приложили все усилия для того, чтобы показать основные концепции, на которых основаны астрономические рабочие определения. Мы исследуем назначения этих шкал времени. В частности, свидетельствующее о вращении Земли всемирное время, является очень важной величиной, требующей постоянных измерений. В действительности, усилия посвященные измерениям всемирного времени соизмеримы с усилиями, направленными на поддержание атомного времени. Что касается пульсаров, то хотя они и не обеспечивают хорошую меру времени, тем не менее их изучению препятствуют неопределенности атомного времени. Изучение пульсаров это одно из наиболее востребованных применений атомного времени.

В главе 9 представлены некоторые применения атомного времени, выбранные за их высочайший уровень требований к стандартам времени и частоты. Естественно что эти применения принадлежат к области фундаментальных исследований. Применение атомных стандартов в более утилитарных программах сопровождается определенной долей жертвенности в вопросах их предельных качеств в пользу повышения надежности и уменьшения стоимости. Однако более практичные инструменты иногда отстают совершенно незначительно от лабораторных приборов в тех случаях, когда необходима точность и стабильность. Хорошей иллюстрацией этому является система GPS. Так как все мы являемся прямыми или косвенными пользователями этих систем, то мы в какой-то степени коснемся деталей этих систем. Мы покажем также в качестве примера как соотносится измерение времени с космической океанографической миссией TOPEX/POSEIDON. В действительности мы увидим, что это подчеркивает многие аспекты этой миссии, основанные на высочайшем уровне качества.

И наконец глава 10 добавлена как практический пример применения измерений времени и частоты в навигационной системе GPS, которая стремительно развивается в направлении массового применения. Здесь даны более подробно физические основы и принципы действия системы GPS и рассмотрены некоторые области применения системы GPS.

Глава 2

Принципы измерения времени

2.1. Введение

Диапазон измерения времени огромен, от планковского времени 10^{-43} с – наименьшей величины, которую мы можем констатировать в физике, до величин встречающихся в космологии, т.е. порядка десятка миллиардов лет или $10^{17} - 10^{18}$ с.

Из всего этого диапазона мы будем иметь дело только с той очень малой частью, где могут проводиться точные измерения. Грубо говоря, наша область простирается от периодов порядка 10^{-15} с, характерных для атомных переходов в видимом диапазоне, до периодов в несколько тысяч лет (~10^{11} с), прошедших с начала астрономических наблюдений. Весь этот диапазон лежит в области типичных лабораторных экспериментов и динамической астрономии, которые, в свою очередь, либо обеспечены, либо должны быть обеспечены эталонами времени: поле деятельности метрологии времени.

Вплоть до рождения релятивистских теорий в начале 20-го столетия, время и пространство составляли заданный извне и неизменный каркас, в рамках которого могла осуществляться наша деятельность. Мы могли воспринимать этот каркас через присутствие и передвижение объектов. В то же время мы были убеждены, что он существовал независимо от своего содержания, и даже, что он мог бы существовать вообще без содержания. Сегодня считается, что пространство – время не может рассматриваться отдельно от массы и энергии и что не существует привилегированного времени, наподобие абсолютного времени. Такая концепция повергает нас в двойное унижение перед природой. Допустим, что мы не знаем что такое время. Мы знаем только, что нам нужны временная координата и три пространственных координаты, чтобы описать физические явления и придумать математические модели для них. Почему эти четыре независимых координаты, а не пять или больше? Просто потому, что четырех координат достаточно, чтобы в наших физических моделях не обнаруживалось никакого измеримого несоответствия.

Освободившись от абсолютного времени, легче принять прагматический подход к определению или даже к измерению времени. И действительно, время определяется таким способом, чтобы математические модели природы оставались простыми. Следует также заметить, что другие физические величины, такие как длина, масса и т.д. должны быть определены таким же путем. Это и есть то, что через проверочный критерий простоты мы способны понять как основы измерения времени, определить единицы времени и выбрать шкалу времени такую, которая могла бы быть принята как мировой эталон для датирования событий.

2.2. Время и воспроизводимость: понятие длительности

Нет ничего проще, чем взять хорошо известную наблюдаемую особенность нашей связи с природой и обратить ее в постулат. Такой особенностью является воспроизводимость любого опыта во времени и пространстве. Именно Пуанкаре первым установил этот постулат в своей дискуссии об измерении времени [2.1] (Перевод): «Когда мы используем маятник чтобы измерить время, как постулат мы косвенно подразумеваем то, что длительность двух идентичных явлений одинакова; или по другому: одни и те же причины требуют одного и того же времени чтобы произвести одни и те же эффекты».

Однако эта простота иллюзорна. Ряд положений требуют более пристального внимания.

2.2.1. Хаос

Опыт учит нас [2.1]: «что примерно одинаковые причины требуют такого же времени, что бы произвести примерно одинаковые эффекты».

Утверждение Пуанкаре здесь может быть до некоторой степени неожиданным. Он был одним из пионеров в изучении детерминистского хаоса, согласно которому примерно одинаковые причины могут привести к совершенно разным эффектам за одно и то же время.

Существование хаотических явлений не противоречит постулату воспроизводимости. Это только напоминает нам, что при его применении должна соблюдаться осторожность. Целью метрологии является обеспечение воспроизводимыми эталонами, и она основана на хорошо известных явлениях, для которых приблизительное применение постулата воспроизводимости остается возможным. И в действительности главная роль метролога состоит в распознавании и использовании таких явлений, которые мало зависят от плохо контролируемых возмущающих причин.

2.2.2. Измерение времени через измерение других величин

Когда Пуанкаре говорит о тех же причинах и тех же эффектах он предполагает, что мы уже имеем эталоны для их оценки. Измерение времени, т.е. в нашем случае длительности, может быть доступно нам только через измерение других величин.

Если мы хотим в его примере с маятником получить единицу длительности – секунду путем простого хода маятника туда и обратно, мы должны определить с большой точностью некоторую опорную точку, через которую этот маятник проходит, и получить сигнал в момент ее прохождения. На практике нет ничего мгновенного. Фотоэлектрический метод может дать сигнал длительностью свыше 0,001 секунды. Тогда секунда будет определена только с относительной точностью 10^{-3}. Однако постулат воспроизводимости становится полезным, когда мы можем повторять эксперимент непрерывно. Хотя время отсчета остается неизменным, ошибка в измерении единицы в этом случае делится в среднем на число совершенных периодов. В примере с маятником она уменьшается до 10^{-8} после одного дня измерений. В настоящее время секунда определена как длительность 9192631770 периодов, соответствующих определенному атомному переходу. Простой счет этих периодов в течение одного дня, при условии что мы ничего не пропустили, приведет к относительной неопределенности секунды 10^{-15}.

В двух вышеприведенных примерах измерение времени обращается в измерение напряжения. Постулат воспроизводимости позволяет перевести это измерение к простой, почти качественной индикации напряжения так, что любая неопределенность в вольте не будет иметь какого бы то ни было влияния на воспроизведение секунды. Здесь также высвечивается преимущество использования явлений с очень короткой длительностью.

2.2.3. Локальная природа явлений

Для того чтобы контролировать причины явлений в надежде, что они будут теми же самыми, когда бы мы их ни воспроизводили, эксперимент должен быть локальным, т.е. в непосредственной близости от экспериментатора.

Существуют ли какие-нибудь естественно воспроизводимые удаленные явления, которые являются достаточно быстрыми, чтобы отвечать нуждам точности? В настоящее время мы не знаем таковых. В течение

длительного времени вращение Земли, казалось бы, отвечало этим требованиям, пока во второй половине 20-го столетия не появилось доказательство его неравномерности. Быстро вращающиеся пульсары с периодом в диапазоне миллисекунд, впервые открытые в 1982 году, могли бы быть отличными эталонами для промежутков времени, если бы их вращение не замедлялось.

Если бы удаленное и хорошо воспроизводимое явление было бы открыто, мы могли бы основать определение единицы длительности на наблюдениях этого явления. Фактически мы бы конечно так и сделали, если бы в этом были какие-нибудь преимущества. Однако этот шаг потребовал бы обращения за помощью не только к теории распространения электромагнитных волн, но и к теории этого явления. Используемое в простейшем виде, без обращения к какой бы то ни было теории, явление однажды могло бы быть оспорено тем, что оно не адаптивно к локальным измерениям и поэтому постулат воспроизводимости заставляет нас опираться не на удаленные, а на локальные измерительные эталоны.

2.2.4. Разное старение?

Постулат Пуанкаре подразумевает причинные связи. Мы можем интересоваться, что случается, когда два явления происходят параллельно, независимо одно от другого. Если химическая реакция подойдет к окончанию, когда наши атомные часы отсчитают 10^{10} периодов, можем ли мы быть уверены, что через миллионы лет та же самая реакция будет происходить за то же самое число периодов? Постулат позволяет нам утверждать, что это будет действительно то же самое число. Фактически нет причин, которые мешают нам установить причинную связь между двумя явлениями, например, путем запуска реакции после n-атомных периодов и рассмотрением такого объединенного объекта восприятия как «счет периодов и химическая реакция». Длительность этого объединенного явления должна быть неизменной. Постулат не оставляет места для разной скорости старения различных естественных явлений. Если спросить, а не стареют ли сами локальные явления и их законы глобально, то постановка такого вопроса предполагает существование какого-то фиктивного времени, посредством которого мы могли бы сделать наши оценки. Существование такого времени является безосновательной гипотезой.

Тем не менее, стоило бы воздержаться от неоправданной самоуверенности, уповая на постулат воспроизводимости. Хотя он еще никогда и не

нарушался, может однажды случиться так, что при дальнейшем прогрессе в экспериментах справедливость постулата может быть поставлена под сомнение.

2.2.5. Воспроизводимость и измерительные эталоны

Принимая во внимание вышеуказанное обсуждение, мы можем прийти к заключению, что определение единицы измерения для локальной физики может основываться на воспроизводимости локальных экспериментов. Проверка определения семи основных единиц системы СИ (Международной Системы Единиц), а именно метра, килограмма, секунды, ампера, кельвина, моля и канделы (определения даны в приложении 3), выявила, что постоянно воспроизводится пять единиц. Килограмм и кельвин являются исключениями. Определение килограмма как массы уникального предмета требует предположения о непрерывности и постоянстве свойств этого предмета, которые, возможно, являются формой воспроизводимости. Что касается Кельвина (температура), то его определение относится к более сложным моделям, а именно к законам термодинамики, основанным на дополнительных постулатах.

Единица длительности секунда, обычно рассматривается нами как единица времени и в ее настоящем определении является единицей, устанавливаемой локально, так же, как и все другие единицы. Это приводит к тем последствиям, что когда мы должны рассматривать измерения в моделях, опирающихся на общую теорию относительности, то часто это необходимо делать с достаточно высокой точностью, т.к. с очень высокой точностью мы можем определить секунду.

2.3. Время, эволюция и шкалы времени

Вернемся теперь к происходящим во времени явлениям, которые мы беспричинно считаем воспроизводимыми. Так как такие явления имеются в большом количестве, то мы выберем те явления, которые относятся к динамике звездных тел, т.к. существуют высококачественные математические модели, разработанные для их описания, а также возможно потому, что в прошлом время всегда связывалось с движением таких небесных объектов.

Для того, чтобы описать эти движения нам нужна только шкала координат времени, т.е. шкала времени, главным качеством которой является то, что каждый может ее принять и что она должна быть доступна всем. (Это сразу же приводит к постановке вопроса о синх-

ронизации часов удаленных наблюдателей? Обсуждение мы отложим до части 3. Представим только на время, что мы можем синхронизировать часы). Однако если мы хотим установить модель для тех движений, которые позволяют нам представить их как астрономические эфемериды, временной аргумент в такой теории должен удовлетворять и другим критериям. Что это за критерии? Давайте еще раз обратимся к собственному суждению Пуанкарэ, а не к попыткам его изложения.

«Они (астрономы) определяют длительность следующим образом: время должно быть определено таким способом, чтобы сохранялись и Ньютоновский закон гравитации, и второй закон движения. Закон гравитации является наблюдаемой истиной и как таковой он не точный. Это показывает, что опять мы имеем только приблизительное определение. Если мы предположим, что мы приняли бы другое средство измерений времени, то эксперименты, основанные на Ньютоновском законе гравитации, имели бы то же самое значение. Это только предположение того, что закон изменился бы, на самом деле он бы просто трактовался на другом языке. Он, очевидно, был бы менее простой. Определение, одобренное астрономами, может быть обобщено в следующем виде: время должно быть определено таким способом, чтобы уравнения механики стали как можно проще. Другими словами нет наиболее правильного способа измерения времени по сравнению со всеми другими способами. То, что общепринято, просто наиболее удобно.

Таким образом, мы имеем дело с двумя определениями для измерения времени, а именно: время, исходящее из воспроизводимости локальных явлений; и время, исходящее из классической динамики. Являются ли эти два определения эквивалентными?

До начала 20 века не предполагали, что любое из этих определений приведет к представлению об одинаковом абсолютном времени, и будет отличаться только уровнем качества. Однако потом на сцене появилась эйнштейновская теория относительности. Согласно этой теории только локальное время может быть непосредственно измерено часами. Другими словами, это и есть *собственное время* этих часов или наблюдателя, находящегося в непосредственной близости от них. Время, которое теперь представляет уравнения движения простыми (или, по крайней мере, менее сложными), в расширенной области пространства, включая, например, Солнечную Систему, является просто координатой, выбранной свободно за те ее достоинства, что она просто под руками и никакая

физическая реальность ей не приписывается. Это есть *координатное время* в произвольно выбранной системе пространственно-временных координат. Мы вернемся к этим вопросам в главе 3. Их возрастающая практическая важность связана с прогрессом в развитии времени. Давайте просто иметь ввиду то, что для наших общих целей существуют теоретические соотношения между собственным временем наблюдателя и временем в различных системах координат, которые он должен определить.

Следовательно, метрология времени состоит в обеспечении практически приемлемыми представлениями:

- для единицы собственного времени, используемой любым наблюдателем для локальных экспериментов и наблюдений (даже если они относятся к удаленным объектам);
- для различных координатных времен, полезных астрономам, геодезистам, инженерам, занятым определением орбит спутников или, говоря короче, всем тем, кому требуются точные модели в пространстве, гораздо большем, чем лаборатория.

Коль скоро мы приняли теоретическую динамическую модель, удовлетворительную в том смысле, что еще не было получено успешных экспериментов, доказывающих ее неправильность, то метрология времени может быть основана или на экспериментальных определениях координированного времени, или на локальных измерениях. Хотя оба метода теоретически эквивалентны, они отличаются по своим результатам до такой степени, что в 1960 г. был поднят вопрос, а не лучше ли основать единицу времени на атомных свойствах (наблюдаемых локально), а определение шкалы времени оставить астрономии. Это есть тот практический аспект, к которому мы сейчас вернемся.

2.4. Два способа измерения времени

Мы больше не будем отличать локальные эксперименты от экспериментов в больших масштабах пространства, т.к. мы приняли, что существует подходящая, всеохватывающая теоретическая канва. Следующее обсуждение должно быть понятно в знакомом контексте классической физики и динамики. В духе этой работы оно относится к основам метрологии: как определить секунду и как определить шкалу времени, используемую в качестве первичного эталона времени. В этой части мы исследуем два способа метрологии времени в зависимости от того или мы начнем с явления, которое каждый может воспроизвести в своей лаборатории, или от шкалы времени, данной нам истинным движением небесных тел.

2.4.1. Использование воспроизводимости

Явление, принятое за эталон, должно быть определено таким способом, чтобы его длительность могла бы рассматриваться как идеальная константа. Это значит, что мы должны быть способны оговорить строго одни и те же условия, иначе мы окажемся в положении, заставляющем нас исключить любую причину, которая может рассматриваться как возмущение. Секунда, определенная как оговоренное число эталонных периодов, является в этом случае идеальным эталоном промежутка времени. Естественно, секунды фактически реализуемые, отличаются от идеальных секунд. Эти отличия могут быть предельно уменьшены путем усреднения во времени. Но так как они не независимы для последующих секунд и они не усредняются к нулю, то существует оптимальный период усреднения, за пределами которого ничего не может быть улучшено. (Мы вернемся к этим статистическим проблемам, когда будем обсуждать стабильность осцилляторов в главе 5). Возможно так же усреднение секунд, реализованных в разных лабораториях.

Как только такие усреднения были приняты, так сразу же в итоговой секунде появились неопределенности от двух источников:

- экспериментальные неопределенности каждой лаборатории, с которыми обычно ничего нельзя поделать потому, что сам по себе эксперимент был в прошлом, и, что единственно возможные поправки связаны с обработкой, ошибки которой были выявлены позже, что случается редко;
- неопределенности из-за примененного способа усреднения, включая неопределенности из-за алгоритмов и сличений эталонов (например, проблема статистических весов).

В любом случае эти неопределенности остаются ограниченными. С годами они постепенно уменьшаются благодаря техническому прогрессу в приборостроении и лучшему пониманию основ явлений, позволяющих избавиться от причин, вызывающих возмущение.

Давайте рассмотрим, как можно сделать шкалу времени. Такая шкала получается путем сложения секунд в то, что иногда называется *интегрированная шкала времени.* Но коль скоро накапливаются секунды, также накапливаются и их ошибки. Потенциальное расхождение между реализованной шкалой и идеальной шкалой, созданной накоплениями идеальных секунд, непрерывно возрастает. В действительности расхождение вскоре принимает перекошенные пропорции. Например, в секунде, кото-

рая была получена в 2000 г. с неопределенностью порядка 1×10^{-15} с, расхождение может казаться совершенно ничтожным. Однако отсюда следует, что интегрированная шкала времени может отклониться на 1 микросекунду за 30 лет по сравнению с идеальной шкалой, которая была бы получена путем накопления строго точных (идеальных) секунд. Например, мы увидим позже, что прибытие импульсов от пульсара может колебаться в пределах 0,3 мкс. Расхождение между практической шкалой времени и идеальной шкалой времени может ограничить применение шкал времени, основанных на вращении этих объектов.

Давайте так же отметим два последствия выбора интегрированной шкалы времени как мирового эталона:

- Для того чтобы убедиться в долговечности шкалы и выбрать наилучшие из существующих эталонов для интервалов времени должно быть предпринято усреднение. Другими словами, используя техническую терминологию, мы должны установить алгоритмы для шкал времени. Различные алгоритмы, обеспечивают нас взаимно расходящимися шкалами времени и нет реального способа указать какой из алгоритмов лучше, а т.к. эталонная шкала времени должна быть единой, то результаты полученные различными институтами, должны быть согласованы.
- Людям свойственно ошибаться и ошибки эти тоже интегрируются. Они могут происходить из-за экспериментальных возмущений и сличений эталонов и влиять на передачу или централизованную обработку данных. В качестве практической меры запрещается изменять шкалу эталонов задним числом (все данные, приписанные прошлым событиям, могут быть фальсифицированы и такие подделки могут быть навсегда вписаны в историю).

И именно это явилось веской причиной того, что интегрированная шкала была непопулярна в 60-х годах, когда только начали устанавливаться первые экспериментальные шкалы, основанные на постулате воспроизводимости, примененном к частотам атомных переходов. Но преимущества атомного времени над астрономическим были подавляющими. И в 1971 г. атомная шкала, спустя 16 лет после создания первого цезиевого атомного стандарта, стала официальным временем, а цезиевый стандарт стал рабочим эталоном. Несмотря на нетерпение хранителей атомного времени, эти 16 лет, как все считают, были действительно коротким периодом для такого важного события.

2.4.2. Использование динамической модели

Основное движение небесных тел при их ежедневном прохождении по небу обусловлено вращением Земли. Однако период вращения Земли, хотя и очень слабо, но все же изменяется и поэтому не является строго воспроизводимым. И хотя количественно причины изменения периода известны, это не значит, что мы способны установить удовлетворительную динамическую модель этого явления. Мы вернемся к этому в главе 8, когда будем обсуждать Всемирное Время. А в настоящий момент рассмотрим динамику Солнечной Системы.

То, чего достигла теория – это эфемериды, дающие пространственные координаты небесных тел (планет и их спутников) как функции независимой переменной – времени, координированного времени теории. Для наших настоящих целей пространственные координаты должны быть геоцентрическими, в невращающейся системе, такой чтобы наблюдатель мог использовать только небольшие поправки на свое положение и скорость, чтобы учесть орбитальное движение и вращение Земли.

Измерение времени в этом случае очень простое, по крайней мере, в принципе. Координаты любого тела Солнечной Системы измеряются в некоторые моменты, которые мы засекаем с помощью часов. Тогда эфемериды дают данные, соответствующие этим координатам и выраженные через независимую переменную – время. Таким образом, из теории мы получаем время в форме поправок к показаниям часов. Операция повторяется столько раз, сколько надо (или возможно), а часы используются только для интерполяции между астрономическими наблюдениями.

Первичные данные таким образом образуют шкалу времени. Единица времени может быть определена как текущий интервал градуировки этой шкалы. Для возможности уточнения эфемерид мы можем также определить единицу времени как единичный временной интервал между двумя специфическими астрономическими событиями. В таком случае это есть просто одна из двух констант интегрирования уравнения движения (другая фиксирует момент начала отсчета времени) и эфемериды обеспечивают единственную физическую реализацию. Однако нечего и говорить, что это долгий путь от принципа до практики. Если мы действительно хотим получить единственную шкалу времени, мы должны учесть множество условий. Измерения должны относиться к одиночному астрономическому объекту, например Солнцу, Луне или чему-нибудь еще. Одна

из частных эфемерид должна служить в качестве опорной по соглашению. Различные измерения, относящиеся к единственным часам, должны быть проанализированы и обработаны организацией, которая отвечает за выдачу официального времени.

Учитывая все эти предосторожности, мы должны поднять вопрос об ошибках, влияющих на произведенную шкалу, т.е. ее отклонение от шкалы теоретического, идеального времени, за исключением линейной функции от времени, которая соответствует произвольному выбору единицы и начала отсчета времени.

- Наблюдательные ошибки, связанные с данными, ограничены (в противоположность с тем, что случается с интегрированным временем). Они могут быть уменьшены при дальнейшем техническом прогрессе. Отметим, что предпочтительнее выбирать небесные тела с быстрым истинным движением.
- Динамическая модель, в настоящее время – общая теория относительности, может быть поставлена под сомнение.
- Применение математической модели навлекает на себя ошибки, связанные с ограничением в количестве членов математических рядов или выполнением численного интегрирования. Однако следует заметить, что, так как наблюдения непрерывно выполняются, то существует возможность улучшения теории и применения ее таким способом, чтобы получить обновленную шкалу.
- И, наконец, наиболее важно то, что эфемериды подвержены ошибкам из-за начальных условий интегрирования, т.е. ошибкам, связанным с начальными положением и скоростью.

Последней в этом списке возникает логика. Как мы можем получить скорости, если мы не знаем, как сначала установить способ измерения времени? Именно этот вопрос занимал умы выдающихся астрономов, установивших *Эфемеридное Время* в первой половине двадцатого столетия. Не вникая в сложности их работы, мы можем сказать, что главная причина их успеха состоит в том счастливом обстоятельстве, что период вращения Земли вовсе не такой уж плохой эталон для интервала времени (включая относительные вариации на уровне от 10^{-7} до 10^{-8}). Благодаря этому счастливому обстоятельству стало возможным установить уравнения движения и довольно точно измерить начальные скорости. Более того, так как орбитальные движения почти периодические или что-то в этом роде, то как только возникают вопросы по измерению времени, в теории привлекаются члены следующего порядка малости, неточность которых оказывает незначительный эффект на конечный результат.

В любом случае динамическое измерение времени покоится на части прошлых измерений, а их неопределенности затем скажутся в будущем. Преимущества обеспечения единственной шкалы, для которой была выбрана опорная эфемерида, до некоторой степени смазываются необходимостью ревизии эфемерид при накоплении прошлых наблюдений. Мы увидим в разделе 4.3 «Эфемеридное Время», что и другие проблемы препятствуют принятию этой меры времени в качестве практического мирового эталона.

Тем не менее, реализации времени, основанные на динамике и их долговременные сличения с часами, основанными на атомных свойствах, дают возможность проверки как динамической, так и атомной теорий. Например, спрашивается, действительно ли гравитационная постоянная является константой.

Глава 3

Измерение времени и теоретические модели

Еще в начале шестидесятых годов, через пятьдесят лет после публикации эйнштейновской общей теории относительности (ОТО), измерение времени, а в действительности и вся метрология, все еще не учитывали эту теорию. Конечно, некоторые эксперименты были направлены на проверку ОТО, но говоря словами С.Вилла [3.1] «отношение к этой теории кажется таким, что в то время как она несомненно важна для фундаментальной теории природы, ее исследовательские контакты ограничены классическими проверками и космологией. Как следствие, ОТО была выброшена на обочину главного направления в физике». Только выводы специальной теории относительности обычно принимались во внимание.

Однако, со временем, интерес к релятивистской теории проснулся снова и начал быстро возрастать. Причина кроется в разработках по теоретической астрофизике и космологии, подтвержденных прогрессом в наблюдениях. Мы увидим, что этот экспериментальный прогресс в значительной степени обязан быстрому распространению время-частотных методов измерений. Одновременно поэтому же появилась действительная необходимость в разработках теоретических основ измерения времени, более удовлетворительных, чем те, которые обеспечиваются классической механикой и специальной теорией относительности.

Однако, хотя огромное число исследователей, ученых и инженеров должно теперь принимать во внимание релятивистские эффекты в своей работе, ОТО от этого не стала более знакомым инструментом. В действительности, связь между метрологией и ОТО всегда была трудной. Реальная причина этого барьера лежит в том факте, что ОТО основывается на дифференциальной геометрии, в то время как метрология имеет дело с конечными величинами. Применение ОТО к измерению времени, или вообще говоря к метрологии и макроскопической физике, приводит нас к фундаментальным принципам теории. Это то, с чем мы сейчас должны хотя бы бегло ознакомиться.. Мы рассмотрим пути и концепции, по которым шло развитие от классической модели до модели, предложенной в

ОТО. Читатель может более детально ознакомиться с этими вопросами, например в [3.2]. Другим полезным источником является доклад [3.3], где обсуждается вопрос о применении ОТО к метрологии и который вдохновил на написание части этой главы и главы 5. Тем, кто найдет объяснение этой теории приводящим в уныние или утомительным, советуем отправиться к выводам в конце этой главы, где они найдут основные моменты, которые они должны знать, чтобы понять эту книгу.

Релятивистская трактовка измерений времени вовлекает пространственные координаты. В этой главе нам надо определиться с необходимыми небесными и земными координатными системы. Физическая реализация этих систем в виде набора числовых значений координат для выбранных объектов известна как *система опорных координат*. Метод конструирования таких систем будет рассмотрен в главе 8. А пока мы укажем их точность и их официальный статус.

3.1. Классическая модель и абсолютное время

Начнем со строгого определения *события*. Это есть строго геометрический объект, а именно, точка в пространстве-времени, которая существует сама по себе. Первый контакт между Армстронговской бутсой и лунной поверхностью является примером события, до тех пор, пока мы пренебрегаем площадью подошвы и временем, требуемым для того, что бы придавить верхние слои пыли. Событие может быть идентифицировано описанием, как представлено в этом примере, и это описание определяет его внутреннюю сущность. Однако часто полезно приписать событию числовой код, а именно его четыре пространственно-временных *координаты*, в *координатной системе*. Последняя выбирается конкретно, исходя из того, что она должна быть наиболее подходящей для поставленной цели и находиться под рукой. В примере, процитированном выше, архивы сохранили *дату* (координату времени) во *всемирной координированной шкале времени*, тогда как телевизионный обозреватель был бы более заинтересован в официальном местном времени. Селеноцентрические пространственные координаты были бы в наибольшей степени подходящими для описания места, где видны отпечатки следов, но инженеры из НАСА должны придерживаться траектории Аполло в геоцентрической системе.

Координаты поэтому являются существенным дополнительным инструментом, но они служат только для того, чтобы маркировать события, а не определять их.

В классической модели пространства-времени предполагается, что существуют подгруппы *одновременных событий*. Одновременность рассматривается как физическое свойство, на которое нельзя повлиять каким

либо способом. Частный срез пространства времени может быть обозначен датой t_U в шкале времени U, или датой t_V в другой шкале V, но соотношение $t_V = t_V(t_U)$ является функцией только t_U. В каждом сечении одновременности пространственные координаты, в рамках эвклидовой геометрии, могут быть выбраны произвольно. Одним из первостепенных последствий важности такого представления является то, что интервал между событиями A и B рассматривается как внутренне присущее свойство этой пары событий. Как только выбрана ось временной координаты, так сразу же однозначно выражается разность между датами A и B и нет необходимости в дополнительной информации. Это кажется столь очевидной истиной, что даже исключительно точные измерения, столь обычные в практике современной физики, основываются на этом замечательном свойстве классического пространства-времени.

Сложив вместе одновременные сечения абсолютного пространства-времени, мы теперь обратимся к динамической модели. Она основана на хорошо известных принципах классической механики. Ограничимся первым законом Ньютона, так как этого достаточно, для того чтобы понять, что такая модель может нам сказать.

Согласно этому закону изолированная материальная точка движется по прямой линии с неизменной скоростью. Нечего и говорить, что это идеализация экспериментов или наблюдений, проделанных в грубой форме. Изолированная точка может быть представлена меткой на горизонтальной плоской поверхности, время может измеряться часами, работа которых основана на понятии воспроизводимости (например, маятник), длина может измеряться жестким стержнем, и мы увидим, что движение приблизительно равномерное. Но что в действительности говорит нам первый закон? Определяет ли он изолированную точку или равномерную скорость? Какие измерительные эталоны должны использоваться, чтобы оценить равномерность? Мы должны признать, что эти вопросы поставлены зря. Закон призывает нас только принять «хорошие» координаты, такие, чтобы пространственные координаты x^i (i = 1, 2, 3) были линейной функцией координаты времени t. Классическая модель, как и любая другая модель, является порождением человеческой мысли. Она логически развивается из минимального набора постулатов, как, например, классические законы механики. Она мотивируется опытом, но, будучи однажды созданной, существует независимо от него. И задача физиков в том и состоит, чтобы сконструировать координаты и измерительные эталоны для времени, длины, массы и угла в соответствии с этой теорией. С этими конструкциями физик может потом проверять, что все измеренные явления соответствуют в пределах экспериментальной неопределенности теоретическим предсказаниям, основанным на этой модели. Если же появляется какое-нибудь разногласие, то это означает, что либо коор-

динатная система и измерительные эталоны были выбраны плохо, либо теорию неправильно применили (не включили возмущающие эффекты, допустили ошибки вычисления и так далее), или даже, как последнее прибежище, что сама теория неудовлетворительна.

Ни на первый закон движения, ни на другие законы классической динамики не влияют преобразования координат вида:

$$t \to t' = t,$$
$$x^i \to x'^i = x^i + v^i t, \tag{3.1}$$

где v^i три произвольных константы. (Они также не изменяются при тривиальном изменении размера единицы и начала координат). Тогда как абсолютное время является существенным элементом, абсолютное пространство, очевидно, не является необходимым понятием в этой динамической модели. Набор систем пространственных координат, определяемых преобразованием (3.1) составляет галилеевские координатные системы и инвариантность динамических законов в этих системах является галилеевским принципом относительности.

Особо подчеркнем тот факт, что (пространственно-временная) координатная система *определяется* уравнениями движения. Именно простейшая форма этих уравнений приводит к галилеевским координатным системам и абсолютному времени. Утверждение, как это иногда делают, что оси галилеевской системы неподвижны относительно звезд и квазаров, является некорректным. Это приближенно наблюдаемая особенность. Если мы скажем, что пространственные оси и абсолютное время градуируются в метрах и секундах в соответствии с определениями этих величин, как это принято в системе СИ, мы ни в коем случае не определяем градуировку этих осей. Это только предполагаемое свойство, которое приблизительно подтверждается опытом.

3.2. Специальная теория относительности

3.2.1. Физика в инерциальных системах

В противоположность динамике, оптика и электродинамика оказываются в таком положении, что они могут определить абсолютное пространство через измерение скорости света. Кризис, который ведет к отказу от абсолютного пространства, начался с экспериментов Физо (1849 г.) и Фуко (1850 г.). Он достиг своего предела с максвелловской электродинамикой (1869 г.) и попытками измерить орбитальную скорость Земли Майкельсоном (1881 г.), а затем Майкельсоном и Морли (1887 г.). Уравнения Максвелла не инвариантны к галилеевским преобразованиям (3.1) и подразумевают, что свет имеет постоянную скорость в абсолютном пространстве. Однако все экспериментаторы воздерживались от одобрения идеи абсолютного пространства.

Эйнштейновская специальная теория относительности (1905 г.) разрешила этот кризис. В его теории пространственные опорные координаты являются все еще галилеевскими координатами, или *инерциальными координатами*, рассматриваемыми в пространстве без гравитационного поля. В этих рамках все движется равномерно и прямолинейно по отношению друг к другу. В последующем мы не будем пытаться описать специальную теорию относительности, начиная с минимального набора постулатов, так как это приведет к уровню утонченности, несоответствующему предстоящим целям. Мы рассмотрим в наиболее простом виде принципы и следствия теории, подчеркивая только те особенности, которые в наибольшей степени связаны с метрологией. Цена, которую надо платить за эти упрощения, состоит в определенной многословности и повторяемости.

(a) Инвариантность скорости света

Для каждой инерциальной системы скорость света в вакууме c является максимальной скоростью любого сигнала или частицы. Эта скорость изотропна и неизменна во всех координатах.

Вернемся к проблеме измерительных эталонов, которая нас уже беспокоила, когда мы рассматривали первый закон Ньютона. Как только вопрос касался времени, мы предполагали, что часы, основанные на воспроизводимости, способом, обсужденным в главе 2, годятся для проверки непосредственно или косвенно того, что скорость света c является константой. То же самое относится и к атомным часам, производящим секунду СИ. Что же касается длины, то исторически эталоном был металлический стержень. Но в 1960 г. метр был переопределен, как определенное число длин волн λ_0 атомного перехода криптона. При этом негласно подразумевалось, что частота ν_0 этого перехода и скорость света были константами. Любые попытки проверить постоянство c с использованием атомных часов и метра, определенного таким способом, были совершенно бесполезны. На 17-й Генеральной Конференции по Мерам и Весам в 1983 г. было ясно обосновано определение метра на постоянстве скорости света c (смотри раздел. 4.4).

(b) Геометрия пространства

В каждой инерциальной системе пространство является эвклидовым. *Расстояние D* между двумя фиксированными точками в какой-либо системе координат является хорошо определяемой и измеряемой величиной. При этом удобно пользоваться прямоугольными декартовыми координатами x^i такими, что

$$D = [(\Delta x^1)^2 + (\Delta x^2)^2 + (\Delta x^3)^2]^{1/2}, \qquad (3.2)$$

где Δx^i – разности координат между двумя точками. В таких условиях координаты сами по себе являются непосредственно измеряемыми величинами при использовании измерительных эталонов, например, согласно процедуре описанной ниже в (d).

(c) Временная координата и синхронизация

Для того, чтобы определить временную координату t мы можем привлечь группу часов, установленную в этой системе. По определению, когда мы снимаем показания часов C, то это есть *собственное время* τ_C этих конкретных часов. Рассмотрим двое конкретных часов C и D , удаленных друг от друга. Световой сигнал излученный из C в момент времени τ_C (излучение) записывается часами C. Затем он принимается и отражается в D в момент времени τ_D (отражение), записывается часами D и возвращается в точку C в момент времени τ_C (возврат) и записывается часами C. В соответствии с *эйнштейновским принципом синхронизации,* C и D синхронизированы если,

$$\tau_D(отражен) = \frac{\tau_C(излучен) + \tau_C(возврат)}{2}.$$

Из принципов (a) и (b) следует, что в любой инерциальной системе синхронизация часов согласно эйнштейновскому определению подчиняется принципу транзитивности. Это означает, что если C и D синхронизированы и если D и E синхронизированы, то C и E также синхронизированы. Именно поэтому возможно синхронизировать все установленные часы в данной инерциальной системе и принять время τ одних из этих часов (их *собственное время*) как временную координату t.

(d) Пространственные координаты

Из постулата о том, что скорость света c постоянна и из существующего определения метра, расстояние между двумя точками A и B, фиксированными относительно инерциальной системы может быть однозначно измерено через время Δt необходимое для светового сигнала, чтобы пересечь пространство в вакууме от A до B. Пространственные оси градуируются по времени распространения света вдоль этих осей (это распространение происходит по прямым линиям).

(e) Эйнштейновский принцип эквивалентности

Теперь мы можем установить фундаментальный принцип физики, который гласит, что не существует негравитационных явлений, которые можно было бы использовать для того, чтобы отличить одну инерциальную систему от другой. Другими словами, математические модели (неграви-

тационной) физики имеют одну и ту же форму во всех инерциальных системах, если мы используем идентичные измерительные эталоны, установленные в этих системах. Это подразумевает, что константы физики являются инвариантами, если они измеряются с помощью этих эталонов. Этот принцип, устанавливаемый различными путями, часто называется *эйнштейновским принципом эквивалентности.*

3.2.2. Лоренцевское преобразование и инвариантность интервала

Принципы, установленные выше, не совместимы с галилеевским преобразованием (3.1) между координатами двух инерциальных систем. Они приводят к лоренцевскому преобразованию. Если система R' находится в поступательном движении по отношению к системе R с некоторой постоянной скоростью υ измеренной в R, вдоль x^1 оси R, то это преобразование дается

$$\begin{aligned} t &\to t' = \beta\left(t - \upsilon^2 c^{-2} x'\right), \\ x^1 &\to x'^1 = \beta\left(x^1 - \upsilon t\right), \\ x^2 &\to x'^2 = x^2, \\ x^3 &\to x'^3 = x^3, \end{aligned} \tag{3.3}$$

где $\beta = \left[1 - (\upsilon / c)^2\right]^{-1/2}$ и начала координат совпадают в момент $t=0 = t'$.

Лоренцевское преобразование гарантирует инвариантность c (это достаточно очевидно) и инвариантность максвелловских уравнений. Однако это не следует рассматривать как математический артефакт, сконструированный для этого случая. Это фундаментальное свойство природы, и его различные последствия никогда не были в противоречии даже с наиболее точными экспериментами.

Более того, если Δt и Δx^i являются разностями координат между двумя событиями A и B, то величина Δs^2 определяемая из

$$\Delta s^2 = -c^2 \Delta t^2 + \sum_i \left(\Delta x^i\right)^2 \tag{3.4}$$

является инвариантой при изменении координат (3.3). Величина Δs называется *интервалом* между двумя событиями. Уравнение (3.4) дает пространственно-временную *метрику* для специальной теории относительности, известную как *метрика Минковского.* Она показывает, как интервал, который является величиной, связанной с парой событий, может быть измерен безотносительно какой-нибудь системы координат. Конечно, в R, например, мы можем измерить временные и пространственные координаты, исключить Δt и из (3.2), получить величину D. Но эти величины

не имеют универсальной реальности, так как они зависят от выбора координат. О них говорят как о *координатных величинах.*

В каждой инерциальной системе распространение света характеризуется величиной $\Delta s^2 = 0$, а его скорость равна c.

3.2.3. Время в специальной теории относительности

Единица времени, используемая в инерциальной системе, должна быть обеспечена эталонами, т.е. установлена относительно этой системы. В итоге можно сказать, что секунда должна рассматриваться как *единица собственного времени* (подразумевая, что мы имеем в виду собственное время системы, в которой оно используется).

Принцип синхронизации включает в себя часы, установленные относительно каких-то координатных систем. Он зависит от выбранной системы в том смысле, что часы, синхронизированные в R, не будут синхронизированными в R'. Как хорошо известно, это ведет к видимому замедлению времени, наблюдаемому в движущихся часах. Для того, чтобы избежать малейшей неоднозначности полезно указать точно, что означают величины появляющиеся в (3.3). В этих уравнениях t и t' так же как x^i и x'^i являются координатами одного и того же события (существующего само по себе независимо от способа его записи). Поэтому они являются *показаниями,* отмеченными по координатным осям таким путем, чтобы (3.3) выражало соотношение между показаниями в двух различных координатных системах.

Что касается времени, мы видели, что t может рассматриваться как эквивалент показаний τ синхронизированных часов, фиксированных относительно R. Мы можем представить себе большое число часов, фиксированных относительно R′, физически идентичных тем, что находятся в системе R и чьи показания τ' эквивалентны t'.

Два последовательных показания τ_1 и τ_2 часов, установленных в R, соответствуют двум событиям, для которых временные координаты есть t_1 и t_2 в системе R и t'_1 и t'_2 в системе R′, удовлетворяя условиям

$$t_2 - t_1 = \tau_2 - \tau_1, \quad t'_2 - t'_1 = \beta(t_2 - t_1) = \tau'_2 - \tau'_1. \tag{3.5}$$

Так как $\beta > 1$, то наблюдатель, фиксированный в R', приходит к заключению, что собственное время часов в R, которые находятся в движении относительно него, возрастает не так быстро, как собственное время его часов.

В этих рассуждениях не было вопросов, касающихся обмена сигналами между R и R′. Однако, если наблюдатель в R′ принимает частоту часов из R, движущихся с относительной скоростью **v** (с модулем π), с помощью электромагнитных сигналов, то он должен учесть классический эффект Доплера. Соотношение между излученной частотой f, измеренной в R и принятой частотой f', измеренной в R′, есть

$$f' = f\left[1 + \frac{\mathbf{v}\cdot\mathbf{n}}{c}\right](1 - v^2/c^2)^{-1/2}, \tag{3.6}$$

где **n** единичный вектор вдоль направления распространения сигнала.

В дополнение рассмотрим два события A и B такие, что часы могут транспортироваться с постоянной скоростью вдоль прямой линии из A в B. В координатной системе R, связанной с этими часами, учитывая, что $t = \tau$,

$$\Delta s^2 = -c^2\Delta t^2 = -c^2\Delta\tau^2. \tag{3.7}$$

В произвольной инерциальной системе R′ собственное время часов, транспортируемых таким способом от одного события до другого, может быть использовано, чтобы измерить интервал между такими событиями. Он связан с первичными координатами через

$$\Delta\tau^2 = \Delta t'^2 - \frac{1}{c^2}\sum_i \left(\Delta x'^i\right)^2. \tag{3.8}$$

3.2.4. Область применения специальной теории относительности

Может показаться, что специальная теория относительности имеет очень ограниченное применение, так как из нее исключена гравитация и, следовательно, строго говоря, исключено любое распределение масс. Однако она играет одну из ключевых ролей в физике и метрологии. Одной из причин такого положения является то, что мы можем получить хорошее приближение к теории в инерциальной лабораторной системе, находящейся в свободном падении. Другая причина в том, что теория может быть применима даже к лабораторной системе, движущейся с ускорением. В обоих случаях гравитационное взаимодействие между массами в лаборатории (лабораторное оборудование и его пользователи), настолько мало, что им всегда можно пренебречь, кроме, конечно, экспериментов, специально поставленных для изучения этого взаимодействия (например, измерение гравитационной постоянной).

В свободно падающей лабораторной системе ускорение почти полностью компенсируется гравитационными эффектами. Остаются только те эффекты, которые обусловлены неоднородностью гравитационного поля (потому что масса, производящая гравитационное поле, находится не в бесконечности). Остаточный потенциал называется приливным потенциалом по аналогии с потенциалом, который создает приливные эффекты в земных океанах и земных массах во время свободного падения Земли в гравитационном поле небесных тел (главным образом Солнца и Луны). Эти эффекты вблизи поверхности Земли составляют величину порядка 10^{-16} на метр в относительных величинах. Ими нельзя пренебрегать при

использовании лучших атомных стандартов частоты в космических лабораториях. При менее строгих применениях локальная лаборатория с хорошим приближением находится в инерциальной системе, при условии, что система не вращается.[1]

В ускоряющейся, вращающейся лаборатории, такой как лаборатория на Земле, где ускорение появляется из-за гравитации и вращения вокруг земной оси, специальная теория относительности все еще может быть применима. Для того, чтобы это было возможно, разность гравитационных потенциалов Земли в различных местах лаборатории не должна производить какие-нибудь заметные эффекты. Это стандартный случай для типовой комнаты. (Однако, для измерений времени гравитационный сдвиг частоты составляет относительную величину порядка 10^{-16} на один метр высоты, и здесь мы подходим к осознанию того факта, что некоторые атомные часы больше не могут трактоваться как «локальные».) В дополнение, инерциальные эффекты, такие как вес, сила Кориолиса и т.д., должны быть учтены классическим способом. Например, мы можем видеть, что в некоторых атомных часах используется пучок атомов в резонаторе, установленном и закрепленным в лаборатории. В этом случае мы должны принять во внимание кривизну пучка, из-за веса составляющих его частиц, и другие возможные эффекты массы. Но в локальных экспериментах было бы ошибкой принимать во внимание релятивистские эффекты из-за наличия гравитационного поля Земли до тех пор, пока поле может считаться однородным. Следует помнить, что траектории фотонов являются прямыми линиями только в инерциальных системах (это их свойство используется в лазерных гироскопах).

3.3. Общая теория относительности

3.3.1. Обзор

Общая теория относительности (далее ОТО) является сложной теорией, особенно когда ее применяют к исследованиям в астрономии и космологии. И это несмотря на факт, что она основана на очень простых принципах, которые, по счастливому случаю, адекватны проблемам метрологии, рассматриваемым в этой книге.

В то время,как специальная теория относительности применима к пустому и поэтому нереальному миру, ОТО является моделью реального мира, в котором масса и энергия занимают отведенные им места. Это теория гравитации. Однако, гравитационные эффекты не рассматриваются как

[1] Проблема вращения является сложной в теории относительности. Однако она не является критически важной для измерения времени. В этой книге мы приняли кинематическое определение для условий отсутствия вращения, т.е. отсутствия видимого вращения по отношению к наиболее удаленным известным объектам – квазарам.

действие на расстоянии, как это делается в «замороженной» системе классического пространства-времени. Они в большей степени проявляются в локальных геометрических свойствах Риманова пространства-времени, искривленного происутствием масс и энергии.

В ОТО не существует теоретически привилегированная система координат (хотя существуют одни системы, которые более удобны, чем другие, в том смысле, что они приводят к уравнениям, с которыми легче управляться). Пусть x^{μ}, где $\mu = 0, 1, 2, 3$, будут выбранными координатами. Мы сейчас не будем беспокоиться, что они обозначают в терминах времени и пространства. Пространство-время имеет метрику, связывающую разности координат между двумя бесконечно близкими событиями с координатами x^{μ} и $x^{\mu} + {+}\mathrm{d}x^{\mu}$, т.е. с величиной $\mathrm{d}s$, называемой *интервал*, как и в специальной теории относительности, кроме конечно того, что теперь интервал является бесконечно малой величиной. Связь с координатами выражается через

$$\mathrm{d}s^2 = g_{\alpha\beta}\left(x^{\mu}\right)\mathrm{d}x^{\alpha}\,\mathrm{d}x^{\beta}, \tag{3.9}$$

здесь производится суммирование по повторяющимся индексам и $g_{\alpha\beta} = g_{\beta\alpha}$, так что всего мы имеем десять независимых компонент $g_{\alpha\beta}$. Элементы $g_{\alpha\beta}$ являются компонентами *метрического тензора* и они являются функциями координат x^{μ}. Следовательно, переход к другой системе координат оставляет скаляр $\mathrm{d}s^2$ инвариантным и он рассматривается как величина, которую можно измерить с помощью обычных средств, таких как часы и стержни.

В классической модели, а также в специальной теории относительности, «хорошими» координатами считались физические величины, которые могли быть измерены однозначно в любом месте и в любое время. Чтобы мы не предпринимали, координаты в ОТО никогда не будут иметь такое физическое свойство в глобальном масштабе. Для метрологии это свойство является ключевым отличием. Для того, чтобы подчеркнуть этот факт, многие авторы рассматривают координаты просто как числа (т.е. без размерностей в метрологическом смысле), например как телефонные номера [3.4] или номера домов. Но на практике, особенно когда необходимо производить измерения, такой подход трудно воспринимать. Фактически в метрологии нет всецело удовлетворительного решения проблемы размерностей и единиц для координат [3.3].

Так как метрика Минковского доказала свое право на существование в целом ряде успешных локальных экспериментов, то для каждого события x^{α} должна существовать замена координат $x^{\mu} \rightarrow x'^{\mu}$ такая, что метрика принимает форму, аналогичную (3.4) с $x'^{0} = ct'$,

$$\mathrm{d}s^2 = -c^2(\mathrm{d}t')^2 + (\mathrm{d}x'^{1})^2 + (\mathrm{d}x'^{2})^2 + (\mathrm{d}x'^{3})^2. \tag{3.10}$$

Выражение в форме дифференциалов напоминает нам, что это есть приближенное применение специальной теории относительности к про-

действие на расстоянии, как это делается в «замороженной» системе классического пространства-времени. Они в большей степени проявляются в локальных геометрических свойствах Риманова пространства-времени, искривленного происутствием масс и энергии.

В ОТО не существует теоретически привилегированная система координат (хотя существуют одни системы, которые более удобны, чем другие, в том смысле, что они приводят к уравнениям, с которыми легче управляться). Пусть x^{μ}, где $\mu = 0, 1, 2, 3$, будут выбранными координатами. Мы сейчас не будем беспокоиться, что они обозначают в терминах времени и пространства. Пространство-время имеет метрику, связывающую разности координат между двумя бесконечно близкими событиями с координатами x^{μ} и $x^{\mu} + {+}\mathrm{d}x^{\mu}$, т.е. с величиной $\mathrm{d}s$, называемой *интервал*, как и в специальной теории относительности, кроме конечно того, что теперь интервал является бесконечно малой величиной. Связь с координатами выражается через

$$\mathrm{d}s^2 = g_{\alpha\beta}\left(x^{\mu}\right)\mathrm{d}x^{\alpha}\,\mathrm{d}x^{\beta}, \qquad (3.9)$$

здесь производится суммирование по повторяющимся индексам и $g_{\alpha\beta} = g_{\beta\alpha}$, так что всего мы имеем десять независимых компонент $g_{\alpha\beta}$. Элементы $g_{\alpha\beta}$ являются компонентами *метрического тензора* и они являются функциями координат x^{μ}. Следовательно, переход к другой системе координат оставляет скаляр $\mathrm{d}s^2$ инвариантным и он рассматривается как величина, которую можно измерить с помощью обычных средств, таких как часы и стержни.

В классической модели, а также в специальной теории относительности, «хорошими» координатами считались физические величины, которые могли быть измерены однозначно в любом месте и в любое время. Чтобы мы не предпринимали, координаты в ОТО никогда не будут иметь такое физическое свойство в глобальном масштабе. Для метрологии это свойство является ключевым отличием. Для того, чтобы подчеркнуть этот факт, многие авторы рассматривают координаты просто как числа (т.е. без размерностей в метрологическом смысле), например как телефонные номера [3.4] или номера домов. Но на практике, особенно когда необходимо производить измерения, такой подход трудно воспринимать. Фактически в метрологии нет всецело удовлетворительного решения проблемы размерностей и единиц для координат [3.3].

Так как метрика Минковского доказала свое право на существование в целом ряде успешных локальных экспериментов, то для каждого события x^{α} должна существовать замена координат $x^{\mu} \rightarrow x'^{\mu}$ такая, что метрика принимает форму, аналогичную (3.4) с $x'^{0} = ct'$,

$$\mathrm{d}s^2 = -c^2(\mathrm{d}t')^2 + (\mathrm{d}x'^{1})^2 + (\mathrm{d}x'^{2})^2 + (\mathrm{d}x'^{3})^2. \qquad (3.10)$$

Выражение в форме дифференциалов напоминает нам, что это есть приближенное применение специальной теории относительности к про-

странству с конечными размерами. Математически эта замена координат возможна благодаря симметрии $g_{\alpha\beta}$ вместе с условием симметрии на собственные значения матрицы, которую $g_{\alpha\beta}$ формирует. Фактически существует бесконечное множество первичных координат, которые удовлетворяют этим условиям таким образом, что выбранная локальная система координат может иметь любую ориентацию и любую скорость.

Выражение (3.10) приводит к исключительно важной особенности в измерениях времени. Траектория в пространстве-времени, или *мировая линия* часов с собственным временем, обозначенным через τ, описывается последовательными значениями всех четырех координат. При бесконечно малом перемещении часов локальная координатная система может быть прикреплена к ним таким образом, что $\mathrm{d}x'^{i}$ обращается в нуль для $i = 1, 2, 3$. Тогда, аналогично (3.7), мы имеем

$$\mathrm{d}\tau^2 = -c^{-2}\mathrm{d}s^2 \tag{3.11}$$

Если разделенные конечным интервалом события A и B таковы, что часы могут транспортироваться из A в B, то возрастание $\Delta\tau$ собственного времени, записанного часами, будет выражаться через

$$\Delta\tau = c^{-1}\int_C \left| g_{\alpha\beta}\mathrm{d}x^{\alpha}\mathrm{d}x^{\beta}\right|^{1/2}, \tag{3.12}$$

где интеграл вычисляется вдоль мировой линии C этих часов. Хотя величина $\Delta\tau$ является и измеряемой, она не выражает какие-либо объективные свойства, присущие паре событий, так как допускается произвольный выбор C.

Элементы $g_{\alpha\beta}$ связаны с распределением вещества и энергии через эйнштейновские уравнения поля, тензорные уравнения, играющие ту же роль, что и уравнения Пуассона в классической динамике. Выраженные через координаты, уравнения Эйнштейна дают шесть независимых дифференциальных уравнений для определения десяти независимых элементов $g_{\alpha\beta}$. Такая недоопределенность отражает тот факт, что мы можем выбрать свободно четыре координаты x^{μ} путем произвольного преобразования координат для какого-нибудь частного решения.

Если мы рассмотрим частицы с пренебрежимой массой, которые не подвергаются каким-либо другим силам кроме гравитационных, (называются как *свободные частицы в контексте этой книги*), то существует другой постулат, согласно которому эти частицы двигаются во времени-пространстве по *геодезическим линиям*. Это мировые линии G, для которых величина l

$$l = \int_G \left| g_{\alpha\beta}\mathrm{d}x^{\alpha}\mathrm{d}x^{\beta}\right|^{1/2} \tag{3.13}$$

имеет экстремум. ОТО является такой гравитационной теорией, которая хорошо подходит, например, для описания движения планет. Мировые линии фотонов характерны тем фактом , что ds^2 всегда равно нулю.

ОТО приводит ко многим предсказаниям, особенно в астрофизике и космологии, и за все время ее существования еще не наблюдалось экспериментальных противоречий с ее выводами. С тех пор как Эйнштейн опубликовал свою работу, было изобретено еще много релятивистских теорий и многие из них проходили тесты на экспериментальную проверку (см. главу 9). Однако ни одна из этих теорий не доказала своей необходимости [3.1]. Для наших скромных метрологических целей мы будем обращаться к ОТО, потому, что это простейшая модель, предсказания которой согласуются с измерениями в границах неопределенности последних.

3.3.2. Пост-ньютоновское приближение

(а) Общая форма метрики

Теперь мы воспользуемся той свободой, которую предоставляет нам теория, для того, чтобы выбрать наиболее подходящие координаты. Мы также сделаем важные допущения, совместимые с лучшими достигнутыми результатами по измерениям времени. Хотя мы только касаемся вопросов метрологии и динамики в Солнечной Системе, будет не очень удобно пользоваться в этой области Вселенной единственной системой координат. Нам потребуется по крайней мере две невращающиеся в пространстве системы, одна с началом координат в барицентре Солнечной Системы, *барицентрическая система,* и другая с началом координат в центре массы Земли, *геоцентрическая система.* Нам также нужна геоцентрическая система, вращающаяся вместе с Землей. Теперь мы опишем способ, каким в 1991 г. была определена невращающаяся система координат международным астрономическим союзом (IAU) [3.5].

Международный астрономический союз рекомендует, чтобы пространственные координаты $x^0 = ct$, x^1, x^2, x^3 были выбраны таким способом, чтобы в каждой координатной системе, центр которой совмещен с барицентром любой системы масс, квадрат ds^2 интервала выражался бы с минимальной степенью отклонения следующим выражением

$$ds^2 = -c^2 d\tau^2 = -\left(1 - \frac{2U}{c^2}\right)(dx^0)^2 + \left(1 + \frac{2U}{c^2}\right)\left[(dx^1)^2 + (dx^2)^2 + (dx^3)^2\right], \quad (3.14)$$

где c – скорость света, τ – собственное время и U – сумма ньютоновского гравитационного потенциала для рассматриваемых систем масс (считающегося равным нулю в бесконечности) и ньютоновского приливного потенциала, производимого телами, находящимися вне системы. Этот по-

тенциал выражается в таком виде, чтобы он исчезал в барицентре. По соглашению этот потенциал считается положительным.

Международный астрономический союз также ставит условия, что для измерения собственного времени и собственной длины (для того, чтобы измерять ds^2), должны использоваться единицы измерения системы СИ, секунда и метр.

Метрика (3.14) является приближенным решением уравнений Эйнштейна. Однако она становится неадекватной для установления точных эфемерид планет. Поэтому на генеральной ассамблее 2000 г. международный астрономический союз рекомендовал принять *гармонические координаты*, дающие возможность представить метрику до более высоких порядков по $1/c$. Он также представил релятивистское выражение для наиболее существенных (скалярного и векторного) потенциалов. Международный астрономический союз также разработал применение этой расширенной метрики ко времени. Однако в этой книге, которая главным образом касается вопросов измерения времени на Земле и около Земли, метрика (3.14) является адекватным приближением новой метрики. В геоцентрической системе она дает относительную ошибку частоты самое большое 10^{-18} на расстоянии 300 000 км от Земли, и эти ошибки значительно меньше погрешностей в современных стандартах частоты. Метрика (3.14) таким образом, будет использоваться в дальнейших разработках и исследованиях.

Отношение U/c^2 всегда мало по сравнению с единицей для нашего современного состояния. На поверхности Земли, например, его величина составляет около $1{,}5 \times 10^{-8}$ в барицентрической системе и $0{,}7 \times 10^{-9}$ в геоцентрической системе. Пространство является квазиэвклидовым и пространственные координаты очень близки к классическим декартовым координатам. Координатное время $t=x^0/c$, близко к собственному времени часов, движущихся в рассматриваемой координатной системе со скоростью, малой по сравнению со скоростью света c. На него ссылаются как на *барицентрическое* или *геоцентрическое координатное время.*

Координаты, определенные таким способом, очевидно имеют размерности длины и времени. Практика метрологии такова, что приписывают одну и ту же единицу СИ ко всем величинам, имеющим одинаковую размерность. Различия между этими величинами могут потом быть выражены через их определения, а не через использование различных единиц. Мы будем следовать этой практике, выражая пространственные и временные координаты в метрах и секундах. Эта практика имеет тот недостаток, что иногда приводит к путанице между *собственными величинами,* которые непосредственно измеряются с помощью стандартов и *координатными величинами* (координаты сами по себе или величины, сконструированные из них), для которых градуировочная шкала меняет-

ся относительно собственных единиц в зависимости от места и времени. Надо иметь в виду эту существенную разницу между такими величинами. Когда это необходимо, единица на координатной градуировочной шкале, будет определяться как *единица шкалы.*

Мы часто ссылаемся на соотношение между собственным временем часов и координатным временем. Это соотношение основано на выражении связи между $d\tau$ и dt. Мы можем положить

$$\frac{d\tau}{dt} = 1 - h(t, \mathbf{x}), \tag{3.15}$$

где $\mathbf{x}$ представляет триплет x^i (i=1, 2, 3), как напоминание, что это соотношение зависит от всех четырех координат часов. На данной мировой линии $\mathbf{x} = \mathbf{x}(t)$ и мы можем написать, упрощая обозначение

$$\frac{d\tau}{dt} = 1 - h(t). \tag{3.16}$$

Приращение $\Delta(t - \tau)$ за координатный интервал времени (t_0, t) выражается

$$\Delta(t-\tau) = \int_{t_0}^{t} h(t)dt. \tag{3.17}$$

Используя метрику (3.14),

$$h(t) = c^{-2}\left[U(t) + \frac{1}{2}v(t)^2\right] + O(c^{-4}), \tag{3.18}$$

где v является координатной скоростью часов, определяемой из

$$v = \frac{\left[(dx^1)^2 + (dx^2)^2 + (dx^3)^2\right]^{1/2}}{dt} \tag{3.19}$$

и $U(t)$ есть величина гравитационного потенциала в месте нахождения часов. Члены порядка $O(c^{-4})$ в (3.18) пренебрежимо малы для наших сегодняшних целей и мы больше не будем их упоминать. Потенциал U и координаты x^i, i = 1, 2, 3 являются функциями от t вдоль мировой линии часов. В интеграле t может быть заменено на τ, т.к. эта замена дает ошибку всего лишь порядка $O(c^{-4})$.

(b) Координатные системы и их реализации

В (3.14) x^α представляет координаты в любой координатной системе. Чтобы избежать путаницы, должны использоваться разные обозначения для каждой системы. Обозначение $t=x^0/c$, x^i будет зарезервировано для невращающихся геоцентрических координат и $t_B = u^0/c$, u^i будет использоваться для барицентрических координат.

Пространственные оси барицентрических координатных систем международный астрономический союз рекомендует располагать в барицентре Солнечной Системы. Они имеют определенное направление по отношению к компактным внегалактическим источникам, известным как *квазары*, с u^3 направленным примерно вдоль оси вращения Земли на 1 января 2000 г. и u^1 направленным примерно в направлении равноденствия на эту дату.

Пространственные оси невращающейся геоцентрической координатной системы располагаются в центре масс Земли (включая газовую оболочку). Они имеют такие же направления по отношению к квазарам, как и в барицентрической системе. В метрике, выражаемой в виде (3.14), U представляет гравитационный потенциал Земли, к которому должен быть добавлен приливной потенциал Солнца – Луны. Преобразование координат между барицентрическими и геоцентрическими координатными системами усложняется при строгом рассмотрении [3.6]. Однако такая степень строгости нам здесь не будет необходима.

Нам так же нужна координатная система X^i, вращающаяся вместе с Землей. Она прикреплена глобально к литосфере с учетом модели движения тектонических плит (несколько сантиметров в год). Ось X^3 указывает направление, близкое к направлению оси вращения Земли. А ось X^1 определяет положение географической долготы таким образом, чтобы долгота Гринвического меридиана была примерно равна 0. Эта система выводится из невращающейся геоцентрической системы путем пространственного вращения, которое учитывает движение осей Земли при вращении как в пространстве, так и при собственном вращении. Геоцентрическое координатное время t одинаково во вращающейся и в невращающейся системах. Во вращающейся геоцентрической системе метрика (3.14) принимает формы

$$\mathrm{d}s^2 = -\left(1-\frac{2U}{c^2}\right)c^2\mathrm{d}t^2 + \left(1+\frac{2U}{c^2}\right)\left[\mathrm{d}r^2 + r^2\mathrm{d}\phi^2 + r^2\cos^2\phi(\omega^2\mathrm{d}t^2 + 2\omega\mathrm{d}t\mathrm{d}L + \mathrm{d}L^2)\right] \tag{3.20}$$

где:

- ϕ – есть геоцентрическая широта (угол между геоцентрическим направлением места и плоскостью экватора),
- L – долгота (положительная по направлению к востоку),
- r – расстояние в геоцентрических координатах

$$r = \left[(X^1)^2 + (X^2)^2 + (X^3)^2\right]^{1/2},$$

- ω – угловая скорость вращения Земли (постоянная величина $\omega = 7{,}292115\times$ $\times 10^{-5}$ рад/с является достаточно хорошим приближением),

- U – гравитационный потенциал Земли плюс приливной потенциал, как в невращающейся системе.

Теперь мы рассмотрим, как эти координатные системы могут быть установлены практически. Что касается времени, то международный астрономический союз выбрал начало для t_B и t. Каждая из этих шкал определяется затем однозначно, т.к. единица шкалы определяется путем использования единиц СИ для интервала ds. На эти шкалы ссылаются как на *барицентрическое координатное время* (TCB) и *геоцентрическое координатное время* (TCG). Чтобы их не путать с соответствующими, чисто теоретическими понятиями, они обозначаются заглавными буквами и акронимами. Физические реализации TCB и TCG могут быть получены из международной атомной шкалы TAI способом, объясненным в главе 7.

Что касается пространственных координат, то, выбрав однажды собственную длину-метр мы также определяем градуировку координатных осей.[2] Остается только сориентировать оси.

Физически барицентрическая система координат реализуется путем присвоения неизменных угловых координат сверхудаленным галактическим источникам, известным как квазары. По международному соглашению используется *международная небесная опорная система координат (ICRF)*. Она заключает в себе список прямых восхождений и склонений около 600 квазаров, опубликованных Международной Службой Вращения Земли (IERS) [3.7]. Неопределенности в координатах составляют от 0,0002″ до 0,0005″. Как уже отмечалось, геоцентрическая система сориентирована таким же способом.

Что касается пространственных координат во вращающейся геоцентрической системе, то физическая реализация, принятая по соглашению, называется *международной земной опорной системой координат* (ITRF). Она включает в себя список координат для фиксированной опорной даты и скоростей примерно для 200 пунктов разбросанных по всей Земле. Величины неопределенностей для координат выражаются сантиметрами [3.7]. Как и международная небесная опорная система координат ICRF, международная земная опорная система ITRF также устанавливается Международной Службой Вращения Земли IERS. Эта служба постоянно поддерживает и уточняет ее. Следует также заметить, что ITRF официально является первичным геодезическим стандартом, по которому постепенно равняются все геодезические системы. Вращение ITRF по отношению к ICRF описывается параметрами вращения Земли.

---

[2] Для их собственных специфических нужд астрономы используют также астрономическую единицу длины, приблизительно равную $1{,}5 \times 10^{11}$ м. Ее соотношение с метром определяется экспериментально. Астрономическая единица может также рассматриваться как собственная единица и использоваться как основа для градуировки пространственных осей.

(c) Соотношения между геоцентрическим координатным временем и собственным временем

В невращающейся геоцентрической системе соотношение между собственным временем часов τ и геоцентрическим координатным временем t дается выражениями (3.17) и (3.18).

В координатной системе, вращающейся вместе с Землей, выражение для $h(t)$ имеет следующий вид:

$$h(t) = c^{-2}\left[\hat{U}_{\mathrm{g}} + \Delta\hat{U}(t) + \frac{1}{2}V(t)^2\right] + 2c^{-2}\omega\frac{\mathrm{d}A_{\mathrm{E}}}{\mathrm{d}t}\,, \qquad (3.21)$$

где:

- $\hat{U}_{\mathrm{g}}$ – потенциал U плюс потенциал центробежных сил на поверхности геоида (геоид – полностью эквипотенциальная поверхность), с величиной $\hat{U}_{\mathrm{g}} = 6{,}263686 \times 10^7$ м^{2}с$^{-2}$,
- $\Delta\hat{U}$ – разность между общим потенциалом (включая потенциал центробежных сил) и $\hat{U}_{\mathrm{g}}$ в месте расположения часов,
- V – координатная скорость часов по отношению к Земле,
- A_{E} – площадь проекции на экваториальную плоскость поверхности, описываемой вектором, берущим начало из центра масс Земли и оканчивающимся на часах. Площадь измеряется во вращающейся системе и считается положительной, если часы движутся на восток.

Соотношения (3.18) и (3.21) используются часто. Мы применим их, в частности, для сличения времени и частоты далеко разнесенных стандартов, а также для установки мирового опорного времени. Вышеуказанные соотношения достаточно точны для этих целей с учетом неопределенностей современных стандартов, при условии, что U и $\Delta\hat{U}$ оценены правильно.

(d) Оценка гравитационного потенциала Земли

На поверхности Земли лунно-солнечный приливной потенциал приводит к относительному сдвигу частоты на величину порядка 10^{-17}, который является пренебрежимо малым. Однако он играет возрастающую роль при удалении от Земли и дает относительный сдвиг частоты до 10^{-15} уже на высоте геостационарного спутника (36 000 км). Мы не будем принимать его во внимание, хотя возможно, это и необходимо для часов, расположенных на борту такого спутника. Заинтересованный читатель может посмотреть ссылку [3.6].

Рассмотрим теперь вклад для U, производимый земным потенциалом U_{T}. Его оценка приводит метрологов в область, в какой-то мере им незнакомую и часто создающую для них определенные трудности.

На расстояниях, типичных для искусственных спутников Земли, требуется разложение земного потенциала на сферические гармоники. Границы точности определяются неопределенностями спутниковых орбит. В лучших случаях они соответствуют относительным неточностям частоты на уровне между 10^{-17} и 10^{-18}.

Если мы рассматриваем земные часы, или часы на самолете, у которых допускается относительная неопределенность частоты порядка 10^{-14}, то достаточно ограничиться первыми членами разложения сферического гармонического потенциала, приняв

$$U_{\text{т}} = \frac{GM_{\text{E}}}{r} + \frac{J_2 GM_{\text{E}} a^2 (1 - 3\sin^2\phi)}{2r^3}, \tag{3.22}$$

где символы и их величины (точные, по крайней мере, до последнего десятичного разряда) объясняются следующим образом:

- GM_{E} – произведение гравитационной постоянной на массу Земли с величиной
$$GM_{\text{E}} = 3{,}986\,004\,42 \times 10^{14}\ \text{м}^3\ \text{с}^{-2};$$
- r – геоцентрическая координата расстояния в метрах;
- J_2 – коэффициент квадрупольного момента Земли с величиной $J_2 = 1{,}082\,636 \times 10^{-3}$;
- a – радиус Земли на экваторе, $a = 6\,378\,137$ м;
- ϕ – геоцентрическая широта.

В (3.21) $\Delta\hat{U}$ дается выражением:

$$\Delta\hat{U} = U - \hat{U}_{\text{g}} + \frac{\omega^2 r^2 \cos^2\phi}{2}, \tag{3.23}$$

где последний член представляет потенциал центробежных сил.

Если известна высота над геоидом h (не путать с функцией h определенной ранее), то $\Delta\hat{U}/c^2$ может быть получена из [3.8]

$$\frac{\Delta\hat{U}}{c^2} = -1{,}08821 \times 10^{-16} h - 5{,}77 \times 10^{-19} h \sin^2\phi + 1{,}716 \times 10^{-23} h^2, \tag{3.24}$$

для h в метрах.

Неточность составляет менее 10^{-16} для высоты ниже 15 км. Однако, надо осторожно подходить к оценке h, так как ее определение может изменяться в зависимости от используемых уровневых сетей и топографических карт. Безопаснее проконсультироваться с геодезистами. В этом контексте будет полезной ссылка [3.9].

(е) Земное время: другое геоцентрическое координатное время

Если идеальные часы установлены на Земле на уровне геоида, то из (3.21) собственное время τ связано с t через

$$\frac{\mathrm{d}\tau}{\mathrm{d}t} = 1 - h(t) = 1 - \frac{\hat{U}_{\mathrm{g}}}{c^2} \approx 1 - 0{,}697 \times 10^{-9}. \tag{3.25}$$

Поэтому оно задерживается на 22 мс в год по отношению к t. С возрастанием высоты, расхождение уменьшается, но очень слабо, согласно (3.24), на 3,4 мкс в год на километр.

Международная атомная шкала TAI была установлена совершенно спонтанно с использованием показаний атомных часов, без каких бы то ни было попыток учесть релятивистские эффекты. Единица ее шкалы, поэтому, очень близка к секунде вращающегося геоида. В 1970 г., когда появилась необходимость релятивистского определения мировой шкалы времени, никто не стал бы рассматривать ни на мгновение введение такого определения, которое бы значило, что все земные часы отставали бы на 22 мс в год. Земное время TT, таким образом определенное, называлось земным динамическим временем (TDT) с 1979 по 1991 г.. Оно включает в себя частотный сдвиг по отношению к геоцентрическому координатному времени таким образом, чтобы единица шкалы TT длилась одну секунду собственного времени на вращающемся геоиде. Мы увидим в главе 7, что TAI является реализацией TT.

Затем было замечено, что запутанность вопроса и временные вариации (например, приливные изменения), присущие определению и реализации геоида являются источником ошибок в определении и реализации TT, а, следовательно, также и в TAI. Во избежание этих трудностей в 2000 г. международный астрономический союз переопределил земное время TT с использованием следующего условия: TT – есть шкала времени, отличающаяся от геоцентрического координатного времени TCG на постоянный коэффициент

$$\frac{\mathrm{d}TT}{\mathrm{d}TCG} = 1 - L_{\mathrm{G}}, \tag{3.26}$$

где $L_{\mathrm{G}} = 6{,}969\,290\,134 \times 10^{-10}$ – определяющая константа. Величина L_{G} была принята равной наилучшей по имеющейся оценке $\hat{U}_{\mathrm{g}}/c^2$ в 2000 г. и неизменной в будущем, хотя величина $\hat{U}_{\mathrm{g}}$ может быть и уточнена. Это новое определение TT не подразумевает изменения скорости TT в 2000 г. Это не имеет практического влияния при обработке данных земных часов, но, в частности, очень важно для будущих космических часов. В дальнейшем обсуждении мы будем помнить о символе $\hat{U}_{\mathrm{g}}$, но определим его как $c^2 L_{\mathrm{G}}$.

Хотя применение ТТ практически очень естественно, оно является источником некоторой сумятицы в динамической астрономии и космической геодезии. Изменение координат, которое оно вызывает, оказывает довольно сложный эффект на оценку потенциала и применяется не всегда корректно. Более того, вплоть до 1991 г. международные организации не давали информации, какие пространственные координаты должны быть использованы.

Пусть $\underline{t}$ есть наземное время ТТ определяемое из

$$\frac{\mathrm{d}\,\underline{t}}{\mathrm{d}\,t} = 1 - \frac{\hat{U}_{\mathrm{g}}}{c^2} \tag{3.27}$$

и из выбора начала координат (см. главу 7). Подставив теперь x^i из их определения через метрику (3.14) для невращающейся геоцентрической системы, получим новую метрику в форме

$$\mathrm{d}s^2 = -\left(1 - \frac{2U}{c^2}\right)\left(1 + \frac{\hat{U}_{\mathrm{g}}}{c^2} + O(c^{-4})\right)^2 c^2\mathrm{d}\underline{t}^2 + \left(1 + \frac{2U}{c^2}\right)\left[(\mathrm{d}x^1)^2 + (\mathrm{d}x^2)^2 + (\mathrm{d}x^3)^2\right] \tag{3.28}$$

Положив

$$\frac{\mathrm{d}\tau}{\mathrm{d}\underline{t}} = 1 - \underline{h}(\underline{t}), \tag{3.29}$$

и сохраняя только члены порядка $O(c^{-2})$,

$$\underline{h}(\underline{t}) = \frac{U - \hat{U}_{\mathrm{g}}}{c^2} + \frac{\underline{v}^2}{2c^2}, \tag{3.30}$$

и где v координатная скорость, выраженная относительно $\underline{t}$. Для применений, рассматриваемых в данном тексте, выражение

$$\underline{h}(\underline{t}) = \underline{h}(\underline{t}) - \frac{\hat{U}_{\mathrm{g}}}{c^2} \tag{3.31}$$

будет достаточно точным.

В 1991 г. международный союз геодезии и геофизики (IUGG) рекомендовал [3.10], чтобы во вращающейся геоцентрической системе сохранялись пространственные координаты X^i и, следовательно,

$$\underline{h}(\underline{t}) = \frac{\Delta\hat{U}}{c^2} + \frac{\underline{V}^2}{2c^2} + \frac{2\omega}{c^2}\frac{\mathrm{d}A_{\mathrm{E}}}{\mathrm{d}\,\underline{t}}, \tag{3.32}$$

с вычислением соответствующей координатной скорости $\underline{V}$ через $\underline{t}$.

(f) Синхронизация и опорное время

В ОТО эйнштейновская синхронизация теряет свою транзитивность и, следовательно, это представляет интерес для решения проблем, с которыми мы встречаемся. Мы будем использовать *координатную синхронизацию*, как определено ниже.

В данной координатной системе два события *координатно одновременны*, если они приписаны к одной и той же дате t в координатном времени системы. Двое часов синхронны, в смысле координатного синхронизма на интервале $t_1 \le t \le t_2$, если их одновременные показания координат равны для всех дат t в этом интервале.

В том же духе, результаты сличений между часами всегда выражаются через разность их показаний в одно и то же координатное время.

Вообще говоря, синхронизация и сличения часов не имеют смысла в других координатных системах (отличных от тех) чем те, для которых эти обозначения приняты. Заметим, однако, что изменения координат типа

$$\begin{aligned} t &\to t' = t'(t), \\ x^i &\to x'^i = x'^i(t, x^1, x^2, x^3) \end{aligned} \tag{3.33}$$

не влияют ни на условия координатной синхронизации, ни на результаты сличения часов (хотя при этом изменяются даты). Это происходит при переходе от t к $\underline{t}$.

Очевидно, для метрологии на Земле или в околоземном пространстве, (например, вне орбит геостационарных спутников), используются вращающиеся или невращающиеся геоцентрические координатные системы. Так как они имеют одинаковое координатное время, выбор между ними является только вопросом удобства. Для земных часов вращающаяся система является более подходящей, в то время как для часов на орбите может оказаться более предпочтительной невращающаяся система. Если однажды мы преуспеем в установке далеко разнесенных лабораторий на Луне, то потребуется определить селеноцентрическое координатное время. Часы, синхронизированные на Луне, не будут синхронизированы в земной системе.

Как только вопрос касается применений, то повсеместное распространение реализованного единственным способом геоцентрического координатного времени, основанного на определении секунды (собственного времени) и единогласно признанная метрика, достигают следующего:

- Это обеспечивает основу для синхронизации часов по всей Земле и в ее непосредственном окружении (например, для искусственных спутников).
- Это обеспечивает временную координату, которая, будучи связанна с соответствующими пространственными координатами, позволяет нам описать орбитальные движения (искусственных спутников или Луны) вокруг Земли.

- Через четырехмерные преобразования обеспечивается практическая реализация других координатных времен, таких как барицентрическое координатное время, необходимое для вычисления орбит планет или траекторий межпланетных зондов.
- Через преобразования (3.17) выводится собственное время, а также секунда СИ, используемая для локальных измерений (собственная секунда).

3.4. Выводы и заключение

В повседневной жизни обычное абсолютное время совершенно удовлетворительно для большинства технических и научных применений. Однако обязанность метрологов состоит в принятии таких теоретических рамок, которые совместимы с точностью и стабильностью лучших стандартов во всем диапазоне пространства, где эти стандарты могут использоваться. Для времени и длины такими рамками является ОТО.

В этой теории существенным является различие собственных физических величин, непосредственно измеряемых с применением стандартов, от координатных величин, которые зависят от произвольного выбора пространственно-временной координатной системы. Причина кроется в том, что невозможно найти координатную систему, в которой градуировочные единицы вдоль осей, называемые как *единицы шкалы*, в каждой точке пространства-времени равные секунде или метру, были бы такими же, как если бы мы получили их в этих точках из определений единиц в СИ. Давайте сначала подведем итоги последствий для времени.

- Единица времени – секунда, необходимая для экспериментов в достаточно малых лабораториях (например, в комнатах обычного размера), должна быть установлена и связана с локальным стандартом, расположенным в лаборатории. Это связано с тем, что может быть гарантирована универсальность локальных законов физики и что это оправдывает применение физических констант в рамках специальной теории относительности (эйнштейновский принцип эквивалентности). Короче говоря, мы говорим, что локальная секунда, полученная таким путем, является *собственной* секундой, (что подразумевает осознание того, что мы говорим о собственном времени лаборатории, в которой эта секунда используется).
- Шкала времени, взятая повсеместно в качестве опорной, должна определяться как координата времени, так как только этот способ позволяет однозначно датировать события. Другими словами это *координатное время*. Пока дело касается земной метрологии, то обычно используется пространственная координатная система, вращающаяся вместе с Землей. Однако, иногда полезна геоцентрическая система,

не вращающаяся в пространстве. Эти две координатные системы, определены таким образом, что они используют одну и ту же временную координату, называемую *геоцентрическим координатным временем*. Соотношение между собственным временем и геоцентрическим координатным временем дается интегралом (3.17), использующим выражения (3.18) или (3.21).

- В теории относительности, синхронизация больше не имеет абсолютного значения. По соглашению, события происходят одновременно в установленной системе координат, если они имеют одну и ту же временную координату. Из этого соглашения следует, что двое часов могут быть сличены по разности их одновременных координатных показаний. Если разность всегда равна нулю, говорят, что часы синхронизированы в том смысле, что они синхронизированы по координате. Для наземной метрологии, этой координатой является геоцентрическое координатное время. Мы видели, что замена координат, в которой t' является только функцией t, не влияет на понятия координатной одновременности или координатной синхронизации.

На практике существуют очевидные причины создания на базе лучшей возможной реализации секунды единственного координатного времени, принятого за стандарт по соглашению. Тогда преобразование этого координатного времени в собственное время в любом частном локальном окружении обеспечивает локальную шкалу времени, проградуированную в секундах. Это то, что теперь достигнуто при использовании международного атомного времени (TAI). Существенно, и это надо иметь ввиду, что хотя TAI градуируется в часах, минутах и секундах, интервал времени между двумя последовательными метками, указывающими секунды TAI (далее называемые как *единицы шкалы* TAI), не равен локальной собственной секунде, за исключением вращающегося геоида.

Те же замечания применимы к единице длины – метру и пространственным координатам. Метр должен рассматриваться как единица собственной длины. В его определении от 1983 г., основанном на определенной соглашением величине скорости света (совместимой с предыдущими экспериментальными величинами), он определяется из собственной секунды путем локального эксперимента. Единица шкалы для пространственных координат хотя обычно и называется метр, не везде равна локальному метру. На практике измерения длины менее точны, чем измерения времени, что приводит к тому, что единица пространственной шкалы и собственный метр не очень то часто должны различаться. Однако это различие начинает иметь важность в глобальной геодезии и оно действительно существенно в некоторых астрономических исследованиях.

Глава 4

Эволюция измерений времени

Когда мы измеряем время так же, как измеряем любую другую физическую величину, то кажется логичным начать с оценки метрологической необходимости в предвидении применений и в последующем поиске способов, как им удовлетворить. Этот подход был действительно испытан на астрономической динамике. Но более часто стандарты были реализованы прежде, чем рассматривалось их использование для того, для чего они создавались. И поэтому их само применение предшествовало официальному признанию стандартов организациями, созданными для работ по метрологии в мировом масштабе. В этом смысле метрология есть странная комбинация прагматизма и строгости.

Основная часть этой книги посвящается представлению снимка измерения времени, как оно практиковалось во время написания. Это контрастирует с настоящей главой, в которой мы описываем как такие измерения развивались как в части техники, так и в части идеи. Эта история отмечена колебаниями, сомнениями и иногда даже несостоятельностью.

4.1. Дата, календарь и час

Дата события в научном смысле представляет собой набор данных, приписанных к временной метке на какой-то ее специфической шкале времени. Традиционно это объединяет некоторые способы определения дней, известных как календарь, вместе с мелкими делениями дня, которые обычно называются часами.

Различные календари (Григорианский, Еврейский, Исламский и т.д.) являются изобретательными схемами использования естественных циклов, имеющих точные астрономические определения:

- истинные *солнечные сутки* это промежуток времени между двумя последовательными проходами Солнца через местный меридиан,
- *тропический год* это промежуток времени между двумя проходами Солнца через точку весеннего равноденствия,
- *лунный месяц* это промежуток времени между двумя последовательными новолуниями.

Мы ограничимся главным образом хорошо известным Григорианским календарем, который находится в хорошем согласии с календарем, основанным на тропическом годе[1].

Сутки подразделяются на часы, минуты и секунды способом, унаследованным от вавилонян [4.1]. Этот способ сопротивляется любым попыткам перевода на десятичную систему. Во Франции 24 ноября 1793 г. был издан декрет с намерением введения десятичного деления суток и *десятичной секунды*, но он был отменен 7 апреля 1795 г., и только для интервалов короче, чем 1 секунда используются десятичные доли, т.е. миллисекунды, микросекунды и т.д.

В этой сложной системе причудливого набора сопротивлений технократическим влияниям, единицы не всегда следовали постоянным соотношениям одна с другой. Добавим еще то, что год может содержать или 365 или 366 суток, и что еще одно оригинальное требование появилось в 1971 г. Оно состоит в том, что сутки, обычно продолжающиеся 86400 секунд, иногда могут насчитывать на 1 секунду больше или меньше во *всемирном координированном времени* (UTC), которое отмеряет нашу жизнь.

Для того, чтобы упростить свою работу астрономы иногда используют *Юлианскую дату* (JD). Она основана на непрерывном счете дней, начиная с 4713 г. до Новой эры. Юлианская дата может быть дополнена также десятичными частями суток. В этой системе сутки считаются с полдня. Например, 1 января 2000 г. в полдень соответствует Юлианской дате JD=2 451 545,0.

В метрологии времени, в изучении вращения Земли и в космических науках часто используется *модифицированная Юлианская дата* (MJD). Она определяется из

$$MJD = JD - 2\,400\,000{,}5. \tag{4.1}$$

Это означает, что MJD=0,0... соответствует 17 ноября 1858 г. в 0 часов. (Заметим, что мы следуем международному применению акронимов. MJD происходит из английского, в то время как международная атомная шкала TAI происходит от французского *Temps atomique international).*

По определению календарь и JD и MJD системы тесно привязаны к изменению дня и ночи. Тем не менее они используются на шкалах времени, основанных на других явлениях, таких как международная атомная шкала, в которой имеется только незначительный дрейф по отношению к истинным солнечным суткам. Неоднозначность исключается при условии, что шкала времени указывается в записях даты.

---

[1] За время от Григорианской реформы в 1582 г. и до 3200 г. Григорианский календарь прибавит грубо на один день по сравнению с календарем, основанном на тропическом годе. Неопределенность в этом определении обусловлена главным образом непредсказуемой неопределенностью во вращении Земли.

4.2. Измерение времени, основанное на смене дня и ночи

4.2.1. Среднее солнечное время

В древние времена день, от рассвета до заката, и ночь, от заката до рассвета, почти неизменно делились на 12 часов каждый. Естественно, что часы имели неодинаковую длительность днем и ночью, за исключением дней весеннего и осеннего равноденствия или на экваторе. И длительность часов изменялась вместе с сезонами и широтой. Эти изменчивые часы, или *сезонные часы,* все еще находились в употреблении в 15 веке, несмотря даже на то, что уже задолго до этого, древние астрономы изобрели и использовали *равноденственные часы*, путем деления видимых солнечных суток между двумя последовательными прохождениями Солнца через местный меридиан на 24 равных периода.

Более точно *истинное солнечное время* (или истинное время) определяется как часовой угол Солнца, т.е. угол между полуплоскостью наблюдательного меридиана и полуплоскостью, определяемой осью вращения Земли и Солнца. Этот угол затем делится на «часы» по 15°. (Астрономы не всегда имеют отчетливое различие между углами и временами, как это было показано тем фактом, что астрономические эфемериды Бюро долготы в Париже называются *Знание Времен* (Connaissance des temps).

Истинное солнечное время наблюдается непосредственно и оно все еще использовалось в ряде стран до начала 20 столетия. Однако это не очень «удобное» время в том смысле, как это надо было понимать из Пуанкаре (см. раздел 2.3). Мы можем сказать, что это не равномерное время. Его неравномерность была известна уже Птоломею (150 г. н.э). Эта неравномерность происходит главным образом из-за элиптичной формы орбиты вокруг Солнца и наклона оси вращения Земли по отношению к эклиптике (плоскости ее орбиты). Она имеет общую амплитуду 30 минут и называется *уравнением времени* (см. рис. 4.1). Астрономы исключают эти вариации шкалы равномерного времени, известного как *среднее солнечное время.* Эта шкала времени, как и истинная шкала времени, связана с местным меридианом. Однако попытки идеализировать систему на этом не остановились. Предельной целью было гарантировать строгую пропорциональность между средним солнечным временем и углом, на который Земля поворачивается вокруг своей оси. В главе 8 мы откроем некоторые тонкости этого требования.

Несмотря на преимущество, которое дает равномерное среднее солнечное время, оно долго рассматривалось только как инструмент защиты астрономов. Сделав однажды свою работу, они проводили такие преобразования, чтобы астрономические эфемериды могли бы выражаться через истинное время. Как пример, среднее солнечное время Гринвича,

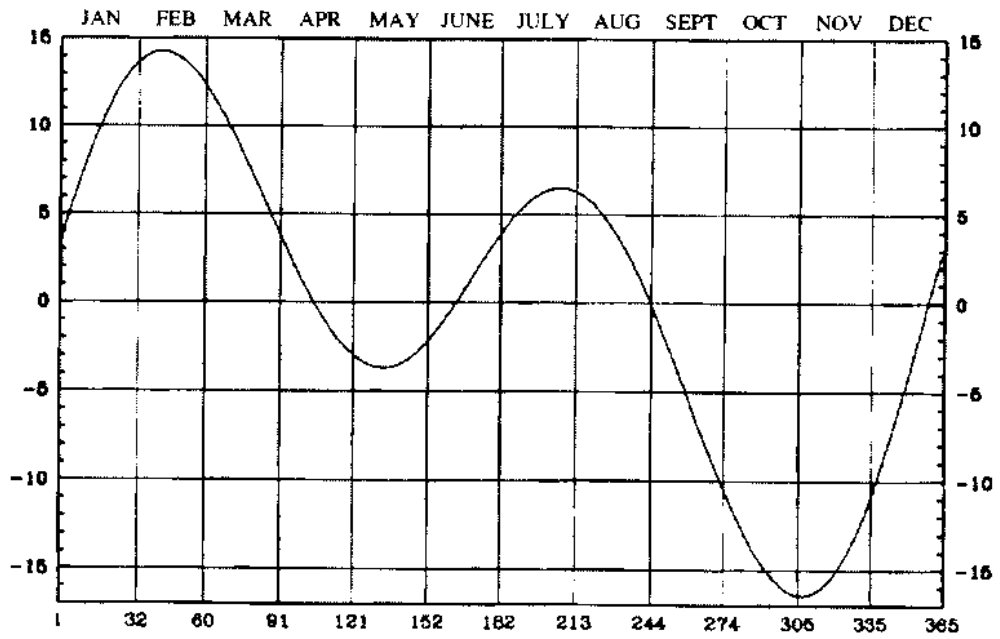


Рис. 4.1. Уравнение времени. Среднее солнечное время минус истинное солнечное время в минутах. (Полученное благодаря бюро долготы, Париж.)

Англия, было введено в *навигационный альманах и астрономические эфемериды* только в 1834 году. На следующий год во Франции Connaissance des temps бюро долготы приняло среднее солнечное время в Париже.

4.2.2. Шкалы всемирного времени и часовые пояса

Появление железнодорожных сетей во второй половине девятнадцатого столетия привело к необходимости использования по крайней мере на национальном уровне, одного единственного определения часа. Затем многие страны приняли определение часа среднего времени меридиана их столиц со сдвигом на 12 часов. Так во Франции актом парламента от 14 марта 1891 года было проголосовано установление часа парижского меридиана.

Однако это решение оказалось неадекватным для некоторых стран, занимающих территории, простирающиеся в большом диапазоне долгот. Идея деления времени таких стран на часовые пояса, таким образом, чтобы время часовых поясов отличалось на целое число часов и чтобы солнечный полдень всегда приходился на 12 часов, была, повидимому, предложена Х.Даудом в США в 1870 г. [4.1]. Вскоре она была опробована практически инженером канадских железных дорог С.Флемингом.

Использование среднего солнечного времени начало таким образом распространяться, хотя все еще и отсутствовал единый мировой стандарт. Соглашение об унификации времени в мировом масштабе было принято в октябре 1884 г. на международной конференции, которая собралась в Вашингтоне «для принятия единственного начального меридиана и всемирного времени» [4.2]. Гринвичский меридиан был предложен

как естественный выбор, потому что он уже был принят как начальный для долготы в большинстве морских лоций, и, действительно, это предложение было одобрено почти единогласно большинством национальных представителей. Всемирное время стало, таким образом, определенным через меридиан. Более того, на конференции было оговорено, в противоположность общей астрономической практике счета суток с полдня, что всемирные сутки должны начинаться в полночь начального меридиана и что часы должны иметь счет с 0 до 24.

В последовавшие затем два десятилетия часовые пояса были связаны с всемирным временем и распространены на всю Землю. Первоначально планета была разделена на 24 временных пояса, каждый из которых занимал 15° долготы, с центром первого пояса на Гринвиче. Ось каждого пояса расположена на *поясном меридиане* и, таким образом, установлено *поясное время* в каждом поясе. Любая страна, занимающая небольшое число зон, могла принять время зоны, в которой находилась ее столица. Постепенно страны начали внедрять новую систему. С 1971 г. она начала ассоциироваться со Всемирным Координированным Временем (UTC), которое будет определено в разделе 4.5 и с бóльшей степенью глубины будет обсуждено в главе 7. Например, 9 марта 1911 г. актом французского парламента во Франции было введено время гринвичского меридиана в терминологии, направленной на сохранение национального достоинства: – «Официальным временем Франции....является среднее Парижское время с задержкой на девять минут и двадцать одну секунду». (Интересно отметить несоответствие терминологии, так как среднее время – это время в полдень в 0 часов). Этот акт действовал до 9 августа 1978 г., когда правительственный декрет связал официальное время с UTC.

Стандартная система времени с тех пор потеряла что-то из первоначальной простоты в связи с постоянным сдвигом между официальным временем и временем соответствующего пояса, а также в связи с летним временем. Например, летнее время появилось во Франции в 1916 г. «для того, чтобы противостоять раздражающей тенденции большинства горожан поздно ложиться и поздно вставать» [4.3]. Очень трудно найти разницу между UTC и официально принятым временем в некоторых странах потому, что не существует организаций, отвечающих за достоверность такой информации.

Астрономы были слишком медлительны для того чтобы принять всемирное время на конференции 1884 г. *Connaissance des temps Бюро долготы Парижа* было согласовано с гринвичским среднем временем (GMT) в 1916 г., но не с всемирным временем, которое, как мы знаем, отличается от GMT на 12 часов.

И только в 1925 г. те эфемериды, которые использовались навигаторами и астрономами, стали отсчитывать начало суток в полночь, а не в

полдень. Новая шкала времени, выведенная из среднего времени путем добавления 12 часов, стала называться *гражданским временем*. Например, гражданское время Гринвича является ничем иным как всемирным временем, подробно представленным в 1884 году. Оно использовалось как независимый временной аргумент для *Connaissance des temps Бюро долготы* с 1925 по1950 гг. Однако общим подходом, даже в астрономии, является обращение к *гринвичскому среднему времени* для новой шкалы, и это является источником большой путаницы. Хотя название *Всемирное Время* происходит непосредственно из терминологии, использованной на конференции 1884 г., оно не использовалось вплоть до 1948 г., когда международный астрономический союз, после долгих колебаний, наконец, сделал твердые рекомендации в его пользу. Это не мешало продолжению неправильного использования акронима GMT там, где должно использоваться UT.

Даже сегодня все еще имеются остатки систем, в которых сутки считается с полдня, например, Юлианская Дата, упомянутая в Разделе 4.1.

4.2.3. К реализации Всемирного Времени

Начнем с соглашения, направленного на упрощение языка. Мы будем использовать термин *всемирное время* только тогда, когда мы говорим о шкале времени, основанной на вращении Земли и соотнесенной с на-

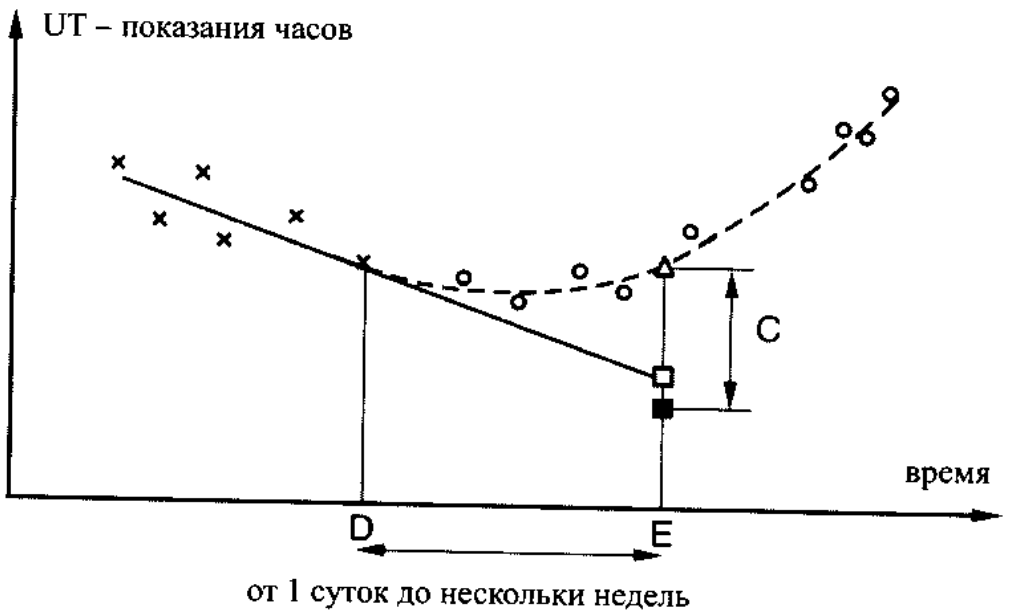


Рис. 4.2. Излучение сигналов всемирного времени UT. × – астрономические наблюдения UT минус показания часов, доступные на момент излучения сигналов времени D и действительные до E из-за влияния вычислений, задержек в линиях передачи и плохой погоды. Для прогноза значения в E (обозначенного через □) используется линейная экстраполяция. Последующие наблюдения (обозначенные через ○) не подтверждают этот прогноз. В дополнение, излученное время ■ фактически может не совпадать с прогнозом. Общая ошибка в излучении сигнал времени равна C.

чальным меридианом. Оно поэтому включает в себя всемирное время (UT) с его современным определением, также как и GMT, сдвинутое на 12 часов по отношению к UT, а также *гринвичское звездное время*. Оно является часовым углом точки весеннего равноденствия, измеренным от начального меридиана и математически связанным с GMT и UT.

С середины девятнадцатого столетия и почти до 1970 г. национальные организации по хранению времени работали более или менее неизменно. Они имели основное оборудование, включая часы, которые должны были быть стабильными настолько насколько это возможно, вместе с рефракторными телескопами, которые служили для определения перехода звезд через меридиан, или, иногда, через равные высоты. Все это дополнялось приборами, излучающими сигналы времени, которые по природе являлись электрическими, оптическими или звуковыми сигналами. Затем примерно с 1910 г. стали использоваться излучатели и приемники радиосигналов. В периоды ясной погоды астрономы тратили свои вечера или даже все ночи на отсчет моментов перехода звезд с помощью локальных часов, принятых за опорные часы. Затем они шли спать прежде чем сделать оформление «их восстановительных наблюдений», которое обычно означало вычисление *поправок часов*. Эти поправки должны были быть добавлены к показаниям опорных часов для того, чтобы на момент астрономических наблюдений, получить местное время, или после учета долготы – всемирное время. Поэтому назначение опорных часов заключалось в обеспечении некоторого способа усреднений для наблюдений и их сглаживания, т.е. попросту говоря, для хранения времени между вечерами, когда проводились наблюдения. Таким образом часы путем экстраполяции, давали приближение к всемирному времени в реальном времени. С помощью этой экстраполяции сигналы времени излучались в форме импульсов в согласованное номинальное время. Рис. 4.2. показывает, как этот метод работает и высвечивает некоторые неопределенности, связанные с ним. Причина, для чего нужен хороший *хранитель времени* становится ясна. Однако, прогресс в разработке хороших хранителей времени в конце концов лишил астрономов их роли держателей времени.

Когда примерно в 1910 г. радиосигналы стали пересекать Атлантический Океан, то было обнаружено, что они могут отличаться на одну или даже на две секунды. Это открытие пришло во времена, когда случайные погрешности в локальных измерениях всемирного времени были не хуже нескольких сотых секунды. Основными причинами, приводящими к систематическим ошибкам, объединенным со случайными ошибками, были ошибки в наблюдении положения звезд, задержки в инструментах, и более всего, ошибки в определении долготы – три причины, которые при объединении вместе, давали неопределенный сдвиг. Всемирное время

действительно единственное по своему определению, но его реализация может расходиться совершенно неприемлемым образом.

В бюро долготы Парижа стало ясно, что в этой ситуации надо что-то предпринять. Поэтому в 1911 г. по инициативе Ферри было предложено, чтобы французское правительство собрало международную конференцию по созданию международного бюро времени (BIH). Его основной задачей должна быть унификация мирового времени. Съезд состоялся в 1912 г., но война помешала установлению положения о BIH. Тем не менее международное бюро времени начало выполнять свои функции немедленно в полуофициальной манере при поддержке Парижской Обсерватории. Оно приобрело официальный статус в 1919 г., когда оно было поставлено в подчиненность международного астрономического союза, и выполняло свой мандат до 1988 года, когда его функции разделились на две части. Измерение атомного времени было передано международному бюро мер и весов (BIPM), в то время как астрономические и геодезические функции были реорганизованы, чтобы сформировать международную службу вращения Земли (IERS). Центральное бюро IERS до 2001 г. находилось в парижской обсерватории прежде чем переместиться в институт картографии и геодезии во Франкфурт, Германия. Мы будем обсуждать еще и другие области деятельности IERS. Ссылки [4.4, 4.5, 4.6] дают исторический обзор и подводят итог деятельности международного бюро времени.

Цель международного бюро времени состояла в обеспечении единственного приближения к теоретическому всемирному времени, известному как *точное время (definitive time)*, и затем давать расхождение между окончательным временем и номинальным временем излучения сигналов времени. Вычисления были исключительно сложными несмотря на то, что они основывались на достаточно простых принципах. Мы можем подчеркнуть это следующим примером. Каждый раз службы сопоставляли астрономические наблюдения и излучения сигналов времени со своими опорными часами. Затем путем обмена сигналами времени международное бюро времени могло сформировать уникальные мировые опорные часы, опираясь на данные наблюдений и излучений сигналов. Потом уже было просто сконструировать точное время путем усреднения астрономических измерений и их сопоставления с излучениями сигналов времени. Результаты публиковались в *Bulletin horaire* каждый месяц или каждый второй месяц в зависимости от периода.

Как пример, рассмотрим результаты на октябрь 1936 г. опубликованные в апреле 1937 г. Точное время 56 ежедневных излучений представляется с точностью до миллисекунд. Поэтому можно было бы ожидать разбросы в диапазоне до нескольких миллисекунд в точном времени из-за неопределенности в оценке задержек при распространении радиосигналов. С учетом этих разбросов, расхождения времен излучения по отноше-

нию к их номинальным величинам по оценке международного бюро времени были значительные, доходя до 0,2 с.

Мы, конечно, можем поинтересоваться – до какой степени точное время международного бюро времени представляет теоретическое всемирное время, определенное астрономами. Едва ли возможно точно дать ответ на этот вопрос, потому что уже упоминались систематические ошибки, и было бы мудрым ожидать неопределенности до нескольких сотых секунды. Это показывает, как важно различать синхронизацию, которую можно точно гарантировать на некоторых согласованных базисах, от реализации теоретической шкалы времени, опирающейся на эти базисы. Мы увидим, что эти отличия объясняют определенное нежелание к установлению атомной меры времени.

Работа, выполняемая в международном бюро времени, мало изменялась примерно до 1960 г., хотя технические улучшения используемых часов постепенно делали возможным уменьшение ошибок синхронизации в излучениях сигналов времени. Можно сказать, что тогда унификация времени могла быть достоверной в пределах 1 или 2 миллисекунд с временной задержкой около одного года, прежде чем эти результаты объявлялись. В реальном времени путем экстраполяции часов и использования локальных наблюдений расхождение в унификации времени было порядка 10 миллисекунд.

4.2.4. Определение секунды до 1960 г: секунда среднего солнечного времени

Измерение времени, основанное на вращении Земли, первоначально покоилось на постулате воспроизводимости, как это обсуждалось в главе 2. Этот постулат был впервые применен к средней продолжительности суток, а потом к периоду вращения. В 1737 году Эйлер показал, что если Земля, рассматриваемая как недеформированное твердое тело, вращается с постоянной скоростью, то среднее солнечное время приобретает вид абсолютного динамического времени. Однако, два столетия спустя, когда было осознано, что вращение Земли вовсе не равномерно, то стало много труднее устанавливать математическую модель и с ее помощью определять систему времени. Только небольшие эмпирические поправки были добавлены с 1955 г. для того, чтобы удалить ежегодные флуктуации (это шкала с поправками была обозначена как UT2). Было осознано, что среднее солнечное время, а следовательно и всемирное время, не являются хорошими мерами времени, и что вместо вращения Земли в качестве мер времени могут использоваться лучшие часы.

Секунда, единица времени, или длительность молчаливо и повсеместно была определена как 1/86400 часть средних солнечных суток. Удивительно, что это определение секунды среднего времени, которое остава-

лось в силе до поздних 60-х, никогда не было официально ратифицировано организациями, отвечающими за мировую метрологию начиная с метрической конвенции 1875 г.

4.3. Время, основанное на динамике Солнечной Системы: эфемеридное время

4.3.1. Первые сомнения в равномерности вращения Земли

В то время как Коперник принял древнегреческую догму о равномерности вращения Земли, Кеплер упомянул о возможности некоторых неравномерностей. Первая попытка демонстрации существования такого явления появилась благодаря Фламстиду. Вскоре после основания Гринвической обсерватории в 1677 г. он установил гигантские маятники в надежде выявить какие-нибудь аномалии, но ничего не обнаружил [4.7]. Позже в 1752 г. берлинская академия искусств и наук под председательством Мопертюи подняла следующий вопрос: «Является ли ежедневное движение Земли всегда происходящим с одной и той же скоростью или нет? Какие средства мы имеем в нашем распоряжении, чтобы ответить на этот вопрос? И если имеется некоторая неравномерность, что может быть ее причиной?» В своем ответе Кант сказал, что это мог бы быть эффект замедления из-за диссипации энергии в приливных движениях океанов. Он был прав, но его идея была подтверждена наблюдениями только два века спустя.

Убеждение, что вращение Земли должно быть идеально равномерным, укоренилось так глубоко, что к этим первым сомнениям едва ли кто-нибудь прислушивался. В 1825 г. Лаплас написал: «Определено, что со времен Гипарха длительность суток не изменилась больше, чем на одну сотую (десятичной) секунды [т.е. 0,00864 обычной секунды]». Как мы увидим заявление необоснованное. В то время как Феррель в 1864 г. а потом и Делоне 1865 г. утверждали, что расхождение между лунными эфемеридами и наблюдениями происходят из-за удлинения суток, Флеминг заявил в 1864 г., что нет движения более равномерного, чем вращение Земли вокруг своей оси.

4.3.2. Принятие неравномерности вращения Земли

Доказательство неравномерности вращения Земли впервые пришло из изучения орбитальных движений планет и Луны. С той степенью точности, которая требовалась для наблюдений, орбитальное движение планет может изучаться в предположении, что эти планеты и Солнце являются точечными объектами. Чисто гравитационное взаимодействие между ними является достаточно простым для того, чтобы вычислить их движения с

высокой точностью, используя, например, классическую механику, что было обычной практикой в это время. (Релятивистские теории использовались только несколько последних лет, однако, разница между ними и классической механикой несущественна в контексте данной книги). Вообще говоря, движение спутника вокруг центрального тела трактовать гораздо труднее. Причина кроется в том, что центральный объект может оказаться не совсем сферическим, а также в том, что могут существовать явления, приводящие к диссипации энергии. Такие явления незначительны в случае Луны, которую мы сейчас будем рассматривать и поэтому проигнорируем эти явления на данный момент. (Дальнейшие объяснения этого подхода будут даны в главе 8). Следующий шаг – это согласование наблюдений и теории. Другими словами, числовые значения должны быть приписаны неизвестным параметрам так, чтобы можно было установить эфемериды, дающие положение тел в форме таблицы в зависимости от времени, например, в терминах среднего солнечного времени. Что случится если время, используемое в наблюдениях, будет отличаться от равномерного абсолютного времени? Появятся расхождения между наблюденными положениями и положениями, даваемыми эфемеридами на момент наблюдения. Если такие расхождения достаточно велики, они не смогут быть объяснены случайными ошибками в наблюдениях. Ясно, что они могут быть минимизированы до некоторой степени путем согласованияи орбитальных параметров, которые по этой причине могут быть смещены. Если даже после этого расхождения до некоторой степени остаются и имеются ошибки в орбитальных параметрах, то в случае использования эфемерид для прогнозирования параметров ошибки должны быстро возрастать со временем.

Именно это и обнаружил Ньюкомб, когда он использовал лунные эфемериды, установленные Хансеном в 1857 г. После проверки правильности применения теории, Ньюкомб рассмотрел возможность неравномерности среднего солнечного времени. Другими словами, он поставил под сомнение вопрос о равномерности вращения Земли. Однако необходимо было сначала проверить эфемериды других планет на предмет выявления подобных же расхождений, причем таким способом, который был бы совместим с гипотезой о неравномерности вращения Земли. Это была трудная задача, так как эти эффекты для планет менее значительны, чем для Луны. И действительно, они пропорциональны истинной скорости орбитального движения. Последовавшие в 1927 г. исследования Брауна и де Ситтера [4.8] не оставили сомнений в том, что вращение Земли не может служить в качестве хороших часов.

Почти в то же время в 1929 г. Данжон в статье, замечательной своей ясностью и правильностью предсказаний, предложил использовать динамику орбитальных движений для измерения времени [4.9]. Давайте процитируем короткие выдержки.

... это законно рассматривать [вращение] Земли как глубинную причину кажущегося беспорядка, все еще царствующего в Солнечной Системе. Хотя закон Ньютона и сохраняется, он подвергается совершенно экстраординарным приключениям: с этого времени призванный измерять прохождение времени, он становится частично непроверяемым и прекращает быть тем, что должно быть строго действующим законом. [...] С тех пор мы бы требовали, чтобы эти (Кеплеровские) законы обеспечили меру для прохождения времени, так как мы не могли бы больше подвергать их экспериментальной проверке не попав в порочный круг. [...] Давайте просто надеяться, что однажды мы можем открыть хороший земной эталон времени, такой, что мы можем оставить эти типично логические трудности позади.

Хороший земной эталон появился в 1955 г. в виде атомного стандарта времени. Данжон мог также думать о таком стандарте.

К сожалению, статья Данжона была опубликована в любительском астрономическом обзоре и не получила внимания людей его круга. Так продолжалось до 1950 г. когда во время международного коллоквиума по астрономическим константам, который проходил в Париже [4.10], Клеменс сделал точное предложение по определению динамического времени. Оно было названо *эфемеридным временем* по предложению Брауэра. Определение эфемеридного времени, обсуждаемого в главе 8, было ратифицировано международным астрономическим союзом (IAU) в 1952 г.

Мы в частности увидим, что эфемеридное время имеет серьезный недостаток – не очень хорошее качество таблицы данных, с неопределенностями порядка 0,1 с. Оно может быть сравнимо с роскошными наручными часами, которые хотя и имеют отличный механизм, но не имеют нанесенной на циферблате градуировки, чтобы помочь нам измерить безупречность их поведения. Таким образом оно способно выявить только крупномасштабные и долговременные вариации во вращении Земли. В 1936 г. неравномерность вращения Земли была впервые открыта одновременно и независимо в двух лабораториях путем сравнения с искусственными часами двух типов, маятниковыми и кварцевыми. Позже было найдено, что ежегодная общая амплитуда составляет 60 мс [4.11, 4.12]. Однако можно увидеть из рис. 4.3, что годовая неравномерность находится на пределе того, что могло быть обнаружено теми часами, которые тогда использовались. Позже атомные часы доказали свою адекватность, продемонстрировав очень много других неравномерностей вращения (см. главу 8).

Все эти неравномерности приводят к изменению длительности секунды среднего времени. Для исключения некоторых кратковременных флуктуаций и годовой неравномерности, а также для уменьшения роли неопределенностей измерения, мы можем рассматривать среднее годовое

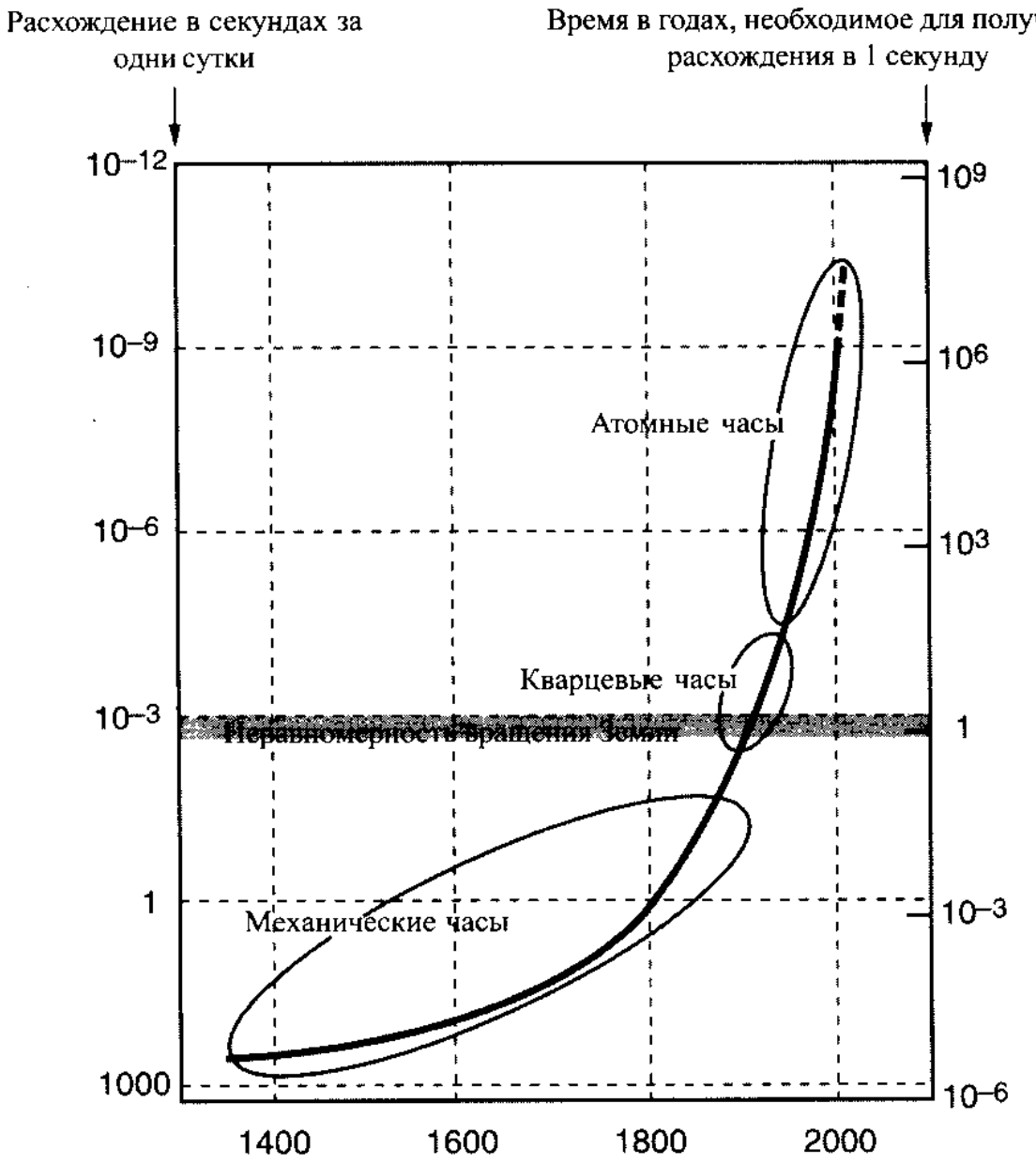


Рис. 4.3. Улучшение качества искусственных часов и сличения с часами, опирающимися на земное вращение

значение этой секунды. Тем не менее остаются непредсказуемые изменения, достигающие почти 10^{-7} с.

4.3.3. Определение секунды с 1960 по1967 гг.: эфемеридная секунда

Как объяснялось в разделе 2.4.2, если мы хотим обосновать определение секунды на динамических явлениях, то достаточно установить продолжительность некоторых частных явлений, представив их в виде таблиц. Коллоквиум 1950 г. рекомендовал поместить такие численные величины в Ньюкомбовские *Таблицы Солнца,* и это предложение не нашло возражений. Однако, выбор опорного явления привел к некоторым сомнениям. Явлением, выбранным международным астрономическим союзом (IAU) в 1952 г.

на генеральной ассамблее, был 1900 звездный год. Его продолжительность измерялась между двумя переходами Солнца через точки установленного равноденствия (т.е. без учета точности точек равноденствия), и поэтому эта продолжительность является виртуальной константой. Однако вскоре стало ясно, что непосредственно наблюдаемый тропический год был бы более предпочтительной опорой, несмотря на его медленные вариации. Это привело к исключительному событию – выработке резолюции международного астрономического союза по переписке.

В 1956 г. международный комитет по мерам и весам, используя право, данное ему 10-й генеральной ассамблеей по мерам и весам (CGPM) в 1954 г., решился на следующее определение:

Секунда есть 1/31 556 925, 9747 доля тропического года на 0 января 1900 г. в 12 часов эфемеридного времени.

Это определение было ратифицировано на 11 CGPM в 1960 г. Век его был скоротечен, и оно было отвергнуто в 1967 г. в пользу атомного определения секунды.

Определение *эфемеридной секунды* оказалось кроссвордом для тех, кто не был знаком с астрономией. Его странное определение происходит из того факта, что средняя долгота Солнца является, согласно Ньюкомбу, квадратичной функцией эфемеридного времени (ET). Тропический год на 0 января 1900 г. в 12 часов эфемеридного времени, является фиктивным годом, соответствующим скорости средней долготы на эту дату.

Рассмотрим более внимательно численные величины, используемые в определении эфемеридной секунды. Ньюкомбовские таблицы были основаны на астрономических наблюдениях 19 столетия. Естественно, что они были датированы с помощью среднего солнечного времени. Продолжительность эфемеридной секунды поэтому примерно равна средней продолжительности секунды среднего времени этого столетия. Однако в 1960 г., когда эфемеридная секунда была принята как единица времени для системы СИ, новая секунда была короче, чем используемая в то время секунда среднего времени (усредненная за 1960 г.) на $1{,}4 \times 10^{-8}$ с и примерная величина этого расхождения была известна во время принятия решения. Принятие такого решения противоречило обычной метрологической практике согласования последовательных определений единиц таким способом, чтобы они давали неизменяемую величину после изменения определения по крайней мере в пределах экспериментальных ошибок. В 1960 г. не предполагали, что это могло бы привести к печальным последствиям для системы всемирного координированного времени, которой мы сейчас пользуемся.

Эфемеридная секунда определяется единственным способом, но ее корни находятся в прошлом 1900 г. Ее физическая реализация на данный момент зависит от астрономических эфемерид. Оценивают, что с 1960 г. она могла бы быть реализована в пределах $\pm 2 \times 10^{-9}$ с путем анализа астрономических наблюдений, занимающих несколько лет. По сравнению с секундой среднего времени воспроизводимость повысилась бы в 50 раз.

4.3.4. Эфемеридное время: шкала сохраненная для астрономов

Красота и простота эфемеридного времени вызывает некоторый энтузиазм. В 1966 г., когда атомные часы уже работали в течение 11 лет, Вуллард и Клеменс писали [4.13]: «Традиционная практика измерения времени по истинным движениям небесных тел может быть теперь основана на точном динамическом фундаменте путем принятия в качестве первичного стандарта меры времени, неявно определенной через законы движения». История эфемеридного времени отягощена длинными сериями проблем и недопониманий, как мы увидим в главе 8.

Эфемеридное время никогда не было доступно и никогда не использовалось в повседневной жизни. Фактическое его применение было ограничено нуждами астрономической динамики. Никогда не было официальных организаций, устанавливающих централизацию и процесс его измерения так, чтобы произвести единственную реализацию, которая могла бы быть принята по соглашению. Давайте подчеркнем, что до 1960 г. официальная шкала времени, всемирное время, и определение секунды (среднего времени) последовательно основаны на вращении Земли. Следовательно, распространение шкалы времени эффективно обеспечивало секунду СИ. Вместе с принятием эфемеридного времени ET начался период несовместимости между шкалой времени UT и единицей времени. Атомное измерение времени положило конец этой путанице в 1971 г.

4.4. Измерение атомного времени

4.4.1. Атомные стандарты частоты и определение атомной секунды (1967 г.)

Наука постепенно познавала, что различные химические образцы состоят из молекул, которые в свою очередь состоят из атомов. Со временем было установлено соотношение между структурой излучаемого спектра и атомным или молекулярным составом возбужденных газов. Предполагаемая универсальность атомных свойств могла затем быть перенесена на эти спектры. Максвелл в 1873 г. и Кельвин в 1879 г. предположили, что длины волн спек-

тральных линий и периоды соответствующих излучений могли быть использованы для того, чтобы определить единицу длины и единицу времени соответственно [4.14]. Эти предложения были воистину пророческими, если мы припомним, что определение метра было основано на длине волны спектральной линии только с 1960 г. (и до 1983 г.). В настоящее время в высокоспециализированных лабораториях оптические излучения используются для создания сетки опорных частот. Благодаря недавно достигнутому прогрессу в технике сличения СВЧ и оптических частот весьма вероятно, что оптические стандарты времени и частоты вскоре будут конкурировать с нынешними атомными стандартами, работающими в сантиметровом диапазоне.

Как и во многих других областях, знания, которые мы имеем сейчас в электромагнетизме, квантовой физике, атомной физике и спектроскопии, и которые привели к изобретению и развитию стандартов времени и частоты, были приобретены в конце девятнадцатого и первой половине двадцатого веков. Мы можем подвести итоги этих лет прогресса ссылками на тех физиков, которые внесли ключевые вклады. Планк создал основы квантовой теории, в то время как Эйнштейн, кроме многих других значительных вкладов, включая теорию относительности, ввел понятия фотонов и стимулированного излучения. Бор применил квантовую теорию к объяснению атомной структуры и ввел понятие энергетических уровней. Герц первым создал радиогенератор и обнаружил радиоволны. Де Бройль, Гейзенберг и Шредингер создали и развили волновую механику. Штерн сделал вклад в кинетическую теорию газов и открыл вместе с Герлахом магнетизм атомов и его пространственное квантование.

Технологический прорыв, который начался в 1930-х годах и который подстегивался далее нуждами радиосвязи и радарным обнаружением во время второй мировой войны, также сыграли ключевую роль. В 1945 г. стала возможной генерация радиоволн до частот 30 ГГц и измерение их частоты. Эти разработки стали отправной точкой для главного прорыва в направлении прецизионной, внутригерцовой спектроскопии, выполненной Таунсом и Паундом.

Первые атомные часы были созданы в 1948 г. Лионом в США в Национальном Бюро Стандартов, известном теперь как Национальный Институт Стандартов и Технологии [4.15]. Опорными линиями служили интенсивные линии поглощения в молекулах аммиака, лежащие в диапазоне вблизи 24 ГГц. Молекулярный резонанс управлял частотой кварцевого генератора, который затем использовался для генерации импульсов времени. Однако резонанс был уширен доплеровским эффектом и столкновениями. Его долговременная стабильность была не лучше, чем у кварцевых часов и исследования в этом направлении были прекращены.

Цезиевые часы являются результатом работы, выполненной Раби и его студентами [4.16, 4.17]. В конце 1930-х Раби работал над техникой

магнитного резонанса в атомных и молекулярных пучках. Этот метод, первоначально разработанный для измерения магнитных моментов, вскоре стал использоваться для изучения радиочастотных спектров атомов и молекул. Первое экспериментальное доказательство существования переходов в сверхтонкой структуре атомов цезия было получено в 1940 году. Возможность использования метода магнитного резонанса в атомных пучках для создания атомных часов уже обсуждалась в группе Раби в 1939 году. Кажется хорошо установленным фактом, что Раби представил эту идею широкой публике на конференции Американского Физического Общества в 1945 году. Фактически он предложил использовать атомы цезия. В 1950 г. Рэмси улучшил условия для взаимодействия между электромагнитными волнами и атомами путем введений двух разнесенных осциллирующих полей. Используя экспериментальные методы, разработанные Раби и Рэмси, в 1955 году в Национальной Физической Лаборатории в Великобритании Эссен и Пэри построили первые надежные цезиевые часы [4.18]. Эти часы регулярно использовались для калибровки частоты кварцевых генераторов. В то же самое время, Захариас, один из студентов Раби, разработал в Массачузетском Технологическом Институте прототип промышленных атомных часов. В этих часах атомный резонанс использовался для подстройки частоты кварцевого генератора с помощью петли обратной связи. Первые атомные промышленные часы на основе этого прототипа были произведены Национальной Компанией в 1956 году под названием Атомихрон [4.14].

Экспериментальное подтверждение принципа усиления путем стимулированной эмиссии излучения впервые было получено в 1954-55 годах Таунсом в США и Басовым и Прохоровым в СССР. Они впервые построили генерирующий мазер (Microwave Amplification by Stimulated Emission of Radiation), используя переход в молекуле аммиака на частоте 24 ГГц. Эти мазеры рассматривались на предмет использования их в атомных часах, но их договременная нестабильность порядка 10^{-10} была слишком велика и идея не получила дальнейшего распространения [4.19]. Тем не менее эти исследования были очень полезны для указания путей создания твердотельного мазера Бломбергеном в 1956 году и затем лазера, разработанного Шавловым и Таунсом в 1958 году [4.17].

Водородный мазер, разработанный Рэмси, является результатом эволюции метода магнитного резонанса Раби. Первоначальная цель состояла в сужении резонансной линии атома цезия путем увеличения времени нахождения между двумя осциллирующими полями [4.17]. Для этого атомы цезия испытывали столкновения и задерживались на некоторое время в ячейке с перегородками и стенками, покрытыми парафином. Позднее было обнаружено, что лучше использовать тефлоновое (фторопластовое) покрытие. Идея состояла в том, чтобы атомы водорода

подвергались огромному количеству столкновений со стенками и не испытывали значительных возмущений. Было также осуществлено детектирование атомов путем стимулированного излучения и, таким образом, в 1959-60 годах были изобретены водородные мазеры. Этот успех основан на том, что достигается очень большое время жизни атомов в ячейке, покрытой фторопластом (около 1 с), и поэтому получается очень большой период взаимодействия с резонансной полостью, что способствует стабильности частоты. Промышленный вариант водородного мазера вскоре был разработан Вессо и изготовлен на Объединении Вариан.

Так как для многих применений требуются компактные атомные стандарты времени и частоты, то с момента открытия оптической накачки проведено исследование возможности ее применения в часах. В 1950 г. Кастлер предложил оптический метод для того, чтобы заменить использовавшиеся в те времена для селекции и детектирования атомных состояний мощные неоднородные магнитные поля производимые в те дни с помощью громоздких магнитов. В 1953 году Дике показал, что доплеровское уширение микроволновых резонансных линий может быть практически устранено путем смешения атомов с нейтральными газами, ограничивающими их скорость рассеяния. В 1956 году Демелт обнаружил, что микроволновый резонанс проявляет себя через уменьшение в интенсивности света накачки, проходящего через ячейку, содержащую атомы и буферный газ [4.15, 4.16]. Это сделало возможным создание компактных рубидиевых часов, и они вскоре стали производиться многими изготовителями. Они и теперь являются наиболее распространенными атомными часами. С 1980-х годов методы оптической накачки стали применяться к цезиевым пучковым часам благодаря разработке полупроводниковых лазеров.

Двух- или трехмерное ограничение ионов в неоднородном переменном электрическом поле разрабатывалось с 1955 г. Паулем. Долговременное удержание, полученное таким способом, является благоприятным для наблюдения резонансов в микроволновом диапазоне. Именно Демелт в 1965 году использовал такие ловушки для прецизионных измерений частот переходов в сверхтонкой структуре иона $^3He^+$. В 1969 году Мэйжор предложил объединить методы оптической накачки и методы удержания ионов для стандарта времени и частоты на ионах ртути $^{199}Hg^+$. Такой стандарт впервые был реализован в 1979 году в лаборатории атомных часов в Орсэ (Франция). В 1983 году несколько улучшенных работающих образцов было произведено на фирме Хьюлетт-Паккард под руководством Катлера. Изучение стандартов частоты на удержанных ионах продолжается с использованием переходов и в микроволновом и в оптическом диапазонах. В некоторых из этих стандартов используется метод лазерного охлаждения, предложенный в 1975 году Вайнлэндом и Демелтом.

Через несколько лет Хэнш и Шавлов описали эффективный способ замедления и уменьшения тепловых движений нейтральных атомов. Метод впервые был использован для замедления атомного пучка. Затем в 1989 году Вьеман и его группа в университете Колорадо преуспели в применении этого метода к атомам цезия, содержащимся в ячейке при малых давлениях. Мячик из атомов с кинетической энергией теплового возбуждения, соответствующей температуре порядка 1 милликельвин был пойман на перекрестье шести лазерных лучей. Несколько групп в США и Франции сделали экспериментальный и теоретический вклады в улучшение техники для охлаждения, удержания и управления движением атомов. Сейчас уже возможен запуск холодных атомов цезия вверх со скоростью порядка 1м $с^{-1}$. Таким образом стало возможно их использование в цезиевых часах, в которых преимущество медленных атомов является решающим. Впервые первичный стандарт частоты на холодных атомах цезия был создан в Париже (Франция) в 1996 году Клероном из лаборатории первичного времени и частоты, Соломоном из Высшей Технической Школы и соответственно их исследовательскими группами.

Нобелевский комитет наградил некоторых физиков, которые непосредственно внесли вклад в изобретение и исследование атомных часов: И.Раби (1944 г.), С.Таунс, Н.Басов, и А.Прохоров (1964 г.), А.Кастлер (1966 г.), Н.Рэмси, В.Пауль и Х.Демелт (1989), С.Чу, С.Коэн-Тануджи и В. Филлипс (1997 г.), даты соответствуют времени присуждения премии.

В главе 6 объясняется, как работают различные типы атомных часов. Мы увидим, что именно цезиевые часы с лазерным охлаждением атомов дают в настоящее время наивысшую точку, но их превосходству угрожают стандарты на холодных атомах рубидия. Точность является ключевой идеей метрологии. Здесь она понимается, как способность воспроизводить частоту, имеющую идеальное соотношение с частотой перехода невозмущенного атома. Она исключает всякие дрейфы частоты. На время написания книги цезиевый стандарт играл привилегированную роль в фундаментальных измерениях частоты. Давайте рассмотрим, как происходила эволюция его точности, и обрисуем последствия этой эволюции для определения секунды.

Первые цезиевые стандарты имели неточности частоты с относительной величиной порядка 10^{-9}. (Неопределенности и малые изменения (вариации) обычно выражаются в относительных величинах. Мы больше не будем заострять внимание на этом в дальнейшем, так как отсутствие названия физической величины будет достаточным индикатором употребления относительной величины). Прогресс достигался стремительно и, как мы увидим, стандарты необходимо было сличать друг с другом, чтобы оценить их взаимное согласие. Очевидно, частота перехода ν_{Cs} должна была быть выражена в терминах единиц времени там, где это возможно. В своих ранних исследовательских работах Эссен и Пэри использовали

секунду среднего времени, которая производилась в реальном времени. Эфемеридная секунда уже была определена, хотя еще и не признана как единица СИ. Однако, как мы уже отмечали, она могла быть реализована только как результат длительного анализа. Этот анализ был выполнен Марковицем и др. [4.20], и в 1958 году было получено значение частоты

$$\nu_{Cs} = 9\ 192\ 631\ 770 \pm 20\ \text{Гц}. \qquad (4.2)$$

Единица Гц, упомянутая здесь, основана на эфемеридной секунде. Неопределенность ±20 Гц была почти полностью обусловлена неопределенностью реализации эфемеридной секунды.

Естественно, что различные лаборатории, изучающие атомные стандарты частоты, охотно согласились на единственное значение частоты для цезия. Значение частоты, полученное Марковицем и др. было немедленно принято по соглашению в том смысле, что она рассматривалась строго как точная опорная частота для выражения других частот. Это неявным образом определяло неофициальную атомную секунду. И действительно, она была признана как таковая на 12–й генеральной конференции по мерам и весам в 1964 году, хотя и чувствовалось, что еще не пришел момент для принятия нового определения секунды в предвидении дальнейшего значительного прогресса, который мог бы вскоре произойти. Однако в осознании неотложности ситуации переход к сверхтонкой структуре цезия 133 был принят для использования в качестве стандарта и вышеуказанное значение было приписано его частоте.

В 1967 году неопределенность частоты атомных стандартов была уменьшена до 10^{-12} и было решено выйти с предложением об атомном определении единицы времени. Это определение было принято на 13-й генеральной конференции по мерам и весам в 1967 году:

Секунда есть промежуток времени, состоящий из 9 192 631 770 периодов излучения, соответствующего переходу между двумя сверхтонкими уровнями основного состояния атома цезия 133.

Определение так называемой эфемеридной секунды было тогда же аннулировано. Как было со временем показано, атомная секунда была согласована с эфемеридной секундой настолько, насколько это было возможно. Последующие измерения эфемеридного времени подтвердили эту согласование. Текущая секунда таким образом имеет длительность, которая соответствует длительности секунды среднего времени, согласно усреднениям данных девятнадцатого столетия.

Улучшение цезиевых стандартов продолжалось. В 1976 году неточности лучших стандартов, находившихся тогда в Германии, были уменьшены до нескольких долей от 10^{-14}. Ситуация сохранялась на этом уровне

вплоть до 1995 года, когда новому этапу в подъеме к улучшенной точности был дан старт тем открытием, что атомы могут быть охлаждены лазером. Эта техника, как мы увидим в главе 6, уже интегрировалась в очень точные цезиевые часы.

Применение измерений времени находится на подъеме потому, что промышленность способна производить высококачественное оборудование для измерений времени. В частности, цезиевые стандарты имеются в продаже и, хотя они менее точны, чем лучшие лабораторные инструменты, зато они менее громоздки, имеют высокую стабильность частоты и могут работать непрерывно с высочайшей надежностью. Строго говоря, это цезиевые часы, нанизывающие вместе секунды. Существование таких часов делает реально выполнимой задачу: – основать мировую шкалу времени на атомных переходах, а не на движениях небесных тел. Это подводит нас к рождению Международного Атомного Времени.

4.4.2. От Всемирного Времени к Международному Атомному Времени

(а) Возраст сличений частоты и интегрированное атомное время

Первый цезиевый стандарт частоты работал спорадически. Он использовался время от времени, чтобы прокалибровать частоту независимых кварцевых часов, которые между такими измерениями сохраняли память об атомной частоте. Затем стало возможным установить шкалу времени, основанную на частоте атомного перехода, т.е. атомное время. Давайте обсудим некоторые детали этой работы. Мы можем тогда вывести свойства атомного времени, реализованного этими методами и ввести полезные определения, как представлено ниже.

Генератор характеризуется *номинальной частотой* ν_0, установленной или объявленной его производителем и соотносящейся с определением секунды. Его фактическая частота несколько отличается и изменяется. Она обозначается через $\nu(\theta)$, где θ дата на шкале времени θ, взятой за опорную. Мы увидим, что от θ требуется несколько метрологических характеристик, которые могут, например, быть взяты из всемирного времени. Так как номинальные частоты различных генераторов не являются одинаковыми, то полезно определить:

- *нормализованную частоту* $\Phi(\theta)$ через

$$\Phi(\theta) = \frac{\nu(\theta)}{\nu_0}, \tag{4.3}$$

которая близка к единице;

- *относительную отстройку частоты* $y(\theta)$ через

$$y(\theta) = \Phi(\theta) - 1, \tag{4.4}$$

которая мала по сравнению с единицей.

Для того чтобы сконструировать атомную шкалу времени должна быть установлена величина номинальной частоты перехода ν_{Cs}. Она может оказаться той величиной, которая будет использована для того чтобы определить секунду, но в начале это было неизвестно и могли использоваться различные величины. Если частота внешних кварцевых часов есть ν_Q, то эпизодические сличения частот в даты θ_i могут быть выражены в терминах разности между нормализованными частотами,

$$y_{Cs}(\theta_i) - y_Q(\theta_i) \equiv [y_{Cs} - y_Q](\theta_i). \tag{4.5}$$

Эти данные являются фактически средними величинами на дату θ_i. Начиная с этих измеренных величин, функция $[y_{Cs} - y_Q](\theta)$ должна быть определена настолько точно, насколько это возможно. Атомное время τ_A затем получается путем коррекции показаний τ_Q кварцевых часов:

$$[\tau_A - \tau_Q](\theta) = [\tau_A - \tau_Q](\theta_0) + \int_{\theta_0}^{\theta} [y_{Cs} - y_Q](\theta)\mathrm{d}\theta. \tag{4.6}$$

Отсюда понятно почему на шкалы времени, полученные таким способом, часто ссылаются как на *интегрированные шкалы времени*. Начало τ_A является произвольным. Хотя вначале оно было выбрано в разных лабораториях по разному, в конце концов в 1960 году оно было установлено по соглашению так, что τ_A должно быть равно всемирному времени UT2 на 1 января 1958 года в 0 часов по UT2. Что же касается значения ν_{Cs}, то все согласились использовать величину, полученную Марковицем и др. Ясно, что шкала атомного времени зависит от того, какие предположения были сделаны при интерполяции частоты, а также какие числовые методы были приняты, чтобы выполнить интегрирование. Другими словами, атомное время определено не однозначно.

Когда в 1955 –1960-х годах первые атомные шкалы были построены таким путем, то в это время уже использовались очень низкие частоты (VLF) [сверхдлинные волны (СДВ), 10-30 кГц], для целей дальней радиосвязи и навигации. Несущие частоты были стабилизированы по кварцевым генераторам. Эти частоты очень точно принимались методом слежения за фазой. Таким образом эти частоты составляли хорошие реперы (опорные частоты) для сравнения различных частот, используемых в лабораториях. Сличения частоты стали таким образом доступны на межконтинентальных расстояниях. Сличения времени однако оставались неопределенными, так как даже если в начальный момент сличения были доступны, практически было невозможно из-

бежать потери некоторых периодов несущей частоты, так что в течение года могло проскальзывать от нескольких микросекунд до десятков микросекунд.

Поэтому пионеры атомного времени имели возможность сличать только частоты стандартов со множеством неопределенностей. Сюда включались недостатки локального способа интерполяции частоты, плохое знание распространения волн и потери фазы. Тем не менее с помощью вычислений можно было сконструировать стандарт средней атомной частоты, который, возможно, был лучше, чем каждый из индивидуальных стандартов, и производил среднее интегрированное атомное время. Это время локально материализовалось через коррекцию показаний одних из часов, участвующих в сличениях. Этот метод использовался, в частности, в международном бюро времени вплоть до 1969 года [4.21].

Такое интегрированное атомное время первоначально было использовано для изучения неравномерности всемирного времени, т.е. вращения Земли. Оно также использовалось для стабилизации сигналов времени передающих радиостанций, позволяя, насколько возможно лучше экстраполировать всемирное время по сравнению с предыдущими часами. Напротив, его применение в других областях, таких как небесная динамика, в частности, не вызвало какого-нибудь энтузиазма. Общее мнение астрономов было хорошо выражено в работе, опубликованной в 1966 году [4.13]:

Атомные часы обеспечивают только стандарт частоты; он определяет единицу времени, но не непрерывный счет единиц, необходимый чтобы определить интервал, прошедший от любой начальной эпохи в прошлом. Астрономические определения времени существенны для определения эпох и опорных моментов времени к ним, так как не существует искусственных часов, которые могли бы бесконечно долго продолжать свою работу наподобие небесных движений.

(b) Эпоха современного атомного времени

Будет правильным, если мы скажем, что подъем атомного времени произошел главным образом благодаря техническим совершенствам, произведенным в конструкциях атомных часов и разрастанию их общего числа. Но этого не произошло бы вовсе без точных методов сличений их показаний, без *сличений времени*, и без возможности передачи времени таким способом, чтобы можно было осуществить *синхронизацию*. Обычные радиосигналы времени, такие, как они появились в начале двадцатого столетия, и которые используются до сих пор, страдают от некоторой неопределенности при учете времени, необходимого для их распростра-

нения. Эта неопределенность, оцениваемая величиной порядка 1 мс, совершенно неприемлема, если мы надеемся использовать свойства атомных часов на больших расстояниях. Например, во времена, когда было принято атомное определение (1967 г.), потребовалось бы тридцать лет усреднения, чтобы получить выигрыш от той высокой точности лучших цезиевых часов с помощью таких сигналов.

Для того чтобы установить порядки величин, относящихся к метрологическим проблемам, вычлененным в этой работе, мы резонно можем предположить, что неопределенность частоты, обусловленная сличениями времени на расстоянии, за один день сличений не должна превышать неточность стандартов частоты. Для стандартов, имеющих точности 10^{-12}, неопределенность сличений должна быть меньше 0,1 мкс, в то время как для стандартов с 10^{-14} неопределенность должна быть меньше 1 нс. Это действительно очень строгие требования. Почти всегда было невозможно согласовать точность сличения времени со стабильностью и точностью атомных часов.

Тем не менее, в 1962 году экспериментальная передача сигналов телекоммуникационного спутника Телстар обеспечила линию передачи времени между Великобританией и США с точностью до 1 мкс (для космического сегмента) [4.22]. Хотя после этого эксперимента последовало еще много других экспериментов, применение этого метода все еще оставалось на уровне исследований. Как упоминается в цитатах в конце последнего раздела, одна из особенностей величины время заключается в том, что его измерение требует непрерывных усилий. Поэтому методы, необходимые для сличений, должны быть простыми и достаточно дешевыми, чтобы их можно было применять изо дня в день.

Практическое решение было найдено примерно в 1968 году путем использования системы Лоран-С. Сигналы передаются на расстояния до 1000 – 2000 км наземными волнами и поэтому имеют хотя и неизвестные априори, но очень стабильные задержки распространения (с возможными вариациями ±1 мкс). Также успешно и даже с бóльшей точностью, использовался локальный одновременный прием обычных телевизионных сигналов. Поэтому появилась идея определения задержек распространения и их эпизодической проверки. Это было сделано с помощью перевозимых по дорогам и с помощью гражданской авиации работающих цезиевых часов. Этот метод будет описан в главе 5. А сейчас давайте только скажем, что большие регионы глобуса, включая Северную Америку и Европу, могут быть покрыты сетью сличений времени, рутинно работающей с точностью до ±1 мкс.

С этого момента и далее средние атомные шкалы могут быть сконструированы из показаний часов. Эти атомные шкалы были реализованы путем вычисления поправок к показаниям участвующих часов. Это при-

вело к новой проблеме. Должны быть установлены оптимальные алгоритмы для того, чтобы построить шкалы времени из данных с переменным качеством (см. главу 7). 1 января 1969 года международное бюро времени заменило свой стандарт средней частоты на среднее атомное время, без какого-нибудь перерыва в его атомной шкале. Для того чтобы следовать в русле технического прогресса в стандартах и передачах времени, в1973 году международное бюро установило алгоритмы и мировую координацию, которая с незначительными модификациями живет и до сих пор (см. главу 7). Значительное улучшение в сличениях разнесенных часов было обеспечено в 1983 году спутниковой системой, известной как глобальная система навигации и определения положения (GPS). GPS постепенно вытеснила Логан-С из лабораторий времени и уменьшила неопределенности до нескольких наносекунд. Недавно стал оперативно использоваться для сличений времени обмен сигналами, переизлучаемыми телекоммуникационными спутниками, что привело к улучшенным результатам, особенно для частотных сличений.

Национальные организации, в частности, Военно-морская Обсерватория США в Вашингтоне (USNO), также производят атомную шкалу времени. Необходимость согласия о единственном репере привело ряд ключевых международных организаций к рекомендации об использовании шкалы международного бюро времени, включая: международный астрономический союз (IAU) в 1967 г., международный научный радиосоюз (URSI) в 1969 году, и международный консультативный комитет по радио (CCIR) в 1970 году. Предельное признание пришло от организаций метрической конвенции с определением *Международной Атомной Шкалы* (TAI), сформулированной в 1970 году международным комитетом по мерам и весам (CIPM) в следующей редакции:

Международное Атомное Время (TAI) есть временная опорная координата, установленная международным бюро времени на основе показаний атомных часов, работающих в различных учреждениях в соответствии с определением секунды, являющейся единицей времени в международной системе единиц.

Это определение, или даже, это официальное признание было ратифицировано на 14-й генеральной конференции по мерам и весам в 1971 году. После переходного периода, который длился с 1955 по1971 годы, измерения времени стали, наконец, опять корректными, с единицей времени и мировой шкалой времени, основанными на одном и том же физическом явлении, на оговоренном атомном переходе. Время, как и другие фундаментальные физические величины, стало теперь измеримым с помощью лабораторных стандартов. Ничто не препятствовало теперь тому,

чтобы оно заняло соответствующее место среди других величин в международном бюро мер и весов (BIPM) в Сэвре (Франция). И это случилось в 1988 году.

Так как атомная шкала международного бюро времени поддерживалась непрерывно с 1955 года, и хотя тех пор вычислительные методы с сильно эволюционировали, она была задним числом названа TAI.

В то время как стабильность атомных стандартов времени и TAI улучшалась, стало ясно, что вскоре потребуется более полное определение TAI в контексте общей теории относительности. Мы должны упомянуть исключительно удачные идеи, выдвинутые на эту тему Беккером еще в 1967 году [4.23]. Но они не были востребованы до 1980 года, когда консультативный комитет по секунде (CCDS) (международный комитет по определению секунды CCDS при международном комитете по мерам и весам CIPM) дал релятивистское определение TAI, определение, которое было довершено международным астрономическим союзом (IAU) в 1991 году в его резолюции A4, и которое было незначительно модифицировано в соответствии с переопределением земного времени (TT) в 2000 году (см. раздел 3.3.2e). (CCDS с 1997 года преобразовался в международный консультативный комитет по времени и частоте CCTF). Важное последствие прогресса в измерениях времени для истории науки состоит в том, что общая теория относительности стала существенным инструментом в метрологии и ее практических приложениях. Действительно, тематика «метрология и относительность» является предметом докладов, подготовленных при покровительстве CCDS (консультативный комитет по определению секунды) [4.24].

4.5. Всемирное координированное время

В тот период, когда устанавливалась международная атомная шкала, постепенно созревала идея координации передач радиосигналов времени (которые постоянно излучали сигналы всемирного времени).

Инициатива была предпринята Великобританией и США. В обеих этих странах с августа 1959 года сигналы передавались так, чтобы пометить локально полученные атомные шкалы времени. В это время значение частоты перехода цезия, которое позже будет использовано для определения секунды, уже было общеупотребительно. Отметим, что это значение связано с эфемеридным временем. Следовательно, было замечено, что шкалы атомного времени идут быстрее, чем всемирное время. Для того чтобы построить атомное время в примерном согласии с всемирным временем, должна быть введена относительная отстройка частоты y_U. Одно и то же значение y_U было выбрано в обеих странах и сигналы времени изначально были синхронизированы. Так как всемирное время имеет не-

предсказуемые вариации, необходимо было подстроить время чтобы избежать слишком частых изменений y_U. Эти подстройки выполнялись с шагом в 50 мс в согласованные времена в соответствии с наблюдениями всемирного времени.

Так как многие сигналы времени имеют секундные маркеры, непосредственно связанные с несущей (например, в излучении на 10 МГц, секундный маркер появляется после каждых 10^7 периодов), то атомная частота могла быть найдена из y_U или путем использования расстояния между маркерами, или путем измерения несущей частоты.

К этим скоординированным усилиям вскоре присоединились другие страны и уже в 1960 году международные радиосоюз (URSI) предложил международному бюро времени, которое отвечало за унификацию всемирного времени, устанавливать отстройку y_U на каждый год. Соответственно, международное бюро времени определяло даты, когда будут вводиться скачки времени.

Эта система довольно быстро прогрессировала. В 1963 году скачки времени в 50 мс были заменены на 100 мс. До 1965 года шкала времени была более или менее общей для всех координированных сигналов, и она была спонтанно поименована как *всемирное координированное время* (UTC), хотя еще и не было дано строгое определение. Например, международное бюро времени рассматривало UTC как усреднение излученных сигналов координированного времени. Важным шагом в развитии UTC было решение международного бюро времени ввести с 1 января 1965 года определение UTC через математическую связь с атомным временем международного бюро, которое впоследствии превратилось в TAI. Эта связь представлена в виде

$$UTC - TAI = y_U(TAI - TAI_0) + B, \tag{4.7}$$

где TAI_0 есть произвольно установленная дата начала и B – константа, изменяемая шагами так, чтобы поддерживать условия

$$|UTC - UT2| < \varepsilon, \tag{4.8}$$

где UT2 является всемирным временем, включающим поправки на его квази-периодические годовые неравномерности и ε есть оговоренное отклонение.[2]

Это определение UTC использовалось до 1972 года. Однако необходимость в отстройке частоты y_U, иногда изменяющейся в начале года, становилась все более и более обременительной. (Наибольшие изменения были

[2] Отметим, что такая шкала времени как UT1 обозначенная своим акронимом, в написании заглавными буквами. Ее чтение является примером спецификации членов шкалы времени написанной курсивом, например, *UT1(θ)*. Аргумент *θ*, значение θ не всегда показаны точно.

сделаны 1 января 1966 года, когда перешли от $-1{,}5 \times 10^{-8}$ к $-3{,}0 \times 10^{-8}$.) Одним из недостатков было то, что некоторые передатчики опирались на несущие частоты, связанные с частотами UTC, так что на этих передатчиках должны были проводиться технические изменения. По этой причине международный консультативный радиокомитет (IRCC) в 1966 году принял решение, что некоторые сигналы времени передавались бы с $y_U=0$ за счет повторяющихся скачков времени. Размер этих скачков был установлен на 200мс и они вводились только в начале месяца, если было необходимо удерживать ε меньше чем 200мс. Это *ступенчатое атомное время* было принято только в Германии, а также на одном передатчике в США. Важно же это, главным образом потому, что был подготовлен путь, одобренный в 1972 году и используемый в настоящее время в UTC.

Мы можем поинтересоваться, почему сигналы времени, а, следовательно, мировое время не были подстроены под международное атомное время, когда эта шкала времени утверждалась официально в 1971 году. Конечно, TAI идет быстрее по сравнению с всемирным временем UT. Через какое-то длительное время 12 часов дня наступят заметно раньше, чем солнце пересечет гринвичский меридиан. Однако такое отклонение порядка одного часа набегает через каждую тысячу лет. Когда мы осознаем, что мы одобрили изменение показаний часов во многих странах на один час дважды в год, и, следовательно, устанавливаем летом время на час вперед по сравнению с солнечным временем, идея подстройки часов на один час назад каждые десять столетий покажется незначительной подстройкой.

Возражения против всеобщего использования TAI являются немного сентиментальными, так как возвращают нас к исходному положению – заветной мечте, что время определяется движениями звезд и немного техникой. Мореходы, использующие астрономические средства навигации были очень сильно настроены против идеи не передавать больше всемирное время, которое им нужно для определения их положения. Поправка, необходимая для того чтобы получить всемирное время UT из TAI может быть предсказана до ближайшей секунды на несколько лет вперед, и этого достаточно для астрономического определения положения в море. Однако морские организации считают, что риск ошибки (например, поправка была сделана в неправильном направлении) слишком велик. Также рискованно передавать два типа сигналов времени, одни в UT и другие в TAI. В 1970 г. после долгих дебатов и больших споров [4.25], Международный консультативный комитет по радиосвязи CCIR был готов рекомендовать, что UTC должно быть определено с $y_U=0$ и B равное целому числу секунд, и изменяемое скачущими секундами. Сначала расхождение ε (относящееся теперь к UT1, а не к UT2), было установлено на уровне 0,7 с, прежде чем установили 0,9 с в 1974 году. Более

того, сигналы времени должны нести кодированную информацию, которая может быть использована, чтобы получить UT1 в пределах 0,1 с.

Новая UTC была принята согласно правилам, детали которых определены CCIR (см. главу 7). Она официально одобрена на 14-й генеральной конференции по мерам и весам в 1971 году и была внедрена в практику с 1 января 1972 года, путем введения скачков времени, равных скачкам, округленным до секунды. С тех пор она и используется. Система с нарастающим напором подвергается критике, потому что скачущие секунды вводятся с интервалом от одного до двух лет в зависимости от причуд вращения нашей планеты, что является действительным недостатком. Набирает силу идея принятия непрерывной шкалы времени, например путем замораживания на какую-то дату разности между TAI–UTC (2002 год). Но все еще невозможно сказать, каким будет UTC. Человечество еще возможно не готово перейти к чисто атомному времени и поэтому вращение Земли продолжает управлять делами в мире, по крайней мере в случаях, касающихся общего использования времени.

4.6. Финальные заметки

С 1972 года измерения времени официально покоятся на сверхтонком переходе в атоме цезия, но даже если это так, вращение Земли все равно следует приближенным путем. Когда в 1982 году открыли миллисекундные пульсары, то некоторые полагали, что астрономическая мера времени вернется обратно. Мы увидим в главе 8, что такие тела не могут обеспечить определение секунды, каким бы регулярным не было их вращение.

Следует отметить то, как одно определение секунды следует за другим. Под давлением научных и технических требований применение нового определения секунды может даже предшествовать официальному принятию.

Положение для шкал времени совершенно другое. Новая шкала создается в связи с текущим определением секунды, но ранее сформированная шкала не исчезает. Это в какой-то мере может быть объяснено их историческими ролями. Они использовались для датирования событий и последовательно остаются в архивах. Естественно, имеются все причины поддерживать такие архивы в их первоначальной форме. Это справедливо для астрономических наблюдений, датированных в соответствующее время тех мест, где они проводились. Если мы хотим их использовать, даты могут быть преобразованы во всемирное время, или эфемеридное время, хотя это приводит в некоторой степени к интерпретации и результат поэтому неоднозначен, в то время как оригинальные документы остаются аутентичными. Что же касается атомного времени, то оно существует только с 1955 года.

Другая причина в том, что разные шкалы времени, основанные на разных физических явлениях, сохраняют присущее им значение. Всемирное время представляет вращение Земли. Эфемеридное время и различные координатные времена, определенные в контексте общей теории относительности, являются динамическими временами, связанными с динамическими теориями. Кто знает, может быть их долговременные сличения с атомным временем приведут к исключительно важным данным? В том же духе теперь конструируется *пульсарная шкала*.

Выживание шкал времени, которые использовались раньше, не является проявлением консерватизма. Также и определение новых шкал не является знаком их совершенства, а требуется их приложениями. Мы должны принять, что мы входим в эру многочисленных шкал времени. К несчастью для читателя, мы не способны к упрощению без чувства вины за пробелы. По крайней мере, должно быть оценено то, что идея всемирного времени была сохранена для общественного использования.

Глава 5

Время и часы

5.1. Введение

Основной единицей международной системы единиц (единиц СИ) в области времени является единица длительности – секунда. Поэтому было бы уместно ссылаться на атомные стандарты, которые обеспечивают эту единицу, как на стандарты времени, или стандарты интервалов времени. Однако на практике эти термины используются редко по сравнению с термином *стандарт частоты*, оказавшимся более предпочтительным, потому что фактически они в действительности обеспечивают нас частотой.

Согласно официальному предписанию [5.1], мы должны отличать *первичный стандарт,* «который спроектирован или широко известен как имеющий высочайшее метрологическое качество и значение частоты которого принято без ссылки на другие стандарты такого же качества», от вторичных стандартов, «значение частоты которых приписано им путем сличения с первичным стандартом такого же качества». Мы будем следовать этим предписаниям даже тогда, когда качественные аспекты в определении первичного стандарта до некоторой степени затруднительны. И действительно, существуют стандарты частоты, которые, конечно же, не вторичные, но которые также ни в каком отношении не могут рассматриваться как первичные.

Любой, кто создает стандарт частоты, с целью обеспечения спектральной чистоты обычно пытается создать такую электромагнитную волну, чтобы она была как можно ближе к синусоидальной, и чтобы значение ее частоты было как можно ближе к номинальному значению. Но различные периоды этой волны не идентифицированы и не пронумерованы. По определению, часы обеспечивают нас дальнейшей информацией через точный счет этих периодов. Это обычно достигается путем нумерации импульсов с частотой 1 Гц, которые называются *секундные метки* или более знакомо *секундные импульсы*. Слово «часы» также подразумевает идею непрерывной работы на долгие периоды, например, несколько лет. Часы поэтому производят шкалу времени (их *собственное время*, в релятивистской терминологии).

Мы начинаем эту главу с обсуждения способов описания нестабильности частоты генераторов, и с флуктуаций шкал времени часов. В этом случае мы охарактеризуем стабильность различных представленных стандартов, включая, в частности, кварцевые генераторы, играющие ключевую роль в измерениях времени на любых уровнях качества.

Будет удобно распространить результаты нашего изучения шкал времени на шкалы, которые не производятся реальными часами, а также на шкалы, имеющие астрономическое происхождение. Мы будем затем рассматривать шкалы, происходящие от придуманных (фиктивных) часов. Например, мы увидим что (частотная) нестабильность международного атомного времени составляет около 10^{-15} за период усреднения 40 дней, хотя слово «частота» очень часто и совершенно несправедливо здесь опускается.

Когда мы рассматриваем координатную шкалу времени, мы должны вообразить, что она происходит от собственного времени фиктивных часов. Затем мы должны в используемой координатной системе применить теоретические соотношения между собственным временем и координатным временем. Тогда характеристики нестабильности и неточности будут относиться к этим фиктивным часам.

Хотя работа по нестабильности была стимулирована прогрессом, достигнутым в стандартах частоты, она имеет применение ко всем тем областям, в которых привлекаются долговременные серии измерений, в особенности в геофизике. На параметры, описывающие вращение Земли можно сослаться как на образец долговременных серий, и одним из параметров является само всемирное время.

Когда измерения времени были основаны на астрономии, первичный стандарт обеспечивался движениями небесных тел. Эти движения обеспечивали нас временем, т.е. таким же временем, как мы его представляем, когда смотрим на свои наручные часы. (Неопределенность терминологии, используемой при разговоре о времени совершенно недопустима!) Частота является величиной, которая выводится из этого времени, так что все стандарты частоты становятся тогда вторичными. Но появление атомных стандартов частоты порождает новую проблему, касающуюся точности их частоты. Точность часто воспринимается как качественное понятие. Однако при изучении частоты она имеет точное определение, что является несколько необычным для этой величины. Неточность может быть охарактеризована количественно, как это будет объяснено в разделе 5.5.

В окончании главы мы рассмотрим, как могут сличаться стандарты частоты и часы. Другой тип фиктивных часов, который будет нам встречаться в контексте этой книги – это часы, связанные с сигналами времени. Эти сигналы, передающиеся по радио или распространяющиеся по

кабелю, или еще каким-либо способом, состоят из меток времени, несущих дату. Поэтому их можно рассматривать как часы. Однако задержки распространения приводят нас к тому, что мы должны указать, когда мы рассматриваем сигналы. Мы можем рассматривать их или в момент излучения, или в момент приема в данном месте.

5.2. Нестабильность частоты и времени

5.2.1. Определения

Флуктуации частоты генератора в окрестности его номинальной частоты появляются в результате возмущений, таких как тепловой шум электронных компонент, шум, присущий резонатору, который обеспечивает генерацию частоты, старение инструмента и изменения окружающей среды. Эти флуктуации составляют то, что мы называем *нестабильность частоты*.[1]

Если генератор работает в качестве часов, то по сравнению с идеальным временем, сконструированным на базе его номинальной частоты, в нем происходят флуктуации в реальной шкале времени, которую он генерирует. Это проявление известно как *нестабильность времени.*

Как для времени, так и для частоты мы можем различать случайные и детерминистические нестабильности. Первые относятся к флуктуациям, которые кажутся случайными. Последние относятся к таким флуктуациям, которые выглядят как функции внешних параметров, таких как возраст и температура, хотя и не известно априори что производит флуктуации (иначе мы могли бы внести соответствующие поправки). В дальнейшем мы будем касаться главным образом случайных компонент, предполагая, что детерминистические компоненты, если таковые будут обнаружены, будут измерены и учтены.

Свойства нестабильности времени и частоты могут определяться либо во *временной области*, либо в *спектральной области*. Во временной области мы изучаем поведение частоты, усредняемой за различные промежутки времени. Естественно, что минимальное время усреднения определяется условиями измерения. В спектральной области мы используем свойства фурье-преобразований функции, представляющей вариацию частоты или девиацию времени. Здесь также мы имеем только дискретную последовательность измеренных величин, которая подразумевает существование (фурье) частоты среза.

Сначала 1960-х годов было выполнено большое число теоретических и экспериментальных работ по характеристикам нестабильности частоты

---

[1] Хотя это и не строго точно, допускается выражение «стабильность частоты» вместо «нестабильность частоты».

и нестабильности времени генераторов. Эти исследования дали возможность специалистам достичь согласия в вопросах, касающихся использования вариаций, позволяющих оценить характеристики нестабильности. Полученные результаты были приняты к употреблению различными профессиональными и международными организациями [5.2, 5.3].

5.2.2. Условные обозначения

Читатели должны сейчас приготовиться к тем трудностям, от которых авторы были неспособны их оградить, а именно к различиям в обозначениях. В этом разделе, посвященном статистическим характеристикам осцилляторов, мы приняли условные обозначения, установленные общим использованием. К несчастью, иногда они отличаются от обозначений, обычно используемых в отношении тех же самых физических величин, но в другом теоретическом контексте, таком как релятивистская физика, шкалы времени и стандарты частоты. Проблема в том, что имеющийся алфавит недостаточен для того, чтобы представить весь диапазон параметров и математических моделей, необходимых для того, чтобы описать физическую реальность. Вероятно, что даже большая путаница может быть вызвана изменением традиционных обозначений, и обозначения специфические для каждой области по этой причине будут сохраняться. Читатель будет предупреждаться, когда символ изменяет значение таким образом, что это может вызвать неправильное толкование.

Сигнал V_C производимый осциллятором C может быть записан

$$V_C(t) = V_{0,C}\sin[2\pi\nu_{0,C}t + \phi_C(t)], \tag{5.1}$$

где мы подразумеваем, что амплитуда постоянна. Величина $\nu_{0,C}$ есть константа номинальной частоты и t есть идеальное собственное время, основанное на определении секунды. Оно эквивалентно τ, используемому в этом смысле в предыдущих главах. Фаза ϕ_C, по сравнению с идеальной периодической вариацией, содержит случайные и систематические девиации.

Мгновенная частота $\nu_C(t)$ определяется как

$$\nu_C(t) = \nu_{0,C} + \frac{1}{2\pi}\frac{d\phi_C(t)}{dt}. \tag{5.2}$$

Относительная отстройка частоты осциллятора, впервые введенная в разделе 4.4.2а дается

$$y_C(t) = \frac{\nu_C(t)}{\nu_{0,C}} - 1. \tag{5.3}$$

Она представляет собой относительную величину мгновенного ухода осциллирующей частоты от ее номинального значения. Эта величина очень мала для высокостабильных осцилляторов, которые мы будем рассматривать.

Осциллятор часто обслуживает диапазон частот с установленными номинальными значениями, например, 1 МГц, 1 кГц, 1 Гц. Благодаря конструкции фактические частоты остаются строго в том же соотношении, что и номинальные частоты и поэтому здесь используется одна нормализованная частота для всех этих значений.

Показания часов возрастают на номинальное значение $1/\nu_{0,C}$ периода каждый раз, когда суммарная фаза $2\pi\nu_{0,C}t+\phi_C(t)$ возрастает на 2π. Поэтому время, показываемое с самого начала, выражается через

$$t_C = t + x_C(t), \tag{5.4}$$

где мы установили

$$x_C(t) = \frac{\phi_C(t)}{2\pi\nu_{0,C}}. \tag{5.5}$$

Величина t_C является собственным временем часов С. Девиация собственного времени из-за неидеальности осциллятора равна таким образом $x_C(t)$. Это временная нестабильность часов по сравнению с идеальной опорной шкалой времени.

Из вышеприведенных определений легко показать, что

$$y(t) = \frac{dx(t)}{dt}, \tag{5.6}$$

где $y(t)$ представляет относительную отстройку частот осциллятора или часов, как функцию от t и от девиации собственного времени $x(t)$.

5.2.3. Измерения во временной области

Было разработано много методов измерения нестабильности частоты и времени [5.4, 5.5, 5.6]. Для каждого метода требуется два генератора частоты (или двое часов), один из которых берется в качестве опорного. Мы начнем с предположения, что опорный генератор имеет пренебрежимо малую нестабильность частоты по сравнению с генератором, который необходимо исследовать. Противоположный случай мы рассмотрим в разделе 5.2.8.

Ограничим себя с одной стороны иллюстрацией измерений нестабильности времени, и с другой стороны измерением нестабильности частоты, двумя общепринятыми и легко интерпретируемыми методами.

(а) Измерение нестабильности времени

Экспериментальная установка схематически показана на рис. 5.1. Исследуемые часы и опорные часы генерируют периодические импульсы, т.е. секунды. Часы подключены к двум входам измерителя интервалов времени. Этот прибор регистрирует моменты, когда величина импульса достигает предопределенного уровня. Счет запускается прибытием импульса от часов 1 на вход 1 и останавливается прибытием импульса от часов 2 на вход 2. Измеритель интервалов времени обычно управляется опорными часами.

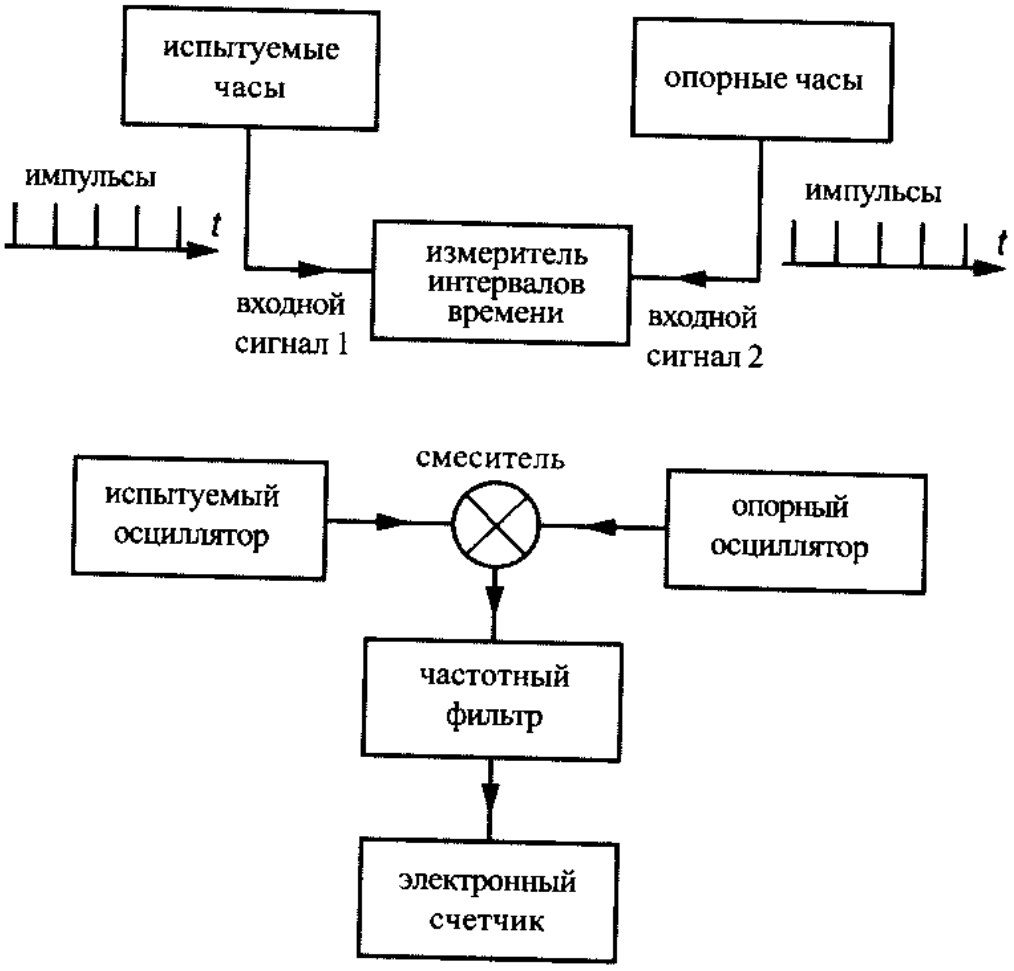


Рис. 5.1. Схема измерения нестабильности времени (вверху) и нестабильности частоты (внизу)

Измерение дает величину x_k от $x(t)$ по модулю одной секунды за время t_k, когда импульс, произведенный испытываемыми часами, поступает на вход 2. При использовании лучших приборов неопределенность измеряемой величины составляет около 10 пикосекунд (пс). Систематические ошибки возникают из-за задержек распространения сигналов в соединительных кабелях. Если кабели одни и те же, то это не имеет значения для статистического анализа временной нестабильности, так как последняя, как мы вскоре увидим, относится к разностям между величинами x_k.

Набор величин x_k, полученных таким способом, составляет временнэю серию, определенную как равно расположенные времена t_k. Следуя общей практике в области определения параметров нестабильности, интервал времени между двумя последовательными величинами t_k обозначим через τ. Строго говоря, это есть интервал собственного времени и согласно обозначениям, введенным в главе 3, он должен быть обозначен как $\Delta\tau$.

(b) Измерения нестабильности частоты

Метод, который будет описан, подходит в том случае, когда два осциллятора, или, вообще говоря, два генератора генерируют различные, но очень близкие частоты. Например, их значения могут быть близки к 10 МГц, в то время как разность их частот составляет около 1 Гц.

Пусть $\nu_C(t)$ есть частота исследуемого генератора и ν_R – частота опорного генератора, которая рассматривается как константа. Как показано на рис. 5.1, генераторы подключаются к смесителю. Смеситель является нелинейным элементом, на выходе которого в первом приближении появляется произведение двух сигналов, подключенных к его входам. Его выходной сигнал содержит спектральные компоненты на частотах $|\nu_C - \nu_R|$ и $\nu_C + \nu_R$. Низкочастотная компонента выделяется с помощью фильтра и мы, таким образом, детектируем частоту биений $\nu_b = |\nu_C - \nu_R|$.

Относительное изменение в ν_b может быть записано в виде

$$\frac{\Delta\nu_b}{\nu_b} = \frac{\nu_C}{\nu_b}\frac{\Delta\nu_C}{\nu_C}. \qquad (5.7)$$

Это значительно больше, чем относительное изменение в ν_C из-за наличия коэффициента умножения ν_C/ν_b. Используя численные значения, предложенные выше, получим, что этот коэффициент составляет по порядку 10^7.

Частота ν_b измеряется с помощью электронного частотомера, у которого временнЬй блок управляется опорным генератором. Частотомер измеряет длительность n периодов входного сигнала. Их длительность равна $\tau = n/\nu_b$. Таким образом, частотомер дает усредненное за интервал τ, значение частоты ν_b, обозначаемое как $\bar{\nu}_b^{\tau}$.

Неопределенность в этом измерении зависит от соотношения сигнал/шум в сигнале биений. С соответствующим оборудованием эта неопределенность может быть близка к $10^{-7}/\tau$, если частота ν_b лежит вблизи 1 Гц и ширина полосы низкочастотного фильтра равна 10 Гц. Отсюда следует, что неопределенность величины $\Delta\nu_b/\nu_b$ составляет около $10^{-14}/\tau$ в рассмотренном здесь примере. Часто ею можно пренебречь.

Чтобы упростить вопрос давайте предположим, что мертвое время между двумя измерениями очень мало, по сравнению с интервалом вре-

мени τ. Последовательные измерения производят, таким образом, серию величин ν_b^τ, с дискретными временами t_k. такими, что $t_{k+1} - t_k = \tau$. Далее мы извлечем среднюю величину ν_b^τ, а, следовательно, и ν_b. Используя (5.3) и (5.7) мы можем установить временну́ю серию значений относительной расстройки частоты поверяемого генератора, усредненной на интервале времени τ.

Так как величина y_k^τ определена между временами t_k и t_{k+1}, то уравнение (5.6) показывает, что выполняется следующее соотношение между измеренными значениями нестабильности частоты y_k^τ и нестабильностью времени x_k:

$$\bar{y}_k^\tau = \frac{1}{\tau}[x_{k+1} - x_k]. \tag{5.8}$$

5.2.4. Измерения в спектральной области

Измерения в спектральной области [5.4, 5.5, 5.6] дают нам спектральную плотность мощности[2] флуктуаций поверяемого времени и частоты. Они обозначаются через $S_x(f)$ и $S_y(f)$ соответственно. Частота f является частотой анализа, известной также как фурье-частота или фурье-компонента, и она не должна смешиваться с частотой генерации. Из (5.6)

$$S_y(f) = 4\pi^2 f^2 S_x(f). \tag{5.9}$$

5.2.5. Моделирование частотных и временных флуктуаций

Экспериментально можно получить только ограниченное число данных y_k^τ, так как общее время измерений оказывается конечным. Их может быть совсем немного в случае, если временная или частотная нестабильность определяется за период τ, превышающий несколько дней. Поэтому можно вычислить только оценку вариации исследуемых флуктуаций (см. раздел 5.2.6). Эта оценка сама по себе является случайной переменной в том смысле, что при повторении измерений будут получаться разные ее значения. Неопределенность увеличивается, когда число измерений уменьшается. Становится трудно, почти невозможно, вывести спектральную плотность мощности частотных или временных флуктуаций для низких значений фурье-компонент частоты. Однако они часто необходимы, когда изучаются свойства атомных стандартов времени и частоты, используе-

[2] Спектральная плотность мощности случайного процесса на частоте f равна $\sigma^2/\Delta f$, где σ^2 является среднеквадратичной флуктуацией процесса, пропущенного через очень узкий фильтр с окном шириной Δf, с центром на частоте f.

мых в таких системах, как отдельные навигационные и телекоммуникационные системы. Это будет рассмотрено в главе 9.

Для того чтобы обойти эту проблему была разработана математическая модель флуктуаций времени и частоты. Справедливость этой модели была оценена проверкой согласия экспериментальных результатов с моделью или ее предсказаниями в пределах экспериментальной неопределенности.

Модель, которая дает удовлетворительный расчет случайных нестабильностей частоты рассматриваемых генераторов, имеет следующие характеристики. Процесс является эргодическим. Нестабильность частоты, представляемая $y(t)$, имеет нулевое среднее значение. Плотность вероятности $y(t)$ является гауссовской. Наконец, спектральная плотность мощности относительных флуктуаций частоты представляется полиномом от f, содержащим до пяти членов. Для $f > 0$,

$$S_y(f) = \sum_{\alpha=-2}^{2} h_\alpha f^\alpha , \tag{5.10}$$

где α – целое число в диапазоне от –2 до +2. Значение констант h_α зависит от источника. Они могут быть равны нулю для некоторых величин α. Соответствующая величина $S_x(f)$ получается из (5.9) и (5.10).

Таблица 5.1 дает общепринятые термины для различных типов независимых шумов, которые могут давать вклад в нестабильность времени и частоты.

Физическое происхождение белого фазового и частотного шумов понятно. Они связаны с тепловым шумом в радиодиапазоне или дробовом шумом от потока частиц. Происхождение других шумов известно в меньшей степени. Однако большое число экспериментальных результатов является свидетельством их реальности.

Таблица 5.1. Термины, используемые для описания пяти шумовых процессов, дающих вклад в нестабильность времени и частоты

α	Шум
2	Белый фазовый шум
1	Фликкерный фазовый шум
0	Белый частотный шум
-1	Фликкерный частотный шум
-2	Шум случайного блуждания частоты

5.2.6. Характеристики нестабильности частоты

Вообще говоря, вариация случайного процесса связана с интегралом от его спектральной плотности мощности по всему диапазону частот. Для $y(t)$ мы находим, что

$$E\{y^2(t)\} = \int_0^\infty S_y(f)\,df\,, \tag{5.11}$$

где $E\{\}$ обозначает математическое ожидание. Сразу же становится очевидным, что, когда $S_y(f)$ имеет вид (5.10), интеграл расходится. Для $\alpha \geq -1$, расходимость имеет место для $f \to \infty$. Однако флуктуации частоты всегда фильтруются низкочастотным фильтром. Поэтому такая математическая трудность не соответствует реальной физической проблеме. Более серьезна расходимость при $f \to 0$, когда $\alpha = -1, -2$. Конечно, частота, равная нулю никогда не может быть достигнута, но можно показать, что математическая расходимость предполагает, что оцениваемое значение вариации зависит от общего времени измерения. Это очень нехорошо. Ситуация становится даже хуже, когда мы рассматриваем $E\{x^2(t)\}$ потому что спектральная плотность мощности $S_x(f)$ содержит члены с $f^{\alpha-2}$. Соответствующий интеграл поэтому расходится на низкочастотной границе для четырех или пяти значений α, т.е. для $\alpha \leq 1$.

Можно показать, что для $\alpha \leq 1$, приращения первого или второго порядка в $x(t)$, а следовательно и в фазовых флуктуциях являются стационарными (см. (5.5)). При этом устраняются вышеуказанные трудности, если для статистического истолкования флуктуаций времени и частоты привлечь средний квадрат этих приращений. Тогда могут быть рассмотрены многие вариации. Мы обсудим только наиболее общеупотребительные.

(а) Определение вариации Аллана

Мы начнем с обсуждения вариации Аллана, известной также как вариация двойной выборки. В ее определении предполагается неограниченное число выборок. Она равна половине средней величины двойной выборки величин y_k^τ, где выборка берется из соседних серий, т.е. с нулевым мертвым временем между последовательными величинами[3]. Мы имеем [5.7]

---

[3] Коэффициент 1/2 происходит от того, что двойная выборка является частным случаем более общей вариации из N выборок. Когда случайные переменные есть относительные частотные флуктуации, усредненные за интервал τ, N-выборочная вариация есть

$$\sigma^2(N,\tau) = \frac{1}{N-1}\sum_{n=1}^{N}\left(\bar{y}_n - \frac{1}{N}\sum_{k=1}^{N}\bar{y}_k\right)^2.$$

$$\sigma_y^2(\tau) = \frac{1}{2} E\left\{[\bar{y}_{k+1}^{\tau} - \bar{y}_k^{\tau}]^2\right\}. \tag{5.12}$$

Используя (5.8) мы получим

$$\sigma_y^2(\tau) = \frac{1}{2\tau^2} E\left\{[x_{k+2} - 2x_{k+1} + x_k]^2\right\}, \tag{5.13}$$

что приводит к приращению второго порядка в x_k. Уравнения (5.12) и (5.13) могут использоваться для характеристик нестабильности частоты во временньй области.

Вариация Аллана может быть выражена также в терминах спектральной плотности мощности $S_y(f)$. Можно показать, что

$$\sigma_y^2(\tau) = \int_0^{\infty} \left|H_A(f)\right|^2 S_y(f)\mathrm{d}f \tag{5.14}$$

где $|H_A(f)|^2$ есть модуль квадрата передаточной функции, которая представляет в частотной области свойства цифровой фильтрации, связанной с вычислением вариации Аллана. Эта передаточная функция дается следующим выражением:

$$\left|H_A(f)\right|^2 = 2\frac{\sin^4(\pi f\tau)}{(\pi f\tau)^2}. \tag{5.15}$$

Когда $f \to 0$, $|H_A(f)|^2$ изменяется как f^2, гарантируя сходимость интеграла в нижнем пределе для всех типичных значений α.

Таблица 5.2. Соответствие между различными компонентами шумов, возмущающих осциллятор, и вариацией Аллана для флуктуаций нормализованной частоты в случае, когда $2\pi f_h \tau >> 1$.

$S_y(f)$	$\sigma_y^2(\tau)$
$h_2 f^2$	$3\, h_2 f_h / 4\, \pi^2\, \tau^2$
$h_1 f$	$h_1[1{,}04+3\ 1\mathrm{n}\ (2\, \pi f_h\, \tau)] / 4\pi^2\, \tau^2$
h_0	$h_0 / 2\, \tau$
$h_{-1} f^{-1}$	$2\, h_{-1} \ln 2$
$h_{-2} f^{-2}$	$2\, \pi^2\, h_{-2} \tau / 3$

Низкочастотный фильтр, с частотой среза f_h, который всегда присутствует на входе или внутри оборудования, измеряющего $\bar{y}_k^{\tau}$, принимается во внимание путем умножения $|H_A(f)|^2$ на передаточную функцию этого фильтра, обозначенную как $|H_F(f)|^2$. Мы положим здесь, что

$$|H_F(f)|^2 = \begin{cases} 1 & \text{для } f < f_h, \\ 0 & \text{для } f \geq f_h. \end{cases} \quad (5.16)$$

Такая фильтрация оказывает влияние на значение вариации Аллана только для $\alpha = 1, 2$, при условии, что

$$f_h >> \frac{1}{2\pi\tau} \quad (5.17)$$

(b) Изменение вариации Аллана с изменением периода выборки τ

Вычисляя интеграл с правой стороны (5.14) с заменой $|H_A(f)|^2$ на $|H_A(f)|^2 \times |H_F(f)|^2$, мы можем установить соответствие между $\sigma_y^2(\tau)$ и каждой компонентой шума, возмущающего осциллятор, при выполнении условия (5.17). Это соответствие показано в таблице 5.2.

Функциональная зависимость $S_y(f)$ и вариация Аллана $\sigma_y(\tau)$ обычно представляются на логорифмическом графике, который показывает их асимптотическое поведение (см. рис. 5.2). Когда данный тип шума доминирует, по крайней мере, на одной декаде значений f или τ, то из такого графика по наклону кривой определяется значение α. Затем оно может быть использовано, чтобы опознать присутствующий шум, если $\alpha \leq 0$. Например, отрезок графика $\sigma_y(\tau)$ с нулевым наклоном связан с фликер-шумом частоты и обычно называется фликерным потолком. Вариация Аллана изменяется как τ^{-1} для белого и фликерного фазовых шумов. Поэтому они не могут различаться таким способом. Единственное средство в этой ситуации воспользоваться разной зависимостью $\sigma_y(\tau)$ от полосы f_h низкочастотного фильтра. Другой способ – использовать модифицированную вариацию Аллана, которая будет обсуждаться ниже.

В вышеприведенном примере предполагалось, что между последовательными выборками y_k^τ отсутствует мертвое время. Если же это не так, как это и случается на практике, то за исключением случая белого частотного шума в вычисления $\sigma_y^2(\tau)$ вводятся смещения. Эти смещения табулированы в работе [5.5].

Рис. 5.6 и 6.11 показывают примеры того, как вариация Аллана изменяется с изменением τ.

(c) Практическое вычисление вариации Аллана

В реальном мире число возможных значений x_k или $\bar{y}_k^\tau$ является конечным. Следовательно, можно получить только оценку вариации Аллана. Теперь мы опишем, как оценка параметра $\sigma_y(\tau)$ может быть вычислена на практике.

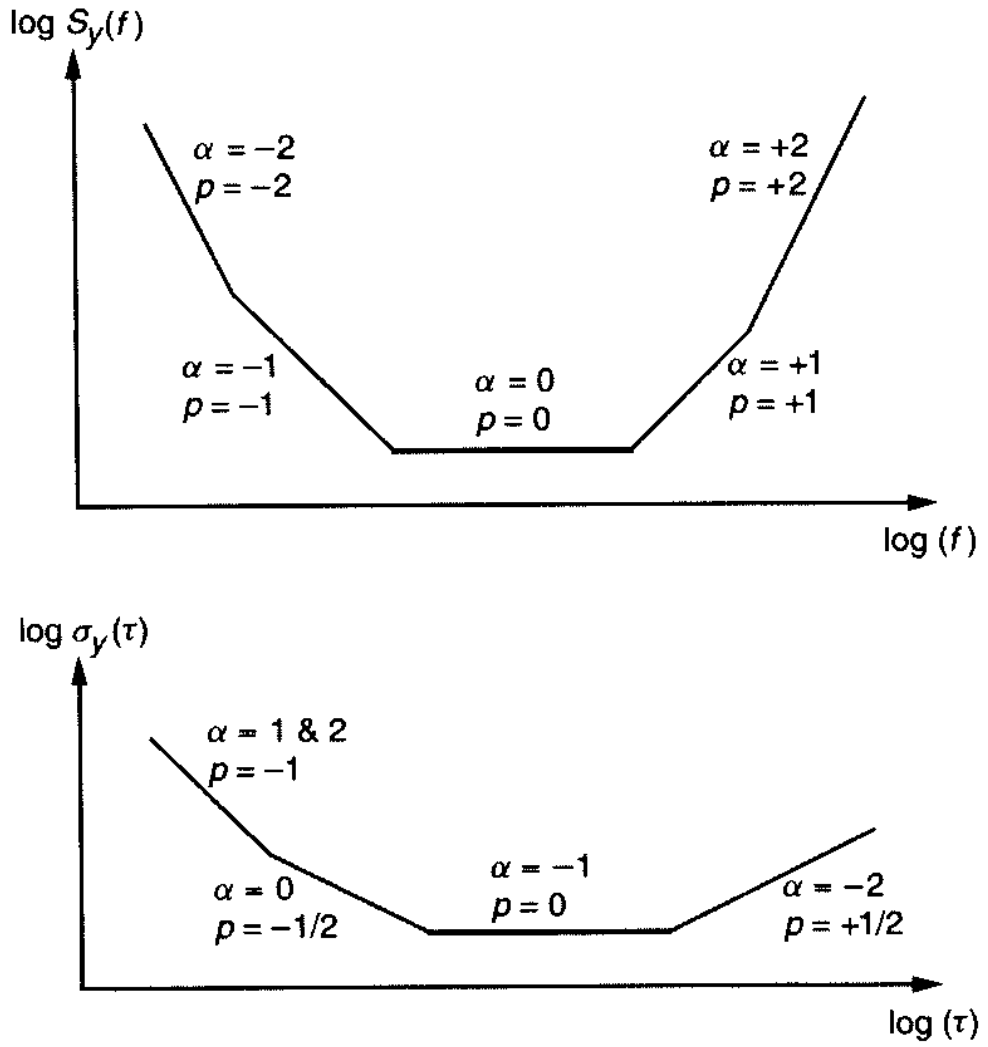


Рис. 5.2. Асимптотическое логарифмическое представление вариации $S_y(f)$ и $\sigma_y(\tau)$ с их аргументами f и τ для различных значений. Наклоны даются значениями p.

Вообще говоря, измерения не повторяются для различных значений τ. Фактически мы определяем временную серию из M+1 значений x_j за времена, разделенные интервалом τ_0, или временную серию из M значений $\bar{y}_j^{\tau_0}$ за период усреднения τ_0. Предположим, что мертвое время пренебрежимо мало. Рис. 5.3a показывает основные временные интервалы длительностью τ_0, за которые определяются значения $\bar{y}_k^{\tau_0}$.

Данные затем группируются вместе для того, чтобы вывести вариацию Аллана из $\tau = n\tau_0$, где n – положительное целое число.

Простейшая и наиболее непосредственная комбинация показана на рис. 5.3b, для n = 3. Временная серия формируется из соседних значений $\bar{y}_k^{\tau}$ путем вычисления средней вариации нормализованной частоты за интервал времени $\tau = n\tau_0$ по одному из следующих соотношений:

$$\bar{y}_k^{\tau} = \frac{1}{n\tau_0}[x_{nk+1} - x_{n(k-1)+1}] \tag{5.18}$$

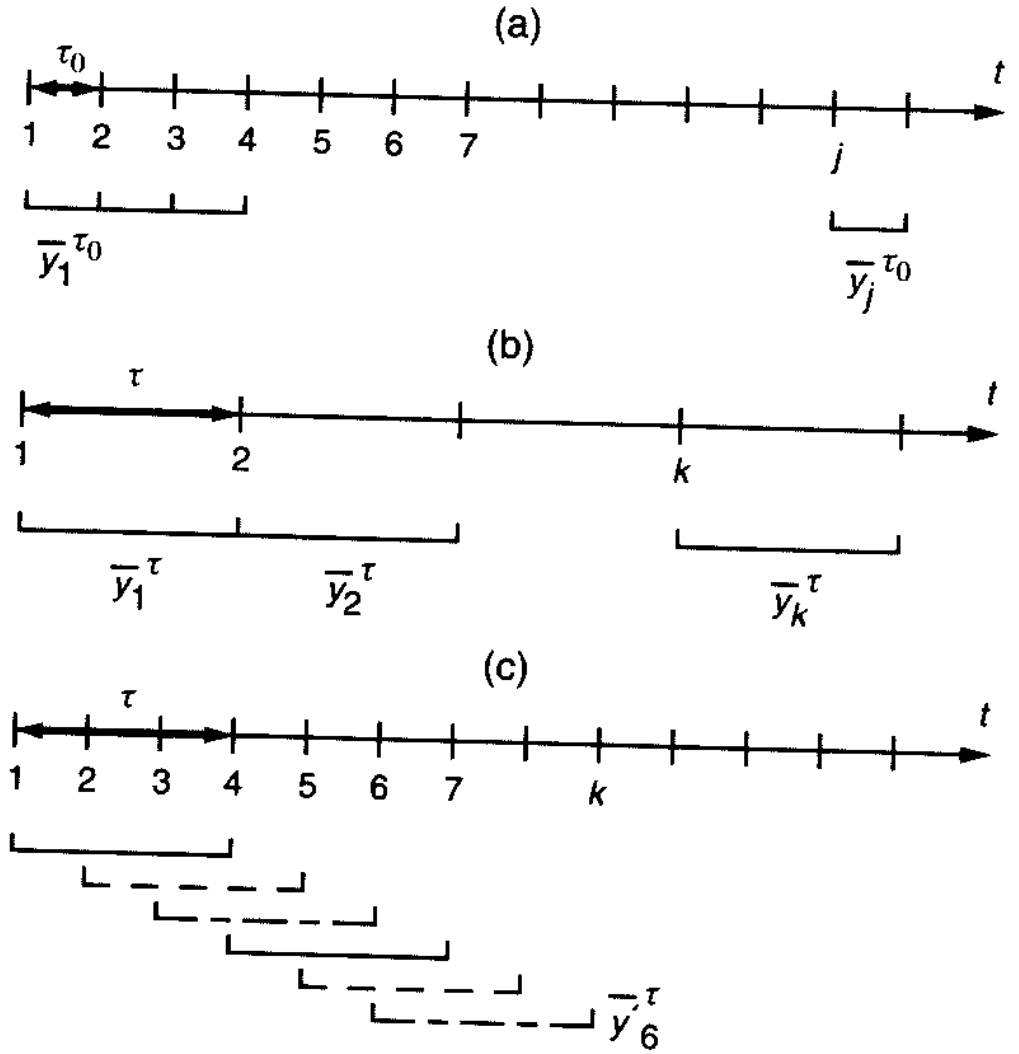


Рис. 5.3. Практическое вычисление вариации Аллана. (а) Интервалы времени, на которых определяется набор начальных данных. (в) Иллюстрация первого метода обработки данных.(с) Иллюстрация второго метода обработки данных

или

$$\bar{y}_k^{\tau} = \frac{1}{n} \sum_{j=n(k-1)+1}^{nk} \bar{y}_j^{\tau_0} . \tag{5.19}$$

Число значений пар y_k^{τ}, которое может быть образовано, дается из

$$p_1 = \text{int}\left(\frac{M}{n}\right) - 1. \tag{5.20}$$

Оценочное значение $\hat{\sigma}_y^2(\tau)$ вариации Аллана вычисляется затем по одному из следующих соотношений:

$$\hat{\sigma}_y^2(\tau) = \frac{1}{2p_1(n\tau_0)^2} \sum_{k=1}^{p_1} \left[x_{n(k+1)-1} - 2x_{nk+1} + x_{n(k-1)+1}\right]^2, \tag{5.21}$$

$$\hat{\sigma}_y^2(\tau)=\frac{1}{2p_1}\sum_{k=1}^{p_1}\left[\bar{y}_{k+1}^{\tau}-\bar{y}_k^{\tau}\right]^2. \qquad (5.22)$$

Определение характеристик одного и того же генератора с помощью различных временных серий одинаковой длины дает различные значения оценки вариации Аллана. Среднее значение оценки равно математическому ожиданию, определяемому из (5.12). Неопределенности в оценке $\hat{\sigma}_y(\tau)$ характеризуются его собственным стандартным отклонением, равным $K_\alpha/(p_1+1)^{1/2}$, где p_1 больше, чем несколько единиц [5.8]. Когда α варьируется от –2 до +2 то параметр K_α занимает диапазон от 0,75 до 0,99.

Другой способ объединения экспериментальных данных более часто используется на практике, потому что он приводит к лучшему использованию этих данных, приводящих к меньшей неопределенности [5.9]. Вместо того, чтобы сформировать набор значений $\bar{y}_k^{\tau}$, как указано выше, другой набор может быть получен из

$$\bar{y}'^{\tau}_k=\frac{1}{n\tau_0}(x_{k+n}-x_k), \qquad (5.23)$$

или, что то же самое, из

$$\bar{y}'^{\tau}_k=\frac{1}{n}\sum_{j=k}^{k+n-1}\bar{y}_j^{\tau_0}. \qquad (5.24)$$

На рис. 5.3с показаны интервалы времени, на которых определены новые значения $y_k'^{\tau}$, для $n = 3$. Эти интервалы перекрываются. Однако, для того, чтобы оценить вариацию Аллана, мы рассматриваем значения $y_k'^{\tau}$, принадлежащие соседним интервалам. Такими значениями будут, например, $\bar{y}_{k+n}'^{\tau}$ и $\bar{y}_k'^{\tau}$. Число таких пар p_2 дается из:

$$p_2=M-2n+1. \qquad (5.25)$$

Новая оценка вариации Аллана за период выборки τ вычисляется теперь путем использования следующих соотношений:

$$\hat{\sigma}_y^2(\tau)=\frac{1}{2p_2(n\tau_0)^2}\sum_{k=1}^{p_2}\left[x_{k+2n}-2x_{k+n}+x_k\right]^2, \qquad (5.26)$$

$$\hat{\sigma}_y^2(\tau)=\frac{1}{2p_2}\sum_{k=1}^{p_2}\left[\bar{y}'^{\tau}_{k+n}-\bar{y}'^{\tau}_k\right]^2. \qquad (5.27)$$

Можно показать, что эти два выражения имеют среднее значение, равное математическому ожиданию, определяемому из (5.12).

Метод оценки неопределенности в аллановской вариации, вычисленный таким способом, приводит к более сложным вычислениям, чем в предыдущем случае [5.10, 5.11]. Главный результат заключается в том, что для данных M и n, число пар p_2 значительно больше, чем p_1 (если мы не приближаемся к $M/2$) и отсюда следует, что относительная неопределенность $\hat{\sigma}_y(\tau)$ уменьшается.

(d) Вариация Аллана в присутствие систематических эффектов

Ничто не мешает нам вычислить вариацию Аллана в том случае, когда появляются относительные отстройки частоты $y(t)$, которые имеют особенности систематических вариаций. Например, возмущение может проявляться в виде линейного дрейфа частоты, типа

$$y(t) = at, \tag{5.28}$$

или синусоидальной модуляции, типа

$$y(t) = a' \cos 2\pi f_M t\,. \tag{5.29}$$

Параметры a и a' представляют наклон частотного дрейфа и относительную амплитуду частотной модуляции соответственно. Частота модуляции f_M такова, что $f_M \ll f_h$.

В случае линейного дрейфа определение (5.12) подразумевает

$$\sigma_y^2(\tau) = \frac{1}{2} a^2 \tau^2. \tag{5.30}$$

На графике с логарифмическими осями, представляющим вариацию $\sigma_y(\tau)$ мы тогда получим прямую линию с единичный наклоном. Добавленная к другим общим шумовым эффектам, она начинает играть доминирующую роль для больших значений τ.

Для синусоидального возмущения определение дает

$$\sigma_y^2(\tau) = a'^2 \frac{\sin^4(\pi f_M \tau)}{(\pi f_M \tau)^2}. \tag{5.31}$$

На графике $\sigma_y^2(\tau)$ его эффект проявляется в переменных вариациях около прямой линии с наклоном – 1. Такое поведение происходит при значениях τ порядка $1/f_M$.

Поэтому, характеризуя частотную нестабильность с помощью вариации Аллана или девиации, мы можем идентифицировать наличие определенных типов систематических возмущений, влияющих на частоту генератора.

(e) Модифицированная вариация Аллана

Одним из недостатков вариации Аллана является то, что она изменяется так же как τ^{-2}, когда шум, воздействующий на генератор, является белым

фазовым шумом ($\alpha = 2$), или фликкерным фазовым шумом ($\alpha = 1$). Для того, чтобы при ее использовании отличить эти два типа фазовых шумов, серии измерений должны быть проведены с разными значениями частотной полосы пропускания f_h измерительной системы.

Было показано, этот недостаток может быть устранен путем введения новой вариации, известной как модифицированная вариация Аллана, и обозначаемая как $\mathrm{Mod}\sigma_y^2(\tau)$. Определяющий алгоритм делает возможным изменение полосы пропускания виртуальным путем [5.5].

При $\tau = n\tau_0$ мы имеем

$$\mathrm{Mod}\sigma_y^2(\tau) = \frac{1}{2} E\left\{\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{k=1}^{n}\bar{y}_{i+k+n}^{\tau_0} - \frac{1}{n}\sum_{k=1}^{n}\bar{y}_{i+k}^{\tau_0}\right)\right]^2\right\}, \quad (5.32)$$

или

$$\mathrm{Mod}\sigma_y^2(\tau) = \frac{1}{2}\frac{1}{(n\tau_0)^2} E\left\{\left[\frac{1}{n}\sum_{i=1}^{n}(x_{i+2n} - 2x_{i+n} + x_i)\right]^2\right\}. \quad (5.33)$$

Для $n = 1$ мы возвращаемся к выражению для вариации Аллана. Таблица 5.3 дает функциональную зависимость $\mathrm{Mod}\sigma_y^2(\tau)$ от τ для различных типов шумов.

Можно увидеть, что эта зависимость различается для белого фазового шума ($\alpha = 2$) и для фликкерного фазового шума ($\alpha = 1$). Эти типы шумов могут таким образом устанавливаться и различаться однозначно. Например, можно измерить уровень белого фазового шума, связанного с конечной разрешающей способностью счетчика интервалов времени, используемого непосредственно в измерениях нестабильности времени. (см. раздел 5.2.3а).

Таблица 5.3. Соответствие между различными компонентами шумов, возмущающих генератор и модифицированной вариацией Аллана для нормализованных частотных флуктуаций когда $2\pi f_h \tau >> 1$. Выражение для $\mathrm{Mod}\sigma_y^2(\tau)$ является точным для белого фазового шума, при любом значении $n = \tau/\tau_0$. Для других типов шумов приведенные выражения справедливы в пределах нескольких процентов при $n \geq 10$.

$S_y(f)$	$\mathrm{Mod}\sigma_y^2(\tau)$
$h_2 f^2$	$3\, h_2 f_h \tau_0 / 4\, \pi^2 \tau^3$
$h_1 f$	$h_1[1{,}04+3\ \ln(2\,\pi f_h\,\tau)] / 4\pi^2\,\tau^2$
h_0	$h_0 / 4\,\tau$
$h_{-1} f^{-1}$	$0{,}936\, h_{-1}$
$h_{-2} f^{-2}$	$5{,}42\, h_{-2}\tau$

Для других типов шумов ($\alpha \leq 1$) не существует особого преимущества при использовании модифицированной вариации Аллана. Отметим, что для $\alpha \leq 0$ числовые коэффициенты не одни и те же для вариации Аллана и модифицированной вариации Аллана.

5.2.7. Характеристики нестабильности времени

Нестабильность времени обычно характеризуется через вариацию Аллана или модифицированную вариацию Аллана. Мы приведем примеры применения каждой из этих вариаций.

(a) Нестабильности в прогнозе времени

При изучении шкал времени и работе с цифровыми телекоммуникационными сетями часто бывает необходимо прогнозировать значение расхождения времени между двумя часами в некоторые поздние даты. Простейший подход состоит в выполнении линейной экстраполяции. Положим

$$\hat{x}(t+\tau) = x(t) + \tau\, \bar{y}_k^{\tau}(t), \tag{5.34}$$

где $x(t)$ и $\hat{x}(t+\tau)$ являются отклонениями времени между двумя часами в настоящее время t и прогнозируемое значение отклонения в более позднее время $t+\tau$, соответственно. Величина $\bar{y}_k^{\tau}(t)$ является значением относительной расстройки частоты между двумя часами, усредненной по времени за период от $t-\tau$ до t. Прогноз не может быть полноценным из-за присутствия случайной ошибки $\Delta x(t, \tau)$, равной

$$\Delta x(t,\tau) = x(t+\tau) - \hat{x}(t+\tau), \tag{5.35}$$

где $x(t+\tau)$ является отклонением времени, которое будет фактически наблюдаться в будущем времени $t+\tau$. Величина $\Delta x(t,\tau)$ является *ошибкой прогноза времени.*

Вариация $\Delta x(t,\tau)$ характеризует неопределенность в предсказании [5.12]. Из (5.8) и (5.13) мы получим

$$E\left\{\Delta x^2(t,\tau)\right\} = 2\tau^2 \sigma_y^2(\tau). \tag{5.36}$$

Стандартное отклонение $\Delta x(t,\tau)$ является, таким образом, независимым от τ для белого и фликкерного фазового шумов. Оно изменяется как $\tau^{1/2}$ для белого частотного шума, оно пропорционально τ для фликерного частотного шума и оно работает как $\tau^{3/2}$ для шума типа случайного блуждания частоты.

(b) Вариация нестабильности времени

Для того чтобы охарактеризовать свойства систем передачи и распространения времени, а также телекоммуникационных систем, считалось по-

лезным ввести вариацию нестабильности времени, обозначаемую как TVAR. (Отметим, что вариация Аллана и модифицированная вариация Аллана иногда обозначаются акронимами как AVAR и MVAR, соответственно.) Она определяется как [5.13]

$$\mathrm{TVAR} = \frac{\tau^2}{3}\ \mathrm{Mod}\sigma_y^2(\tau). \tag{5.37}$$

5.2.8. Определение вклада от каждого генератора

Для измерения нестабильности времени и частоты необходимо два генератора. Мы предполагаем, что измерительная система не дает вклада в наблюдаемые флуктуации.

Если нестабильность генератора 1 много меньше, чем генератора 2, вся измеренная нестабильность приписывается генератору 2. Если два генератора изготовлены одинаковым способом, тогда, исходя из желания иметь лучшую гипотезу, мы можем предположить, что они имеют одинаковое качество. К тому же нет причин предполагать, что их флуктуации коррелированы, так что вариации измеренных флуктуаций равны удвоенным флуктуациям любого индивидуального генератора.

Есть возможность разделить индивидуальный вклад, если мы располагаем более чем двумя генераторами. Давайте, как пример, рассмотрим случай с тремя генераторами. Пусть σ_i^2, σ_j^2, σ_k^2 являются вариациями Аллана, присущими каждому генератору для данного τ, и σ_{ij}^2, $\sigma_{jk}^2, \sigma_{ki}^2$ являются вариациями Аллана для всех попарных сличений за то же самое время τ. При условии, что флуктуации не коррелированы, мы находим

$$\begin{cases} \sigma_{ij}^2 = \sigma_i^2 + \sigma_j^2, \\ \sigma_{jk}^2 = \sigma_j^2 + \sigma_k^2, \\ \sigma_{ki}^2 = \sigma_i^2 + \sigma_k^2. \end{cases} \tag{5.38}$$

Мы можем далее вывести индивидуальные вариации

$$\begin{cases} \sigma_i^2 = \frac{1}{2}\left(\sigma_{ij}^2 + \sigma_{ik}^2 - \sigma_{jk}^2\right), \\ \sigma_j^2 = \frac{1}{2}\left(\sigma_{jk}^2 + \sigma_{ij}^2 - \sigma_{ik}^2\right), \\ \sigma_k^2 = \frac{1}{2}\left(\sigma_{jk}^2 + \sigma_{ik}^2 - \sigma_{ij}^2\right) \end{cases} \tag{5.39}$$

Этот метод дает удовлетворительный результат, если выполняются следующие условия:

- генераторы имеют сравнительно одинаковое качество;
- их флуктуации не коррелированы;
- измерения выполняются одновременно;
- неопределенность в оценке вариаций $\sigma_{ij}^2, \sigma_{jk}^2$ и σ_{ki}^2 достаточно мала.

Этот метод известен как *метод треуголки* (шляпы).

5.3. Механические осцилляторы

Наилучшие достижения в механических осцилляторах были получены в маятниковых часах в астрономических обсерваториях. Эти часы достигли своего предельного уровня идеальности в начале двадцатого столетия. Мы не будем описывать эти прекрасные инструменты. Интересная ссылка по ним дана в [5.14]. Давайте только напомним некоторые меры предосторожности, предпринятые для того чтобы обеспечить регулярность их работы. Часы размещались в герметически закрытых камерах и поддерживались при давлении слегка ниже атмосферного давления. Они хранились в комнатах с постоянной температурой и защищались, насколько это возможно, от вибраций. Их ход поддерживался гирями, которые поднимались электрическим способом, гарантируя, таким образом, постоянство связи и, следовательно, постоянство амплитуды осцилляций маятника. Часто они не имели циферблата, выдавая только каждую секунду электрический импульс. Несмотря на все эти меры предосторожности нестабильность их частоты за период усреднения в один месяц могла достигать от 5×10^{-8} до 10^{-7}.

В парижской обсерватории «средние маятниковые часы» были собраны путем выбора лучших маятниковых часов, расположенных в подвалах обсерватории в глубоких катакомбах. Это отвечало нуждам службы времени и международному бюро времени. С 1950 года со средними маятниковыми часами стали объединяться и кварцевые часы (отметьте), и эти два типа часов использовались в тандеме до 1954 года, когда с последними маятниковыми часами утратили связь и трансформация была закончена.

5.4. Пьезоэлектрические осцилляторы

Кварцевые пьезоэлектрические осцилляторы были первыми генераторами и времени и частоты, использовавшими выгоду от прогресса, произошедшего в электронных технологиях. Их работа основывается на явлении пьезоэлектричества, открытым П.Кюри в 1830 году, и изобретением первого электронного усилителя, триода, Ли и Форестом в 1907 году. Такие осцилляторы использовались в системах связи и в метрологии времени и частоты с 1920-х годов. Сегодня высокоточные кварцевые осцилляторы используются как опорные генераторы частоты и хранители вре-

мени в огромном количестве в радионавигационных и телекоммуникационных системах. Кроме того, они всегда используются в атомных часах, где их частота контролируется по атомному резонансу и сигналы, используемые потребителями, происходят от этих осцилляторов.

В контексте возрастающих требований электронных систем и продолжающегося улучшения характеристик атомных часов много усилий было направлено на улучшение конструкций резонаторов и оптимизацию связи с электронными цепями. Много исследований было направлено на понимание причин происхождения кратковременной нестабильности, чувствительности к окружающим условиям и эффектам старения [5.15, 5.16].

5.4.1. Резонаторы

Резонатор это вибрирующая пластинка. Различные типы деформаций (например, изгиб, растяжение, сдвиг) могут распространяться через тело материала. Резонанс получается, когда акустические волны удовлетворяют граничным условиям. Его частота, таким образом, зависит от размеров пластинки, типа возбуждающих деформаций и свойств используемого материала.

Механический резонанс может поддерживаться электронными цепями через пьезоэлектрический эффект. Последний связывает электрические поля и механические деформации в кристаллах, имеющих центр симметрии, таких как кварц. Электрическое поле производится путем использования электродов, нанесенных на поверхности пластинок, или расположенных очень близко к ним, как показано схематически на рис. 5.4.

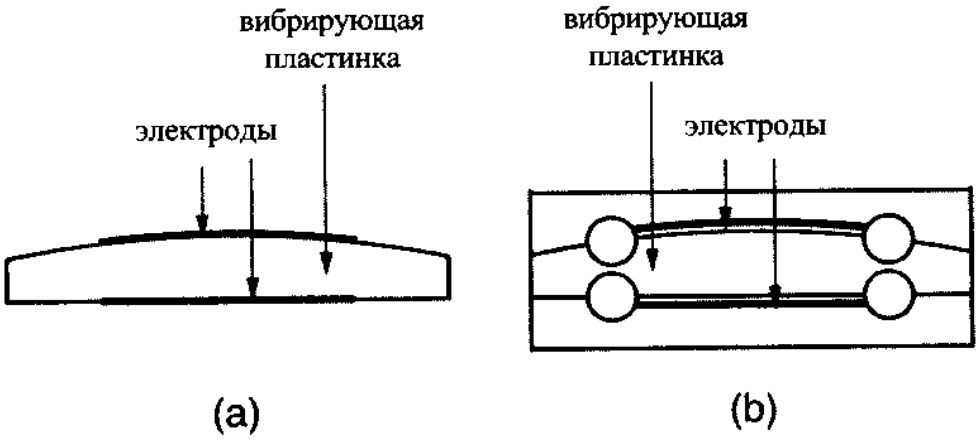


Рис. 5.4. Поперечное сечение кварцевого резонатора. (а) Электроды нанесены на вибрирующую пластинку. (в) Электроды не касаются вибрирующей пластинки. Соединения не показаны

Естественные кварцевые кристаллы не используются. Вместо этого кристаллы изготавливаются в больших количествах промышленным спо-

собом при относительно низкой цене, в автоклавах, где они выращиваются из зародышей. Полученный материал имеет исключительную химическую чистоту и его кристаллическая структура почти не имеет недостатков. Диэлектрические потери также очень малы.

На рис. 5.5а показана эквивалентная электрическая цепь пьезоэлектрического резонатора. Кварцевая пластинка, расположенная между электродами, действует как конденсатор с емкостью C_0. Индуктивность, емкость и сопротивление связанные с динамическим поведением пластинки, представлены через L_1, C_1, R_1 соответственно. Внешний конденсатор с емкостью C_e, соединенный с резонатором последовательно, используется для подстройки частоты.

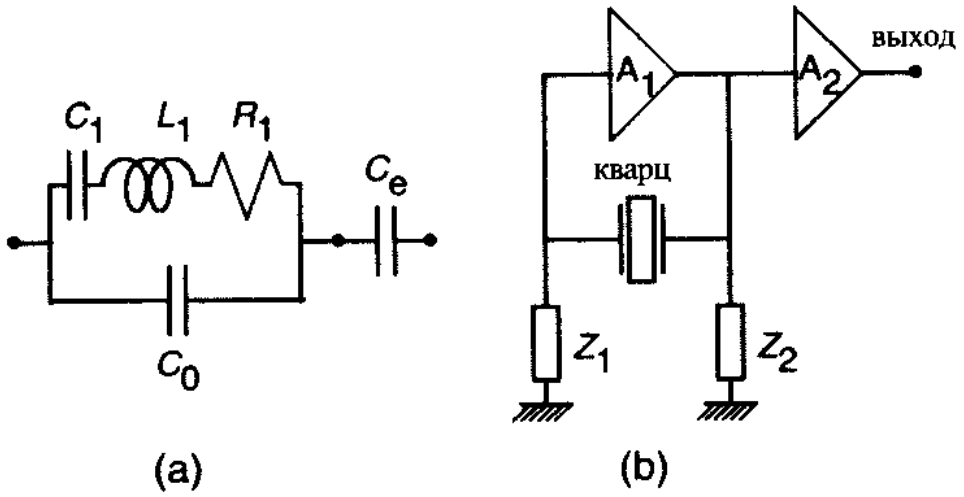


Рис. 5.5. Кварцевый осциллятор и резонатор. (а) Эквивалентная диаграмма для кварцевого резонатора. (в) Схема кварцевого генератора

Угловая частота ω_0 электрического резонанса определяется из следующего выражения:

$$L_1 C_1 \omega_0^2 = 1 \tag{5.40}$$

и добротность из:

$$Q = \frac{L_1 \omega_0}{R_1}. \tag{5.41}$$

Для прецизионных резонаторов в метрологических приложениях частота $\omega_0/2\pi$ обычно равна 5 или 10 МГц, а добротность составляет около 10^6.

Благодаря конденсатору C_0 в цепи на рисунке 5.5а эта цепь имеет два резонанса, последовательный и параллельный, с очень близкими значениями частот. В первом случае резонатор обладает низким импедансом, близким к R_1. Во втором случае импеданс высокий. Возбуждается тот или иной

резонанс в зависимости от того, подключен ли резонатор к цепи с низким или высоким импедансом Z_1 или Z_2, как показано на рисунке 5.5b.

Резонатор спроектирован так, чтобы свести к минимуму эффекты от возмущений, испытываемых вибрирующей частью кварцевой пластины. Ориентация пластины относительно кристаллографических осей выбирается таким образом, чтобы резонансная частота имела экстремум по отношению к температурным изменениям. Резонатор размещается в термостате, который поддерживает температуру при таком значении, при котором исчезают вариации частоты. Выбор ориентации пластины также происходит с учетом уменьшения вариации резонансной частоты, вызванного напряжениями, обусловленными или механическим креплением резонатора, или воздействием ускорения на осциллятор.

В большинстве современных кварцевых резонаторов электроды не напыляются на вибрирующие поверхности [5.17]. Они располагаются на вспомогательных пластинах, как показано на рис. 5.4b. Они размещаются в нескольких микронах от активной пластины. Любое затухание резонанса, вызываемое электродами, таким образом подавляется. Более того, вибрирующие пластины больше не страдают от эффектов старения, связанного с медленной релаксацией напряжений и перемещением материалов, происходящим в области между кварцем и металлом.

5.4.2. Осцилляторы

Резонатор помещается в цепь петли обратной связи между входом и выходом усилителя A_1, как показано схематически на рис.5.5b. Если в осцилляторе предпочтительнее работа с последовательным резонансом, то импедансы Z_1 и Z_2 должны иметь низкое сопротивление.

Общий фазовый сдвиг в цепи обратной связи всегда равен $2k\pi$, где k целое число. Следовательно, если имеются изменения фазы между входом и выходом усилителя, они компенсируются в кварцевом резонаторе. Такие изменения могут быть произведены шумами усилителя A_1, которые накладываются на электрические осцилляции. Если это возмущения на частоте, близкой к резонансной частоте f_0, то на зажимах любого резонатора они приводят к изменению фазы $d\phi$, которое дается выражением:

$$d\phi = -2Q\frac{df}{f_0}. \tag{5.42}$$

Таким образом, возмущение сопровождается изменением осциллирующей частоты. Однако соответствующее изменение уменьшается, если добротность увеличивается. Это объясняет, почему предпочтительнее использовать резонаторы с высокой добротностью при создании генераторов с низкой частотной нестабильностью.

Дополнительный конденсатор с емкостью C_e используется для тонкой настройки частоты генерации. Он обычно состоит из двух конденсаторов, один с фиксированной емкостью, и другой – варикап, соединенный с этой емкостью последовательно. В варикапе используется изменение емкости диода под действием запирающего напряжения. Частота генерации может таким образом управляться с помощью электрического напряжения. Эта особенность используется для автоматической подстройки частоты кварцевого генератора по частоте атомного резонанса (см. раздел 6.4.1c и 6.5.3d).

Электронные цепи, связанные с кварцевым генератором охвачены также системой термостабилизации. Усилитель A_2 увеличивает уровень сигнала и изолирует цепи генератора от возмущений, которые могут появляться в цепях усиления.

5.4.3. Характерные величины

Для применения в метрологии высокого уровня используются кварцевые генераторы с частотами 5 или 10 МГц или с близкими к ним частотами.

При времени выборки между от 1 до 10^3 секунд нестабильность частоты прецизионных кварцевых генераторов определяется частотным фликкерным шумом. Для лучших приборов достигается значение девиации Аллана лучше, чем 1×10^{-13}. На больших интервалах начинает проявляться дрейф частоты из-за эффектов старения. Часто он бывает меньше чем 10^{-11} за один день. Кратковременная нестабильность частоты порядка 1×10^{-14} на интервалах τ от 1 до 10^3 с и долговременный дрейф порядка 1×10^{-12} возможно будут достигнуты в ближайшем будущем.

5.5. Атомные стандарты времени и частоты. Представление о точности

В разделе 4.4.1, мы объясняли как атомные стандарты времени и частоты вошли в обиход, в то время как в главе 6 мы представим принципы их работы.

В настоящей части мы только дадим четкое определение идеи точности, которое уже использовалось в разделе 4.4.1. В атомных стандартах опорная частота ν_0, постулируется как неизменная по природе и обеспечивается переходом в невозмущенном атоме или ионе (или даже молекуле, но мы намерены использовать слово «атом» во всех случаях). Она является собственной частотой (в релятивистском смысле) в очень малом пространственном объеме, в принципе в бесконечно малом объеме, который связан с атомом в его движении. Набор аппаратуры, используемый для наблюдения этой атомной частоты, вносит свои возмущения в частоту стандарта, как это будет показано в главе 6. Этот набор также преобра-

зует частоту в другие частоты, с которыми удобнее работать. Выходная частота стандарта обозначается как ν_{out}. Это тоже собственная частота в малом пространственном объеме, который окружает, например, разъем, с которого снимается сигнал с частотой ν_{out}. Для того, чтобы гарантировать точность, соотношение

$$R = \frac{\nu_{out}}{\nu_0} \tag{5.43}$$

должно быть определено как можно более точно. Относительная неопределенность σ_E, характеризующая неточность стандарта, равна относительной неопределенности в оценке R.

Как и любая физически реализуемая частота, частота ν_{out} подвержена нестабильности. Первая мера предосторожности состоит в ее усреднении на таких интервалах, на которых обеспечивается минимальная нестабильность. Мы видели, что эти периоды соответствуют режиму с доминирующим фликкер-шумом. С этого момента мы примем, что ν_{out} будет такой усредняемой частотой.

Искусство физика состоит в установлении всех возможных источников возмущений. Следующая задача состоит в оценке каждого возмущения, часто с привлечением дополнительных измерений, и с определением неопределенности в этой оценке. Если выполнить это все для каждого возмущения i, то можно получить поправочный коэффициент r_i, который может быть представлен в форме:

$$r_i = (1 + c_i) \pm \sigma_i , \tag{5.44}$$

где σ_i есть соответствующая неопределенность. Тогда оценка $\hat{R}$ из R равна

$$\hat{R} = F r_1 r_2 \ldots r_n \tag{5.45}$$

для всего набора n возмущений. F это коэффициент преобразования частоты и предполагается, что он точно известен, как это обычно и бывает. Так как c_i много меньше, чем единица, мы можем принять следующее выражение:

$$\hat{R} = F(1 + c_1 + c_2 + \ldots + c_n). \tag{5.46}$$

Обычно предполагается, что различные возмущения взаимно независимы, так что неточность нормализованной частоты есть:

$$\sigma_E = \left(\sum \sigma_i^2\right)^{1/2}. \tag{5.47}$$

Для удобства, когда мы используем стандарты, поправки c_i могут быть введены физически на уровне выходной частоты. Это не оказывает влияния на неточность стандарта.

Понятие неточности не связано со спецификой атомных стандартов частоты, а больше является общим понятием, применимым к любой физической величине, определяемой по отношению к некоторым естественным явлениям, которые предполагаются неизменными.

5.6. Сличения времени и частоты

5.6.1. Общие наблюдения

Для всех физических величин сличения между стандартами являются одним из краеугольных камней в метрологии. Главная цель сличений состоит в проверке соответствия между первичными стандартами в различных лабораториях и затем во вкладах в направлении их дальнейшего совершенствования путем определения причин ошибок, которые в ином случае могут ускользнуть от внимания. Сличения могут также использоваться для создания среднего первичного стандарта, лучшего, чем каждый из его участников. Что же касается потребителей, то они обеспечиваются средствами доступа к первичным стандартам через организацию поверочных схем.

В области измерения атомного времени ключевой величиной для создателей стандартов является частота. Внутри лаборатории легко поверять различные стандарты частоты на постоянной основе и, следовательно, проверять что они работают правильно. Далее возможно осуществление эпизодических сличений частоты между разнесенными лабораториями, так же как это делается для других физических величин.

Однако применение стандартов частоты к измерениям времени значительно более сложная задача, как это было подчеркнуто в разделе 4.4.2. Требуются почти непрерывные сличения времени, чтобы создать и распространить однозначную шкалу времени, с помощью которой потребители могут синхронизовать свои часы и обеспечивать себя единицей времени – секундой. Мы также рассмотрим примеры систем, требующих внутренней синхронизации, настолько точной, насколько это возможно, и которая не может быть обеспечена без настройки за интервалы менее чем один день.

В следующих параграфах мы начнем с испытания методов сличения времени, а затем и методов сличения частоты. Мы будем касаться главным образом принципов сличения, потому что техника развивается очень быстро и сильно зависит от тех возможностей, которые она сама себе предоставляет. Последнее заявление требует некоторых объяснений. Конечно, некоторые системы специально предназначены для распространения времени и для сличений времени и частоты, например, радиопередача сигналов времени с начала двадцатого столетия на частотах, занимающих диапазон от нескольких килогерц до нескольких десятков

мегагерц. Такие передачи все еще существуют и их несущие частоты часто стабилизируются по атомным стандартам таким способом, чтобы они распространяли также отличную опорную частоту, которая легко может быть использована с помощью дешевого приемника. Но чтобы достичь лучшей стабильности, необходимо было бы установить более дорогостоящую рабочую систему, неотрывно связанную со сличениями часов. Вместо этого было принято решение использовать преимущества существующих систем, созданных для других целей, и все это без дополнительной стоимости. Такие системы включают в себя навигационные системы, системы определения положения и коммуникационные системы. Мы проверим направления, в которых методы сличения времени и частоты наиболее широко применяются при использовании этих систем.

Для того чтобы осуществить сличения лучших разнесенных стандартов частоты и лучших часов, обычно требуется релятивистская трактовка измерений. Не существует простых способов установления пределов точности с помощью классической трактовки, потому что эти пределы зависят от множества факторов. Тем не менее, мы можем сказать, что для стандартов и часов, установленных на уровне земли, для всех методов, описанных ниже, классическая трактовка адекватна при условии, что мы можем согласиться с неопределенностями на уровне несколько единиц 10^{-13} в нормализованной частоте и одной микросекундой во времени.

5.6.2. Соотношение между частотой и приращением собственного времени

Когда вопрос касается результатов, то соотношения между показаниями часов и их частотой часто приводят к умственной гимнастике потенциально опасной природы, даже если эти соотношения выполняются специалистами! Если τ_C представляет собственное время (показания) часов, выраженное в секундах, и τ есть собственное время идеальных часов, расположенных в том же месте, т.е. часов, у которых секундные метки разделены интервалом в одну секунду в строгом соответствии с определением в системе СИ, то из (5.6) следует:

$$\frac{\mathrm{d}x_C(\tau)}{\mathrm{d}\tau} = y_C(\tau). \tag{5.48}$$

Как объяснялось в разделе 5.2.2

$$\tau_C(\tau) = \tau + x_C(\tau) \tag{5.49}$$

и, следовательно, вводя нормализованную частоту $\Phi = 1 + y$, определенную в разделе 4.4.2, получим

$$\Phi_C = \frac{d\tau_C}{d\tau} . \tag{5.50}$$

Следовательно, для двух часов C и D

$$\frac{\Phi_C}{\Phi_D} = \frac{d\tau_C}{d\tau_D} . \tag{5.51}$$

5.6.3. Сличения времени

(a) Собственное время часов и инструментальные задержки

Часы являются сложным набором оборудования, и возникает вопрос, где в точности расположено собственно время. Фактически это будет шкала времени, созданная на импульсах, генерируемых каждую секунду, когда они появляются на определенном разъеме. Эти импульсы, имеют падающие и возрастающие участки напряжения, которые должны быть сделаны как можно короче, тем не менее импульсы имеют какое-то не пренебрежимое *время нарастания.* Поэтому необходимо установить высоту импульса.

Импульс, прибывающий откуда-нибудь, датируется через обращение к локальным измерениям, как объясняется в (b) ниже. Но прежде чем он придет к часам, он подвергается многим преобразованиям, которые приводят к *инструментальной задержке.* Это случается, например, в радиоприемниках. Так как всегда необходимо определять прибытие сигнала до преобразований, дата, обеспечиваемая часами, должна корректироваться на инструментальную задержку.

Уравнения, установленные в последующем для сличений времени, не будут учитывать инструментальные задержки. Их роль будет упомянута в приложениях. Мы должны подчеркнуть, что их трудно определить, и что глобальная неопределенность в сличениях часов обычно определяется неопределенностями этого происхождения.

(b) Локальные сличения времени

Давайте припомним из раздела 5.2.3a, что *измеритель интервалов времени*, в своей собственной шкале τ_{TIC} измеряет интервал времени $\Delta\tau_{TIC}$, прошедший между секундным импульсом часов C и первым, пришедшим после этого, секундным импульсом часов D. Чтобы интерпретировать это измерение можно условиться, что $\Phi_{TIC} = \Phi_{C.}$ Для этого можно для работы счетчика подать на него опорную частоту от C. Другой подход состоит в проведении вычислений, основанных на измерениях девиации частоты. Тогда мы имеем:

$$\Delta\tau_{\text{TIC}} = [\tau_{\text{C}} - \tau_{\text{D}}](\theta_2) \quad \text{по модулю 1 секунда,} \tag{5.52}$$

где θ_2 – дата в некоторой опорной шкале времени θ, когда измеритель останавливается секундным импульсом от часов D[4]. Затем добавляется целое число секунд. Современные приборы могут выполнять такие измерения с минимальной погрешностью около 10 пс (пикосекунд). В современных применениях мы, конечно, должны учитывать задержки в кабелях.

(c) Сличения разнесенных часов

Когда часы слишком далеки друг от друга чтобы их соединить кабелями, сличение времени выполняется одним из следующих трех методов или их комбинацией:

- физической транспортировкой работающих часов;
- односторонней передачей электромагнитных сигналов между излучателем и приемником;
- квази-одновременной передачей электромагнитных сигналов вдоль идентичных траекторий, соединяющих часы A и B, от A до B и от B до A (так называемый двусторонний метод передачи времени).

Уравнения в релятивистской форме опираются на договоренные определения сличений частоты, как это объясняется в разделе 3.3.2f. Мы видим, что они всегда делаются как «классическая часть», такая же, как она получилась бы в рамках абсолютного времени вместе небольшими дополнительными членами. Разделение происходит из формы метрики и зависит от принятой степени приближения. Оно удобно в том смысле, что так называемые релятивистские члены можно просто опустить в пределах допустимого уровня, представляя их пренебрежимо малыми. Однако в этом есть один существенный недостаток, так как разделение такого типа может привести нас к уверованиям, что общая теория относительности, это что-то такое, что надо только добавить к классической теории, тогда как на самом деле это совершенно новое толкование. Бесполезно и неблагоразумно различать априори классические эффекты, релятивистские эффекты и гравитационные эффекты. Это может привести к ошибкам и, в частности, к неоправданному добавлению «эффектов ускорения».

Релятивистское истолкование всегда обращается к тому, чтобы приращение в координатном времени (в каких-то обозначенных координатных системах) было бы определено или путем транспортирования часов, или

---

[4] Напомним, что выражение $[\tau_{\text{C}} - \tau_{\text{D}}](\theta)$ понимается как среднее показание часов C минус показания часов D на дату θ.

путем передачи сигналов. В этой ситуации используются геоцентрические координатные системы, или невращающиеся, или вращающиеся вместе с Землей. Если неопределенности времени составляют несколько сотен пикосекунд и допустимая неопределенность нормализованной частоты составляет 10^{-16} для земных часов или часов на орбите на расстояниях до геостационарных спутников (42 000 км от центра Земли), то применим анализ, основанный на [5.18], за несколькими исключениями, которые будут указаны. Более точный анализ можно найти в [5.19, 5.20]. Мы можем использовать или геоцентрическое координатное время t, или его модифицированную форму $\underline{t}$, определенную в разделе 3.3.2е. Формулы будут представлены в терминах t. Они остаются справедливыми при замене t на $\underline{t}$ и функции $h(t)$ на $\underline{h}(\underline{t})$, как это показано в (3.18), (3.21), (3.30) и (3.32). Если способ измерения требует, то можно пользоваться выражениями в терминах $\underline{t}$.

Согласно (3.17) и (5.51) приращение Δt в t между двумя событиями отмеченное показаниями $\tau_C(t_1)$ и $\tau_C(t_2)$ часов С, дается следующим выражением:

$$\Delta t = \frac{\tau_C(t_2) - \tau_C(t_1)}{1 + y_C} + \int_{W(C)} h(t)\mathrm{d}t, \tag{5.53}$$

где y_C относительная расстройка частоты по отношению к идеальному стандарту частоты и интеграл вычисляется вдоль мировой линии W(C) часов от t_1 до t_2.

Уравнение (5.53) может показаться до некоторой степени удивительным, так как мы должны бы иметь $\Delta t = t_2 - t_1$. Для того чтобы понять употребление этого уравнения мы должны оценить роль t следующим образом. Неопределенность в Δt дает непосредственный вклад в сличения времени. В показаниях, таких как $\tau_C(t_2)$, это есть подтвержденная точность в τ_C, в то время как t_2 служит только для того, чтобы пометить событие «второе показание часов С», которое мы должны также описать в этих терминах. В интеграле мы определенно нуждаемся в t и в пределах t_1, t_2, но однако достаточно физической реализации t при условии, что расстройка частоты от t не слишком велика. Например, если эта расстройка составляет 10^{-9}, и так как $h(t)$ порядка 10^{-9}, то ошибка, вводимая в Δt, будет меньше чем $10^{-18}\Delta t$, что совершенно пренебрежимо для современных измерений. Если мы предположим, что опорная шкала θ, уже использовавшаяся ранее, удовлетворяет последнему условию, то t может быть заменено на θ везде в правой стороне уравнения (5.53). Шкала может быть, например, всемирным координированным временем (UTC), легко доступным в пределах нескольких десятых секунд (с учетом возможной скачущей секунды) или в пределах 10^{-9} по отношению к норма-

лизованной частоте. Однако в дальнейшем мы будем придерживаться выражений в терминах t, оставляя на потребителя работу по определению степени приближения к t, которая была бы приемлема для него, что никогда не составит ему труда.

(1) Сличения времени с помощью перевозимых часов. Пусть A и B часы для сличений. Здесь метод вовлекает перенос часов M из ближайшей окрестности A в окрестность B, где осуществляются локальные сличения $[\tau_A - \tau_M](t_1)$ и $[\tau_B - \tau_M](t_2)$. Такое толкование легко понять в модели абсолютного времени. Часы M служат для экстраполяции показания часов A. Это их применение требует предварительной стадии, в которой должна быть оценена разность нормализованных частот $y_A - y_M$. Требуется также круговая поверка: M устанавливается вблизи A и подтверждается, что наблюденное время и разность частот с A совместимы с тем, что ожидалось в пределах оцениваемых неопределенностей. Если это действительно так, то наблюденная *замыкающая ошибка* также должна быть принята во внимание. Эти действия определяются скорее здравым смыслом экспериментатора, а не статистическим истолкованием.

Релятивистское истолкование явилось одним из первых практических применений общей теории относительности к обычно выполняемым измерениям. Из (5.53), при $\Delta t = t_2 - t_1$, следует что

$$\tau_M(t_2) - \tau_M(t_1) = (1 + y_M)[\Delta t - \int_{W(M)} h(t)\mathrm{d}t]. \qquad (5.54)$$

Те же уравнения используются для получения $\tau_A(t_2) - \tau_A(t_1)$. Далее, извлекая одно из другого и переходя к измеряемым величинам, получим:

$$\begin{aligned}[\tau_B - \tau_A](t_2) = {} & [\tau_B - \tau_M](t_2) - [\tau_A - \tau_M](t_1) - (y_A - y_M)\Delta t \\ & - (1 + y_M)\int_{W(M)} h(t)\mathrm{d}t + (1 + y_A)\int_{W(A)} h(t)\mathrm{d}t\end{aligned} \qquad (5.55)$$

Первые три члена с правой стороны представляют классическое выражение. Релятивистские эффекты представлены остальным членами, в которых интеграл вычисляется вдоль мировых линий W(M) и W(A) часов M и A между датами t_1 и t_2. С пренебрежимой неточностью по сравнению с неопределенностями измерения мы обычно можем принять $y_M = 0 = y_A$ в коэффициентах умножения при этих интегралах.

Впервые предпринятая в 1967 г. инициатива Хьюлетт – Паккард и Смитсонианской Астрономической обсерваторией по сличению времени с помощью перевозимых часов стала обычной практикой, примерно до 1988 года. Она применялась для синхронизации пунктов контроля навигационных систем, таких как Лоран-С, станций слежения за искусственными спутниками Земли и национальных лабораторий по хранению времени. Для того времени это был самый правильный путь, гарантирующий точность сличения шкал времени, в отличие от всех других

методов, включающих в себя распространение и инструментальные задержки, которые нельзя было оценить непосредственно без внесения дополнительных порций неопределенностей.

Обычно промышленно изготовленные часы транспортировались по дорогам или на самолетах. Их размер и вес (около 40 кГ) не препятствовал тому, что бы они могли располагаться на пассажирском месте. Если емкости их собственных источников питания было недостаточно, они, по соглашению с авиакомпаниями, могли подключаться к бортовой сети самолетов. Например, в 1978 году в ежегодном докладе международного бюро времени перечислено сорок три передачи времени таким способом, которые охватывали все континенты и имели неопределенности от 10 до 300 наносекунд. Транспортировка часов с 1986 года стала производиться реже, так как все больше и больше сличения времени стали выполняться с помощью Глобальной Системы Позиционирования (GPS). Более того, примерно в то же время стали появляться проблемы с правилами безопасности на воздушном транспорте, так как было осознано, что цезий склонен к спонтанному возгоранию под воздействием воздуха. Попытки перевозить рубидиевые часы вскоре также были прекращены, так как часы не обладали стабильностью в таких условиях.

Интересно отметить, что такой способ передачи времени породил массу анекдотов и несколько несчастных случаев. Мы можем упомянуть и авиабилеты, выданные на мистера Атомные Часы, и общее удивление авиаперсонала, и сложности, вызванные употреблением слова «атомные», которое быстро было заменено на «электронные», или, бесконечная путаница и недопонимание таможенников.

Для того чтобы учесть релятивистские эффекты, должна была быть известной траектория полета самолета. Такие эффекты при межконтинентальных перелетах могли достигать несколько десятков наносекунд. Например обнаружилось, что для перелета Париж – Оттава из результатов классических вычислений надо отнять 32 нс, в то время как при обратном перелете и в предположении той же самой траектории, надо было отнимать только 7 наносекунд. Это можно было использовать для доказательства того, что Земля действительно вращается.

(2) Сличения времени с помощью односторонней передачи электромагнитных сигналов

Принципы и теория. Этот метод начал использоваться с начала двадцатого столетия (см. главу 4) и основан на теории распространения сигналов. Он часто выбирается для передачи времени с низким уровнем точности (сигналы времени для навигационных целей, говорящие часы и т.д.). Он также удовлетворяет строгим требованиям по точности, таким, которые мы здесь рассматриваем, при условии работы в пределах прямой видимо-

сти радиопередатчиков. Для больших расстояний это означает необходимость применения антенн космического базирования.

Мы начнем с теории, оставив на потом основы работы по сличению двух часов в пределах прямой видимости, потом мы рассмотрим некоторые применения, в частности такие, которые делают возможным использование глобальной системы позиционирования (GPS)

Электромагнитные сигналы в вакууме в инерциальных системах распространяются по направлениям прямых линий (т.е. в системах с равномерным прямолинейным перемещением без вращения). Если мы возьмем невращающуюся геоцентрическую систему, то определение в главе 3 устанавливает, что такая система является совершенно достаточным приближением для сличений времени. Если фотоны распространяются от события E – излучение сигнала с координатами $(t_E, \mathbf{x}_E)$ до события R – прием сигнала с координатами $(t_R, \mathbf{x}_R)$, то приращение Δt в t во время распространения сигнала составляет величину порядка c^{-3}, которая удовлетворяет

$$\Delta t = \frac{d}{c} + (\Delta t)_U , \tag{5.56}$$

где d координатное расстояние между E и R, вычисляемое из

$$d = \left[(x_R^1 - x_E^1)^2 + (x_R^2 - x_E^2)^2 + (x_R^3 - x_E^3)^2\right]^{1/2}, \tag{5.57}$$

которое мы приведем к аббревиатуре $|\mathbf{x}_R - \mathbf{x}_E|$, хотя $\mathbf{x}_R - \mathbf{x}_E$ не имеет свойств вектора в пространстве-времени. Член $(\Delta t)_U$ находится путем интегрирования вдоль направления распространения, принимая что $U = GM_E/r$, с $r = |\mathbf{x}|$, и имея ввиду, что $GM_E = 3{,}986 \times 10^{14}$ м^{3}с$^{-2}$. мы получим

$$(\Delta t)_U = \frac{2GM_E}{c^3} \ln \frac{r_E + r_R + d}{r_E + r_R - d} . \tag{5.58}$$

Этот член мал. В случае передачи от геостационарного спутника до наземных часов, расположенных на одной с ним вертикали, величина эффекта составляет 56 пикосекунд.

При использовании геоцентрической системы, вращающейся вместе с Землей и, при рассмотрении неподвижных приемных станции на Земле с координатами $\mathbf{X}_R$, для Δt получим из

$$\Delta t = \frac{|\mathbf{X}_R - \mathbf{X}_E|}{c} + \frac{2\omega A_E}{c^2} + (\Delta t)_U , \tag{5.59}$$

где $\mathbf{X}_E$ оценивается на дату t_E и $(\Delta t)_U$ и все еще дается формулой (5.58). Член $2\omega A_E/c^2$ описывает эффект Саньяка. Площадь A_E, определенная сра-

зу после (3.21), здесь относится к траекториям фотонов. С достаточно хорошей степенью приближения это площадь экваториальной проекции треугольника с вершинами в центре масс Земли и в точках с координатами $\mathbf{X}_E$ и $\mathbf{X}_R$. Она считается положительной, если распространение направлено на восток. Часто эффект Саньяка достигает значительной величины. Например, при распространении с точки на экваторе до геостационарного спутника с восхождением 45° восточной долготы, его величина составляет 137 нс. Далее к Δt надо добавить задержку из-за рефракции. Она оценивается независимым способом.

Предположим, что при излучении сигнала мы отметили показания $\tau_S(t_E)$ локальных часов S. Показания можно передать экспериментатору обычными средствами. (В качестве альтернативы мы можем организовать сличения таким образом, чтобы излучение было прикреплено к таким показаниям часов S, для которых известна номинальная дата сигнала). Проделаем ту же самую процедуру с локальными часами A, при приеме сигнала снимая показания $\tau_A(t_{RA})$. Используя (5.53) мы получим разницу времени между часами S и A на дату t_E из соотношения

$$[\tau_A - \tau_S](t_E) = \tau_A(t_{RA}) - \tau_S(t_E) -$$

$$-(1+y_A)(\Delta t)_A + (1+y_A)\int_{W(A)} h(t)\mathrm{d}t, \qquad (5.60)$$

где $(\Delta t)_A$ дается выражением (5.56) или(5.59) и интеграл вычисляется вдоль мировой линии W(A) часов A между датами t_E и t_{RA}. Как обычно величина t заменяется на показания реализованной шкалы времени[5].

Когда двое часов A и B должны сличаться, но они не находятся в прямой видимости, мы можем использовать прием одного и того же сигнала с антенны, которая находится в пределах прямой видимости для обоих часов. Переписывая аналогичные уравнения в (5.60) для часов B и затем вычитая (5.60) мы получим:

---

[5] Как исключение, проверим, что замена $t \to \underline{t}$ не изменяет величину в левой стороне. Из (3.27)

$$\Delta t = \Delta \underline{t}\left(1 + \frac{\hat{U}_g}{c^2}\right) \qquad (5.61)$$

и из (3.31)

$$\underline{h}(t) = \underline{h}(\underline{t}) + \frac{\hat{U}}{c^2}. \qquad (5.62)$$

Член $\Delta \underline{t}\hat{U}_{g/c^2}$ происходящий из интеграла действительно компенсирует изменение в оценке t.

$$[\tau_A - \tau_B](t_E) = \tau_A(t_{RA}) - \tau_B(t_{RB}) - (1 + y_A)(\Delta t)_A + (1 + y_B)(\Delta t)_B$$
$$+ (1 + y_A)\int_{W(A)} h(t)\mathrm{d}t - (1 + y_B)\int_{W(B)} h(t)\mathrm{d}t \qquad . \quad (5.63)$$

Среди четырех последних членов с правой стороны в (5.63) часто необходимо только сохранять $(\Delta t)_B - (\Delta t)_A$. Например, для антенны, установленной на геостационарном спутнике, Δt меньше, чем 0,14 с и если мы имеем $|y|<10^{-8}$, то $|y\ \Delta\ t| < 1{,}4$ нс. Так как $h(t)<7\times10^{-10}$, интегралы никогда не превышают 100 пс, то значением y в коэффициентах умножения можно пренебречь. Тем не менее полезно проверять каждый случай отдельно, чтобы убедиться, что не пропущены важные составляющие.

Отметим, что факт приема сигналов не означает, что часы в S не могут обойтись без него. Этот сигнал является совершенно произвольным. Однако присутствие стабильных часов в S позволяет избежать выполнения требования строго общей видимости антенны. В этом случае сличения времени между A и S, и B и S, выполненные в разные даты, могут затем соотноситься с датами прямой видимости через частотные оценки. Заметим также, что нам надо знать только задержки на приемниках из-за рефракции.

Приложение к Лоран-С. Сличения времени через односторонний прием широко применялись с 1967 года при приеме сигналов от системы Лоран-С. В 1980-е годы система постепенно перестала использоваться для целей метрологии из-за предпочтения к GPS. Определение положения на основе приема сигналов Лоран-С основано на синхронизации (цезиевых) часов на наземных передающих станциях, работающих на частоте 100 кГц. Желающий определить свое положение измеряет разницу во времени приема сигналов, вызванную задержками распространения. Каждая пара станций обеспечивает информацию о своем положении на уровне моря. Искомое положение находится при приеме сигналов от трех станций. Конечно, навигационные приемники выполняют при этом все вычисления. Точность такой системы определяется стабильностью задержек распространения *земных радиоволн*, которые могут приниматься на расстояниях до 1000 – 2000 км. Для неподвижного наблюдателя вариации таких задержек могут составлять около 1 мкс. Фактическая задержка не может быть определена с тем же уровнем точности. В применении к сличениям времени она определялась с помощью транспортируемых часов. В измерениях определялись и учитывались задержки в приемниках. Для сличений времени производились специализированные промышленные приемники. Они дали огромный вклад в решение поставленной задачи благодаря своей высокой надежности и низкой стоимости. Было опреде-

лено, что общая неопределенность сличений времени через Лоран-С лежит в окрестности 1 мкс. Синхронизация различных сетей Лоран-С в северной Атлантике в конце 1960-х годов позволила установить линии непрерывной передачи времени между Северной Америкой и Европой. И это сыграло ключевую роль в развитии измерений времени.

Применение в коммерческом телевидении. При приеме коммерческого телевидения, сигналы, содержащиеся в тестовых строках, или используемые для синхронизации изображения, также широко использовались для сличения вместе с Лораном-С [5.21]. Вблизи, в пределах видимости одной и той же антенны, неопределенности сличения времени могли быть уменьшены до десятков наносекунд после калибровки передачи и инструментальных задержек путем транспортировки часов. Были установлены линии связи протяженностью в несколько сотен километров при использовании радиорелейных линий и направленных радиоволн. Накопленные ошибки достигали нескольких сотен наносекунд, в то же время возникало много проблем с заменяемыми сегментами оборудования и с трассой распространения сигнала.

Применение в GPS и ГЛОНАСС. Вернемся теперь к системе GPS, США, которая широко используется для сличения разнесенных часов. Существует русский аналог называемый *Глобальная Навигационная Спутниковая Система (ГЛОНАСС).* Эта система дает сравнимые по точности результаты и также широко используется. Так как две системы отличаются незначительными техническими особенностями, мы ограничимся объяснениями, применимыми к GPS.

Как и Лоран-С эти системы основаны на синхронизации излучаемых сигналов, и основное отличие заключается в том, что излучающие антенны теперь расположены в космическом пространстве. В главе 9 мы будем объяснять, как они работают. Для наших целей на этой стадии достаточно рассмотреть спутники GPS как передатчики импульсов времени. Кроме этих импульсов спутники передают также данные, позволяющие определить их положение в земной координатной системе на момент излучения импульса времени. Сигналы излучаются на частотах, близких к 1,6 и 1,2 ГГц. Ионосферная рефракция приводит к задержкам, которые варьируются в широких пределах днем и ночью, или во времена солнечной активности. Задержки могут достигать сотен наносекунд при нормальных условиях работы. Они являются простой функцией от частоты и поэтому могут быть определены с точностью до единиц наносекунд через разницу в задержках сигналов на каждой из этих частот. Если же принимается только одна частота, то можно использовать модель ионосферной рефракции, которая также передается спутниками и которая позволяет получить точности до нескольких десятков наносекунд.

На спутниках установлены цезиевые или рубидиевые часы, синхронизированные по GPS времени в пределах нескольких наносекунд. Хотя это и несущественно для работы GPS систем, тем не менее время GPS поддерживается в синхронизме с TAI + 19 секунд в пределах сотен наносекунд или около этого. Существуют коммерческие приемники, специализированные для измерений времени. Приемники первого поколения работают на одной частоте. Они программируемы и успешно отслеживают заданные спутники за времена до 13 минут (почему, мы увидим в главе 9). В дальнейшем, в 1999 году в использовании появились многоканальные приемники, которые позволяли следить сразу за всеми видимыми спутниками одновременно. Все эти приемники производят вычисления в реальном масштабе времени и выдают данные по разнице показаний локальных часов и часов GPS.

Давайте рассмотрим метод, в котором используется прием одного и того же импульса часами, которые необходимо сличить. Этот метод часто называется *методом общего видения.* После обмена данными, сличения часов получаются простым вычитанием. Неопределенности на любых расстояниях могут быть уменьшены до нескольких наносекунд путем усреднения более чем за один день. Однако, чтобы иметь такие точности, необходимо выполнить ряд предосторожностей, которые существенно усложняют задачу.

Координаты приемных антенн должны быть известны с точностью до нескольких десятков сантиметров в той же координатной системе, что и координатная система спутников (за одну наносекунду сигнал проходит расстояние в 30 см). Координатной системой спутников GPS является *Мировая Геодезическая Система (WGS)*, версия WGS 84, которая согласуется, в пределах нескольких дециметров, с международной опорной координатной системой ITRF, которая в свою очередь определяется и поддерживается международной службой вращения Земли (IERS) (см.3.3.2b). Эта система принята для очень точных работ.

Задержки, возникающие в приемниках, измеряются производителями с помощью имитаторов, которые имитируют излучение сигналов спутников. Предполагается, что эти задержки неизменны, но безопаснее было бы их проверять. Для сличений времени было бы достаточно прокалибровать разницу в задержках между приемниками. Это можно сделать путем транспортировки приемника, используемого как передаточный стандарт. Этот метод используется без особой уверенности, так как могут доминировать неопределенности, вызванные чувствительностью приемника к окружающим условиям. Эти неопределенности составляют несколько наносекунд.

Поправки на ионосферу должны измеряться либо двухчастотным приемником для измерений времени, или применением специально разработанной техники.

Одна из проблем, которая формально исчезла в мае 2000 года, состояла в использовании так называемого *Селективного Доступа (SA* см. раздел 9.2.2b). Этот метод лишал неуполномоченных потребителей той потенциальной точности, которую они могли бы получить из GPS. Для сличений времени деградация сигнала, вызванная применением режима селективного доступа, могла быть преодолена строгим соблюдением графика приема [5.22] и использованием точного вычисления эфемерид в ретроспективе. Это оказывало давление на строгость графиков приема, требования к которым теперь ослаблены, но точные эфемериды все еще необходимы для наивысшей точности. Они устанавливаются научной службой, называемой международная служба GPS (IGS), главным образом для того чтобы удовлетворить нужды геодезии и мониторинга вращения Земли.

Разработка многоканальных приемников улучшает точность сличения времени за счет увеличения объема данных. Однако ограничивающими факторами являются слабые места в приемных системах, такие как чувствительность к температуре (наблюдаются вариации в пределах нескольких наносекунд) и возмущения, вызываемые отражениями от объектов вблизи приемной антенны. Были предприняты попытки преодоления этих препятствий и повышения точности не простым приемом кодированных сигналов а путем слежения за фазой несущей, так как это делается в геодезических применениях GPS.

Русская система глобальной навигации ГЛОНАСС работает по тем же принципам, что и GPS. То, что она широко не используется для сличений времени, может быть объяснено ее более поздним появлением и отсутствием подходящих приемников.

Успех GPS предполагает запуск аналогичных гражданских систем (глава 9). Сличение и распространение времени от этого могут только выиграть.

Применение к квазарам. Другой тип источников является также предметом применения метода общего видения (common view) хотя эти источники времени очень удалены. Речь идет о квазарах. Интерферометры со сверхдлинной базой (РСДБ) являются действительно выдающимся воплощением применения атомных часов (см. раздел 8.1.4a и главу 9). Их использование в геодезии и изучении движения Земли основано на измерении разности времени приема случайных сигналов от квазаров в местах, удаленных друг от друга от нескольких сотен до нескольких тысяч километров. Расположенные в этих местах часы не могут быть синхронизированы с требуемой точностью (неопределенности около 10 пс). Разница показаний этих часов является одним из многих параметров, которые крупномасштабно подстраиваются в процессе измерений. Однако подстройка не может отличить разницу в показаниях часов от инструментальных задержек и ошибок в оценке атмосферной рефракции. Ка-

либровка задержек с использованием интерферометров на нулевых базисах может дать точность сличений времени с неопределенностями вплоть до сотни пикосекунд. К сожалению станции РСДБ часто расположены очень далеко от лабораторий времени.

(3) Сличения времени при помощи двусторонней передачи электромагнитных сигналов. Рассмотрим двое часов A и B. Передача сигналов от A к B и от B к A выполняется почти одновременно с незначительными отличиями в траекториях распространения. Мы определим:

t_{EA} – момент излучения сигнала от A к B,

t_{RB} – момент приема сигнала в B,

t_{EB} – момент излучения сигнала из B в A,

t_{RA} – момент приема сигнала в A,

и

$$t_{av} = \frac{1}{2}(t_{EA} + t_{EB}).$$

Используя (5.60) и предполагая, что можно удовлетвориться линейной экстраполяцией $\tau_B - \tau_A$ между t_{EA} и t_{EB} путем усреднения, мы получим

$$\begin{aligned}[\tau_B - \tau_A](t_{av}) &= \frac{1}{2}[\tau_B(t_{EB}) + \tau_B(t_{RB})] - \frac{1}{2}[\tau_A(t_{EA}) + \tau_A(t_{RA})] \\ &\frac{1}{2}[(\Delta t)_{AB} - (\Delta t)_{BA}] + \frac{1}{2}\varepsilon,\end{aligned} \tag{5.64}$$

где

$$\begin{aligned}\varepsilon = {} & y_A(\Delta t)_{BA} - y_B(\Delta t)_{AB} + \\ & + (1 + y_B)\int_{W(B)} h(t)\mathrm{d}t - (1 + y_A)\int_{W(A)} h(t)\mathrm{d}t,\end{aligned} \tag{5.65}$$

интеграл берется вдоль мировой линии часов во время распространения сигналов. Член ε обычно очень мал.

Приращение координатного времени $(\Delta t)_{AB}$ на траектории от A к B и $(\Delta t)_{BA}$ на траектории от B к A вычисляются с использованием (5.56) или (5.59). В формуле используются только их разности. Следовательно, если мы способны оценить эту разность, то метод освобождается от проблем получения точных координат часов, траекторий распространения и атмосферных задержек.

Метод, обозначенный как TWSTFT, обычно применяется для сличений наземных часов, использующих сигналы, переизлучаемые геостационарными спутниками. В таком случае удобно пользоваться координатной системой, вращающейся вместе с Землей. При этом достаточно легко учесть остаточное движение спутника по отношению к Земле [5.19]. В

(5.64) вклад от члена $(\Delta t)_U$, определяемого из формулы (5.68) много меньше чем 1 пс. Члены в (5.65), содержащие интегралы, стремятся уравновесить друг друга так, что их вклад меньше 1 пс при условии, что $|y| < 10^{-2}$. Может оказаться необходимым оценить члены $y\Delta t$. Здесь важен эффект Саньяка, так как он меняет знак вместе с изменением направления распространения. Например, если А находится в Париже, а В в Токио и если геостационарный спутник висит на 67,5° восточной долготы, то $(\Delta t)_{AB} = -(\Delta t)_{BA} = 299$ нс.

Если мы пользуемся этим методом, то мы должны определить суммарные инструментальные задержки при приеме и излучении сигналов на каждой станции. На одной и той же станции достаточно просто измерить эти разности в задержках путем локального эксперимента. Если они остаются постоянными, мы можем предположить, что каждая из этих двух задержек постоянна. Однако сами измерения этих задержек являются достаточно деликатным предметом. Как последнее средство, метод двусторонней передачи может быть повсеместно прокалиброван с помощью другого метода сличений времени. На результаты сличений могут также влиять различные задержки из-за спутниковых радиосистем. Дальнейшие проблемы происходят, если для линии связи времени используются несколько антенн на спутнике.

Иногда встречаются административные возражения, когда необходимо спрашивать разрешение на излучение сигналов. Организации, которым принадлежат спутники, могут потребовать лицензионную плату.

Эти проблемы, а также большое удобство пользования системой GPS, к несчастью замедлили развитие метода передачи времени с помощью двухсторонних передач несмотря на его перспективы в улучшении точности. Тем не менее, после первых испытаний этого метода в 1962 г., упомянутых в разделе 4.4.2 [4.22], последовало еще множество экспериментов. Мы можем сослаться на первые работающие линии передачи времени между Францией, Канадой и, затем, Германией с 1978 по 1982 гг благодаря экспериментальному телекоммуникационному спутнику Симфония [5.23]. Общая неопределенность в этих передачах составляла около нескольких сотен наносекунд и была меньше, чем неопределенность, которая получалась при использовании Лоран-С. Поэтому международное бюро времени использовало эти передачи, чтобы построить TAI.

Последние несколько лет рабочая группа в международном бюро мер и весов (BIPM) разрабатывала критерии и подходы к определению параметров системы связи для времени, работающей по методу двухсторонних передач. Эти критерии и подходы включают в себя выбор спутников и частотных полос, определение особенностей наземных мониторинговых станций, методов калибровки, графиков измерений, обмена и обработки данных. Несколько линий связи работает на регулярной основе.

Некоторые из них используются BIPM для вычисления TAI. Нестабильность сличений времени при усреднении в один день составляет 100 пс, но неточность, которая зависит от калибровки задержек, лежит в диапазоне от 1 до 10 нс. Такая, какая она есть, TWSTFT не имеет себе равных для сличения частот на больших расстояниях (неопределенность 10^{-15} за один день) и является отличным инструментом для изучения первичных стандартов частоты. А что касается сличений времени, то вопрос о конкуренции с GPS все еще открыт.

Существуют другие применения метода двухсторонних передач, в которых не используются геостационарные спутники. На близких расстояниях метод может быть использован с применением лазерных сигналов, передающихся по воздуху или через оптические волокна. На средних расстояниях русские ученые использовали отражения сигналов от метеоритных следов на частоте 57 МГц. Диапазон мог достигать около 1000 км и неопределенности от 20 до 30 нс после калибровки инструментов с помощью перевозимых часов [5.24].

(4) Комбинированный метод сличения времени: эксперимент ЛАССО. Эксперимент ЛАССО (лазерная синхронизация со стационарных орбит) является примером смешанного метода, направленного на повышение точности. Он использует способности лазерных дальномеров определять расстояния до объектов когда стреляют по искусственным спутникам или по Луне в целях изучения глобальной геодинамики, океанографии, астрономии и т.д.

В проведенном эксперименте геостационарный спутник S имел часы, обозначенные также S. Спутник был оборудован уголковыми отражателями и детектором, который мог датировать прибытие лазерного импульса на S.

Лазерные линии связи между часами A на некоторых станциях и часами S могли сравниваться с классическими линиями двухсторонней передачи. Если не брать в расчет различные инструментальные задержки, результаты сличений между A и S могут быть получены из формул (5.64) и (5.65), где B заменяется на S, а время излучения и приема в S одно и то же. Точно так же осуществляются сличения B и S. Сличение времени между A и B включает «транспортировку» часов S между датой прибытия сигналов, излучаемых A и B. При хорошей синхронизации лазерных выстрелов продолжительность этой транспортировки может быть такой малой, как это необходимо, и поэтому главной характеристикой, требуемой от S, является кратковременная стабильность. Кварцевые часы являются вполне подходящими на эту роль. Даты прибытия сигналов передаются на координирующие пункты.

Такой эксперимент был осуществлен европейским космическим агентством. Первоначально он планировался на спутнике Sirio 2, который

был потерян во время неудачного запуска в 1982 г. Эксперимент был предпринят снова в 1988 г. со спутником Метеосат P2. К этому время рубиновые лазеры были заменены на Nd-YAG-лазеры с улучшенными возможностями определения расстояний. Хотя эти лазеры имели в качестве преимущества более короткие импульсы, они передавали меньше энергии и многие станции потеряли все шансы приема отраженного сигнала от геостационарных спутников оборудованных отражателями, спроектированными для LASSO. В дополнение для стабилизации спутников использовалось вращение (период 0,6 с) и поэтому лазеры должны были стрелять в узкие по времени (когда видны уголковые отражатели) окна (33 мс), что сильно усложняло устройства измерения

Как и во всех сличениях времени, определение инструментальной задержки является существенной и трудной задачей. Задержки на S уничтожаются. Чтобы найти задержки на наземных станциях была построена портативная лазерная станция. Этот калибровочный инструмент располагался рядом и стрелял в направлении той же локальной мишени (на расстоянии примерно километр) что и калибруемый прибор. Каждый прибор был способен принимать обратное отражение от другого лазера в дополнение к отражению собственного сигнала. Путем пересечения выстрелов всеми возможными способами можно было получить достаточное число уравнений для определения всех имеющихся задержек.

На установленной LASSO линии между лунной лазерной локационной станцией во Франции и Мак Дональдом (Техас) в 1992-1993 г. было продемонстрировано, что неопределенности передачи времени без учета калибровочных поправок составляют меньше 100 пс. Калибровочные поправки были выполнены с неопределенностью 1,5 нс, которая и оказалась доминирующей [5.25].

Метод очень сложный и сильно зависит от условий облачности. Однако технические усовершенствования по сравнению с его первоначальным проектом открывают некоторые перспективы. Он может служить как калибровочный способ для исследований или для передачи данных на землю от высококачественных атомных часов на орбите. Требуемое оборудование, легкое и компактное, может быть легко установлено на геостационарных или на низкоорбитальных спутниках, уже оборудованных часами и дистанционными приборами.

5.6.4. Сличение частоты

Давайте припомним, что все измерения частоты являются усредненными измерениями, несмотря на то что слово «усредненный» часто опускается. Частота, связанная с экспериментаторами, является *собственной частотой*, т.к. она относится к локальной секунде. Величины Ф и y должны пониматься в этом смысле.

Когда сличаются два разнесенных стандарта частоты, то интуитивно ясно, что величина, которая должна быть определена, это разность между их нормализованными частотами в предположении, что эти стандарты расположены рядом. К сожалению, эта интуитивная идея не соответствует точным определениям. И все это потому, что результат может зависеть от траектории движения стандартов. То же самое можно сказать для любой физической величины. По соглашению частоты сличаются в тех же самых интервалах координатного времени в определенной координатной системе, в нашем случае геоцентрической системе.

(а) Сличения частоты основанные на сличениях времени

(1) Прямые сличения стандартов. Рассмотрим двое часов А и В, нормализованные собственные частоты которых должны сличаться. Используя один из методов, описанных в разделе 5.6.3, мы измеряем разность собственных времен $[\tau_B - \tau_A](t_1)$ и $[\tau_B - \tau_A](t_2)$ для дат t_1 и t_2. В принципе даты измерений выражаются в координатном времени t, но в грубом приближении здесь достаточно θ. Мы вычислим

$$M_B - M_A = \frac{[\tau_B - \tau_A](t_2) - [\tau_B - \tau_A](t_1)}{t_2 - t_1}. \tag{5.66}$$

Затем мы получим разность нормализованных собственных частот, усредненных за (t_1, t_2) из

$$y_B - y_A = M_B - M_A + \langle h(t, \mathbf{x}_B) - h(t, \mathbf{x}_A) \rangle, \tag{5.67}$$

где символ $\langle\ \rangle$ означает усреднение по времени. Функция h дается (3.18) или (3.21). Она зависит только от t вдоль мировых линий $\boldsymbol{x}(t)$ часов (и t может быть заменено на θ для всех практических целей). Ясно, что в локальных экспериментах два члена, обозначенные в h уничтожаются.

В применении к наземным часам уравнения дают

$$\langle y_B - y_A \rangle = M_B - M_A + 1{,}091 \times 10^{-13}(h_B - h_A)/км, \tag{5.68}$$

где h_A и h_B высоты в километрах над геоидом (уровень моря). Неточность в (5.68) в этом случае меньше чем 10^{-15}.

Когда сличения частоты получаются через сличения времени, то они не зависят от инструментальных задержек при условии, что последние остаются постоянными, и это иногда необходимо проверять.

Хотя неопределенности из-за калибровок могут быть опущены, остается шум в сличениях времени. Этот шум может рассматриваться как белый шум, характеризующийся стандартными отклонениями σ_C. Соответствующая неопределенность в разности частоты, выводимая из двух сличений времени на даты t_1 и t_2 равна

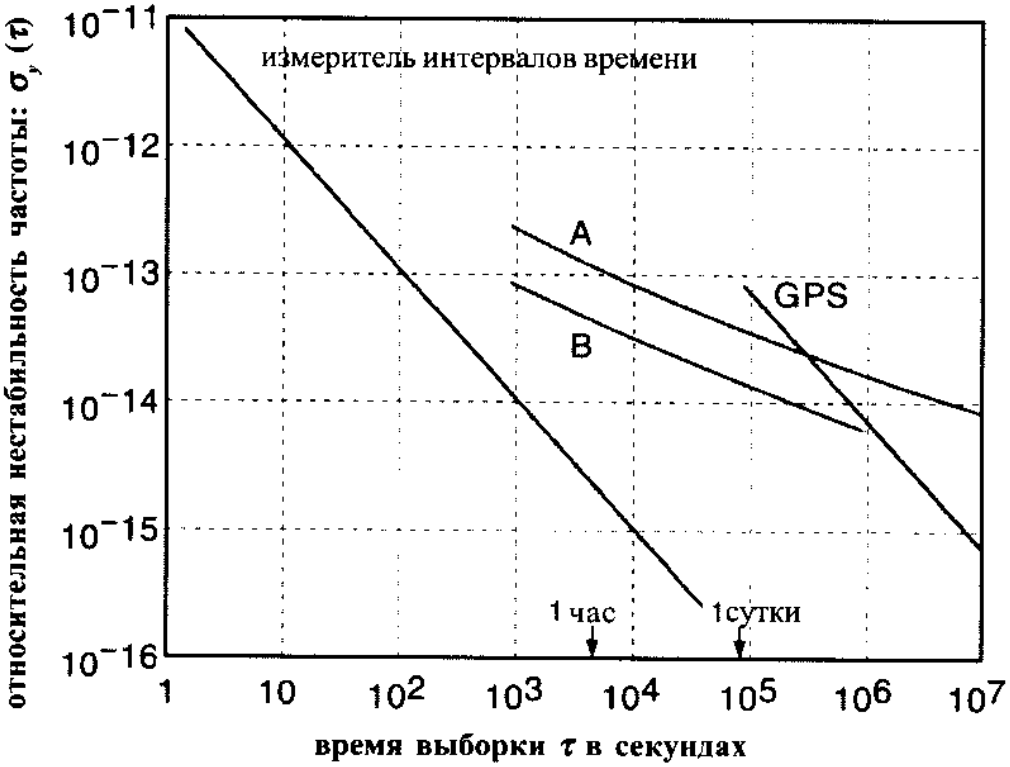


Рис. 5.6. Нестабильность частоты, обусловленная сличениями времени: GPS (±5 нс) и измерители интервалов времени (±10 пс). Для сравнения показана взаимная нестабильность первичных стандартов CS1 и CS2 в PTB (Германия) на кривой A и CS2 и водородного мазера на кривой B.

$$\sigma_y = \frac{\sqrt{2}\sigma_C}{t_2 - t_1}. \tag{5.69}$$

В обычном графике, показывающем нестабильность частоты осцилляторов, нестабильность, выведенная через сличения времени между двумя идеальными стабильными осцилляторами, имеет форму прямой линии с наклоном –1 (см. рис. 5.6). Мы сразу же можем выбрать такой интервал измерений, на котором осцилляторы не кажутся деградирующими.

Точность сличений частоты возрастает при увеличении интервала (t_1 и t_2). Но заметим, что неопределенность в 1 нс при сличениях времени за интервал в 1 день ведет к нестабильности $1{,}6 \times 10^{-14}$ в нормализованной частоте. В настоящее время (2000 год) трудно сличить разнесенные часы с неопределенностью лучшей, чем 100 пс. При частотных сличениях лучших первичных стандартов это требует усреднений, продолжающихся в течение нескольких дней, и при этом исключается изучение кратковременных вариаций частоты. Тем не менее ничего лучшего сделать нельзя. Это есть пример важности сличений времени и ограничений, продиктованных неопределенностями, связанными с ними.

Эти методы являются подходящими для изучения точности и кратковременной стабильности. Метод двухстороннего приема с отличным раз-

решением по времени обеспечивает наилучшие результаты. Прием GPS сигналов является менее точным, но более надежным методом.

(2) Сличения через международное атомное время. В лабораториях, оборудованных временными приемниками GPS, международное время может быть установлено в пределах нескольких наносекунд, но с задержками от 30 до 60 дней в зависимости от даты месяца, которая нас интересует (это происходит от того, что используется ежемесячная организация вычислений для определения TAI). Если нет необходимости в спешке, то это простой способ сличения часов в лабораториях. TAI можно заменить всемирным координированным временем (UTC) применяя меры предосторожности по возможному включению скачущей секунды. Эти две шкалы времени являются представлениями $\underline{t}$ (см. раздел 3.3.2 E). В следующих формулах мы обозначим TAI или UTC через $\underline{t}^*$. Мы вычислим скорости N_A часов A по отношению к TAI через

$$N_A = \frac{[\tau_A - \underline{t}^*](\underline{t}_2^*) - [\tau_A - \underline{t}^*](\underline{t}_1^*)}{\tau_A(\underline{t}_2^*) - \tau_A(\underline{t}_1^*)}, \tag{5.70}$$

так, что

$$N_A = 1 - \left\langle \frac{\mathrm{d}\underline{t}^*}{\mathrm{d}\tau_A} \right\rangle = 1 - \left\langle \frac{\mathrm{d}\underline{t}^*}{\mathrm{d}\underline{t}} \right\rangle \left\langle \frac{\mathrm{d}\underline{t}}{\mathrm{d}\tau} \right\rangle \left\langle \frac{\mathrm{d}\tau}{\mathrm{d}\tau_A} \right\rangle. \tag{5.71}$$

Мы увидим в главе 7, что длительность единицы шкалы TAI в секундах на вращающемся геоиде может быть получена из публикаций BIPM. Пусть D_{TAI} численное значение среднего времени за рассматриваемый интервал, тогда

$$\left\langle \frac{\mathrm{d}\underline{t}^*}{\mathrm{d}\underline{t}} \right\rangle = (D_{TAI})^{-1}. \tag{5.72}$$

Используя (3.29) и (5.50) мы получим

$$\langle \Phi_A \rangle = (D_{TAI})^{-1} + N_A + \left\langle \underline{h}(\underline{t}^*, \mathbf{x}(\underline{t}^*)) \right\rangle \tag{5.73}$$

с достаточным уровнем точности.

Произведя то же самое с часами B на том же самом интервале $\underline{t}^*$, мы получим $\langle \Phi_A - \Phi_B \rangle$, которое не зависит от величины D_{TAI}.

Этот метод, также как и предыдущий, является единственным реально подходящим для точности и долговременной стабильности по крайней мере для стандартов высокого качества. Стабильность TAI достаточно хороша, так что нет необходимости в строгой одновременности измерений. Стоит

отметить, что (5.73) дает прямой доступ к собственной секунде в пределах точности частоты TAI. Мы вернемся к этому вопросу в главе 7.

(b) Сличения частоты, основанные на фазовых измерениях

Локальные сличения частот двух осцилляторов могут быть выполнены путем использования методов, описанных в разделе 5.2.3.

Для сличений разнесенных частот в качестве промежуточных можно использовать частоты радиопередач. Как объясняется в главе 4, этот метод не представляет больше интереса для фундаментальной метрологии. Излучение стабильных частот является простым и экономичным способом чтобы поделиться стандартами частоты с потребителями, которые не нуждаются в предельных точностях.

Давайте тем не менее запишем соотношения между нормализованной собственной частотой $[\Phi_E]_E$ излучателя E, измеренной в пункте E и нормализованный собственной частотой $[\Phi_E]_A$ в пункте приема A,

$$\frac{[\Phi_E]_A}{[\Phi_E]_E} = \frac{1 - c^{-1}(\mathbf{v}_A \cdot \mathbf{n}) + c^{-2}[U(\mathbf{x}_A) + v_A^2/2]}{1 - c^{-1}(\mathbf{v}_E \cdot \mathbf{n}) + c^{-2}[U(\mathbf{x}_E) + v_E^2/2]}. \tag{5.74}$$

В этом соотношении скорости выражены в невращающейся геоцентрической координатной системе. $\mathbf{v}_E$ это скорость излучателя и $\mathbf{v}_A$ скорость приемника в моменты излучения и приема соответственно. Потенциал U оценивается в пунктах E и A и $\mathbf{n}$ единичный вектор EA, сконструированный из пространственных координат в направлении распространения. Это уравнение справедливо примерно до $\pm 10^{-14}$ в вакууме для антенн, установленных на спутниках на расстояниях вплоть до геостационарных спутников.

Используя тоже самое соотношение к приемнику B, в пункте B мы будем иметь соотношение нормализованных частот, полученных для того же самого излучения:

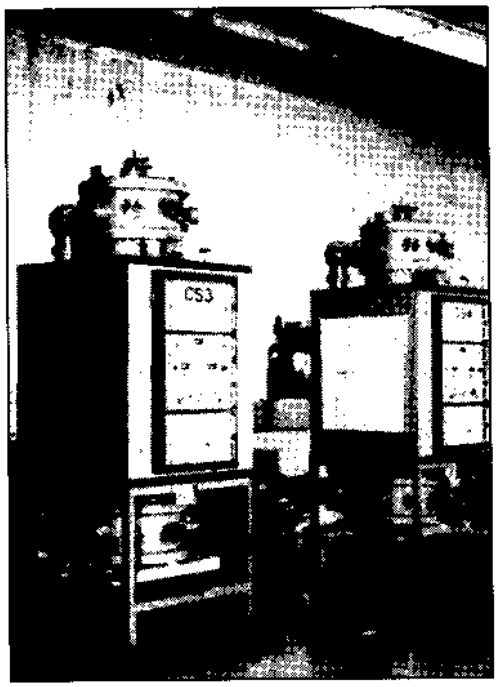
$$\frac{[\Phi_E]_A}{[\Phi_E]_B} = \frac{1 - c^{-1}(\mathbf{v}_A \cdot \mathbf{n}) + c^{-2}[U(\mathbf{x}_A) + v_A^2/2]}{1 - c^{-1}(\mathbf{v}_B \cdot \mathbf{n}) + c^{-2}[U(\mathbf{x}_B) + v_B^2/2]}. \tag{5.75}$$

Сличения между нормализованными собственными частотами Φ_{HA} и Φ_{HB} стандартов, расположенных в A и B, выражается через

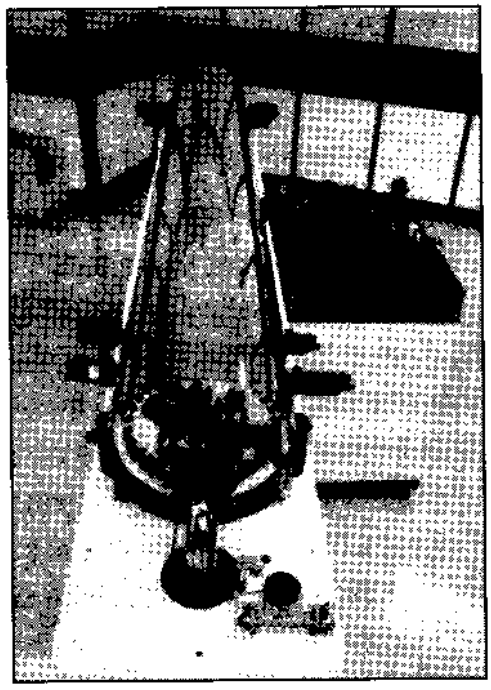
$$\frac{\Phi_{HA}}{\Phi_{HB}} = \frac{\Phi_{HA}}{[\Phi_E]_A} \frac{[\Phi_E]_B}{\Phi_{HB}} \frac{[\Phi_E]_A}{[\Phi_E]_B}. \tag{5.76}$$

Два первых соотношения с правой стороны находятся из локальных измерений, в то время как третье вычисляется через (5.75).

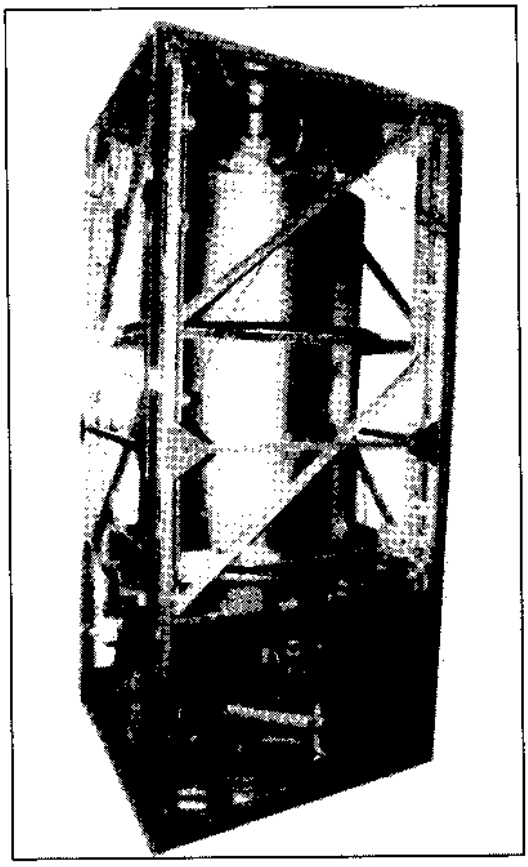
В применениях к односторонним передачам со спутников точность ограничена примерно до 10^{-13} из-за эффектов рефракции.



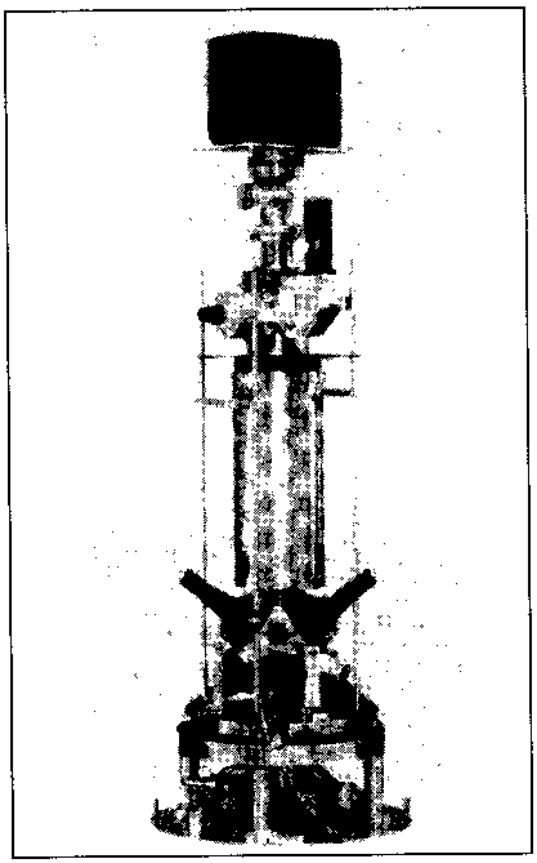
Первичный цезиевый стандарт CS3 и CS4 с магнитным отклонением пучка. (благодаря любезности А.Бауха, PTB)



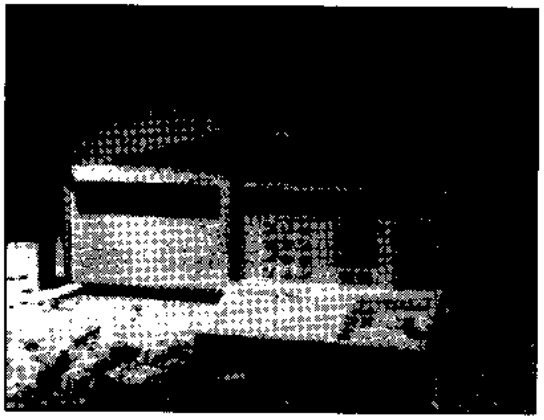
Первичный стандарт NIST7 с оптической накачкой пучка (благодаря любезности Р.Е.Друллингера, NIST)



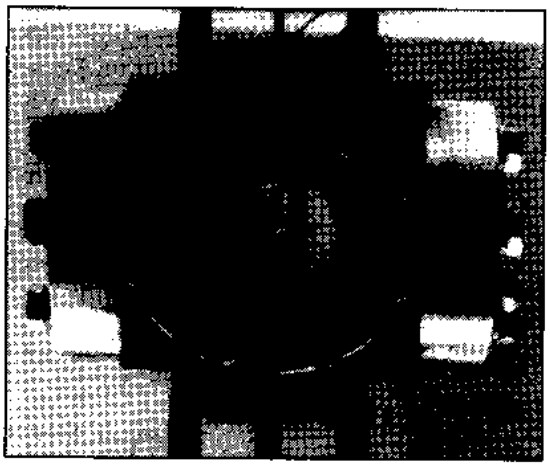
Первичный стандарт F01, основанный на фонтане охлажденных лазером атомов цезия (благодаря любезности А.Клерона, LPTF)



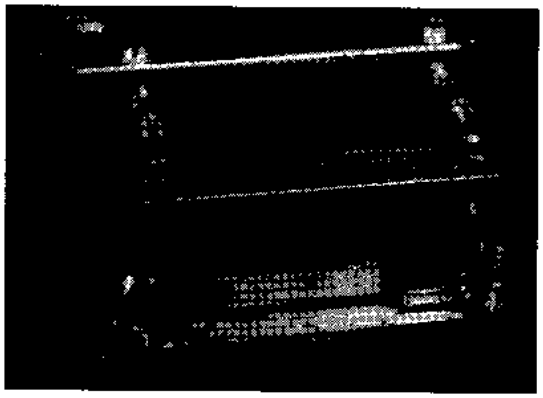
Прототип космических атомных часов на холодных атомах (благодаря любезности А.Клерона и П.Лорана, LPTF)



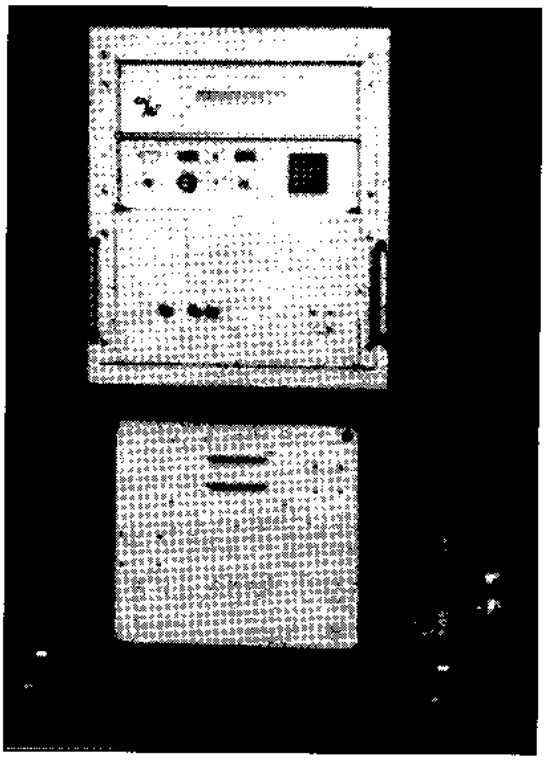
Промышленные цезиевые часы (благодаря любезности Д.Феранда, Хьюлетт Паккард)



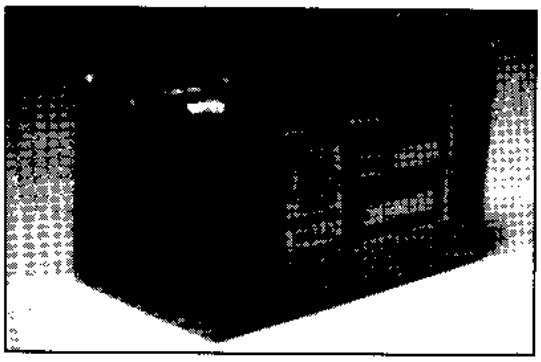
Линейная радиочастотная ловушка для удержания ионов ртути (благодаря любезности Ж.Д.Престижа, JPL)



Компоненты компактного прототипа оптически накачиваемых цезиевых часов (благодаря любезности П.Серэ, LNA)



Водородный мазер (благодаря любезности Р Барийе и П.Пти, LNA)



Компактные рубидиевые часы (благодаря любезности П Роша, TNT)

Глава 6

Атомные стандарты частоты

Цель этой главы состоит в описании принципов работы и особенностях основных типов атомных часов, таких как: цезиевые часы, водородные мазеры, рубидиевые часы и часы на удержанных ионах. Основная часть обсуждений будет связана с цезиевыми часами, так как они используются для реализации современного определения секунды.

Читатель также найдет общее описание этих стандартов частоты например в [6.1, 6.2, 6.3], а для глубокого изучения в [6.4].

6.1. Спектроскопические основы атомных стандартов частоты

6.1.1. Инвариантность и универсальность атомных свойств

Вплоть до современных дней все экспериментальные наблюдения в спектроскопии доказывали совместимость с гипотезой, что атомные свойства одни и те же во все времена и во всех местах, если эти свойства оцениваются наблюдателем, находящимся в непосредственной близости с атомом и движущимся вместе с атомом одинаковым путем. Этот постулат является очевидным характерным примером эйнштейновского принципа эквивалентности, сформулированного в разделе 3.2.1e.

Этот постулат дает возможность построить приборы, которые, используя определенный атомный переход, способны в реальном времени производить сигнал с одной и той же частотой везде и всегда, при условии, что релятивистские эффекты из-за несовпадения атома и наблюдателя учтены должным образом. Частота зависит только от фундаментальных физических констант.

Тот факт, что частоты переходов различных атомов остаются в постоянном соотношении друг с другом в пределах экспериментальной неопределенности, является очень точным подтверждением принципа эквивалентности в области спектроскопии.

6.1.2. Спектральные свойства щелочных атомов

Основные атомные часы используют преимущество переходов в сверхтонкой структуре атомов водорода, рубидия, цезия или ионизированной ртути. Щелочные металлы, включая рубидий и цезий, имеют один непарный электрон на внешней электронной оболочке. Их внутренние электронные подоболочки либо заполнены полностью, либо пусты. Ион ртути Hg^+ имеет похожую электронную конфигурацию и мы можем назвать его «щелочным» продолжением. То же самое относится к атому водорода, обладающего только одним электроном.

Давайте сделаем обзор основных свойств щелочных атомов в той степени, в которой это необходимо для описания работы атомных часов [6.5].

(а) Тонкая структура

Самый дальний (внешний) электрон в электронной оболочке характеризуется главным квантовым числом n. Для тех уровней энергии, которые мы будем рассматривать, величина n приведена в таблице 6.1.

Орбитальный угловой момент, обозначаемый через $\hbar\mathbf{L}$, связан с движением электрона вокруг ядра. Здесь $\hbar$ обозначает постоянную Планка h деленную на 2π·L – соответствующее квантовое число. В щелочных металлах общий орбитальный момент уменьшается до момента одиночного внешнего электрона. Мы будем рассматривать только два значения L, а именно: $L = 0$ и $L = 1$.

Для $L = 0$ энергия внешнего электрона минимальна. Говорят, что атом находится в *основном состоянии*. При использовании стандартных спектроскопических обозначений[1] это состояние обозначается как $^2S_{1/2}$. Под действием соответствующей энергии возбуждения атом может перейти в *возбужденное состояние*. Первое такое состояние соответствует $L = 1$. Разность энергий между основным состоянием и первым возбужденным состоянием составляет порядка нескольких электрон-вольт (eV). Отсюда следует, что если осуществляется переход между основным и возбужденным состоянием путем поглощения или испускания электромагнитного излучения, то длина волны этого электромагнитного излучения лежит в оптической области электромагнитного спектра.

---

[1] Спектроскопические обозначения установились из давних соглашений. Заглавная буква характеризует величину квантового числа орбитального углового момента. Например, S и P соответствуют L=0 и L=1. Буква S не должна смешиваться со спиновым квантовым числом. Верхний индекс перед буквой дает величину $2S$+1, называемую мультиплетностью, здесь S теперь обозначает спиновое квантовое число. Нижний индекс, расположенный после буквы определяет значение квантового числа J для общего углового момента. Если перед этим обозначением стоит цифра, то ее значение определяет главное квантовое число. Например, основное состояние цезия обозначается так $6^2S_{1/2}$.

Все электроны обладают присущим им внутренним свойством, известным как спин. Спиновое квантовое число S имеет значение 1/2. Угловой момент $\hbar\mathbf{S}$ связан с электронным спином. Спиновый угловой момент щелочных атомов такой же как и у непарного электрона.

Общий угловой момент $\hbar\mathbf{J}$ электрона является суммарным вектором его орбитального углового момента и спинового углового момента. Согласно комбинационным правилам атомных угловых моментов соответствующее квантовое число J может принимать только значения $L+1/2$ или $|L-1/2|$. Это означает, что $J=1/2$ в основном состоянии и $J = 1/2$ или $J = 3/2$ в первом возбужденном состоянии. Двум различным возможным значениям J в этих возбужденных состояниях соответствуют два слабо отличающихся энергетических уровня. Они называются *уровнями тонкой структуры* и в спектроскопических терминах обозначаются как $^2P_{1/2}$ и $^2P_{3/2}$. Разность энергии между ними может быть выражена через постоянную тонкой структуры α, значение которой примерно равно 1/137. Эти уровни важны для метода оптической накачки, который мы вскоре будем обсуждать. Две очень близкие спектральные линии в спектре излучения или поглощения щелочных атомов обозначаются через D_1 и D_2. Они обусловлены переходами между $^2S_{1/2}$ и $^2P_{1/2}$ для первой линии, и $^2S_{1/2}$ и $^2P_{3/2}$ для второй. Длины волн для дублетов тонкой структуры водорода, рубидия, цезия и иона ртути Hg^+ приведены в таблице 6.1.

Таблица 6.1. Основные спектроскопические особенности щелочных атомов, использующихся для создания атомных стандартов времени и частоты. Значение частоты сверхтонкого перехода в цезии 133 установлено по определению секунды. Для других атомов указываются неопределенности измеренных значений частот переходов.

Атом (или ион)	Атомная масса	n	I	F	$^\lambda D_1$ (нм)	$^\lambda D_2$ (нм)	ν_0 (Гц)
H	1	1	1/2	0;1	121,5673	121,5668	1 420 405 751,770(3)
Rb	87	5	3/2	1;2	794,8	780,0	6834682610,904335(12)
Cs	133	6	7/2	3;4	894,3	852,1	9 192 631 770
Hg^+	199	6	1/2	0;1	194,2	165,0	40507347996,8416(4)

(b) Сверхтонкая структура

Протоны и нейтроны также обладают спином со значением 1/2. Ядерный угловой момент $\hbar\mathbf{I}$ находится из определенной комбинации составляющих нуклонов. Значение I, связанное с квантовым числом, является не-

четным кратным от 1/2 для щелочных изотопов с нечетной атомной массой (см. таблицу 6.1).

Общий угловой момент $\hbar\mathbf{F}$ атома является векторной суммой углового момента электронного облака и ядра. Квантовое число F, связанное с общим угловым моментом, принимает такие значения, что $|I-J| \le F \le I+J$ и F может изменяться только на целое число в указанных границах. В основном состоянии, а также в состоянии $^2P_{1/2}$ F равно $I-1/2$ или $I+1/2$. В состоянии $^2P_{3/2}$ число возможных значений F больше 2, если $I > 1/2$.

Проекция углового момента $\hbar\mathbf{F}$ на ось квантования может принимать только дискретный набор значений равный $\hbar m_F$. Числа m_F являются магнитными квантовыми числами, связанными с F. Они принимают $2F+1$ возможных значений от $-F, -F+1, \ldots, F-1, F$.

На рис. 6.1 показаны уровни энергии атома цезия в состояниях $^2S_{1/2}$, $^2P_{1/2}$ и $^2P_{3/2}$, а также возможные значения F для этих случаев. В дальнейшем F' будет обозначать значение F в возбужденных состояниях $^2P_{1/2}$ и $^2P_{3/2}$.

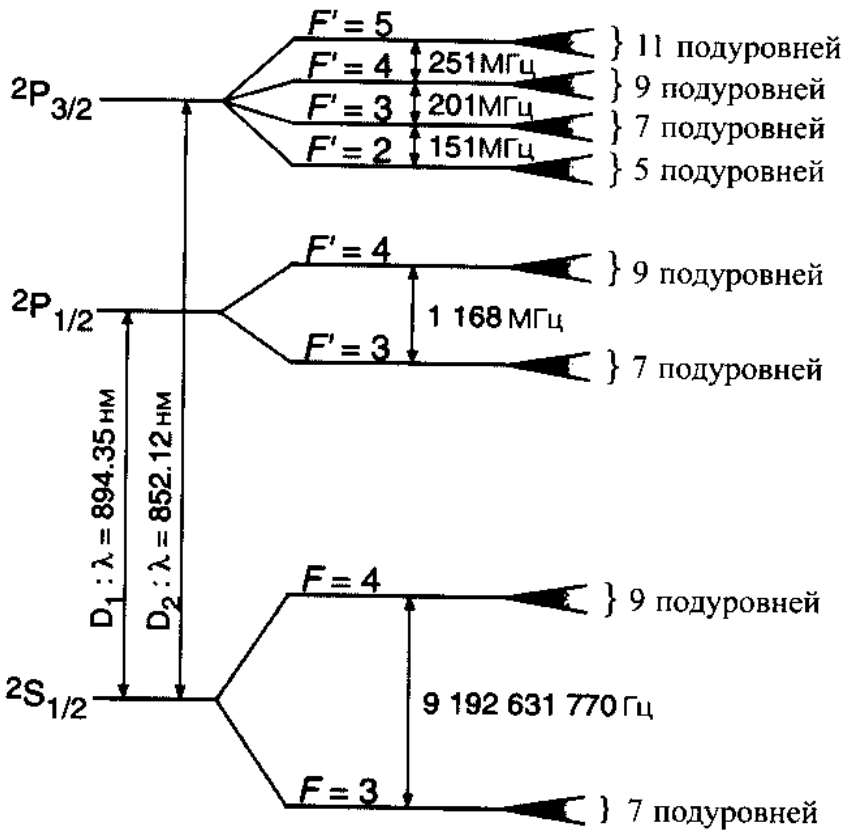


Рис. 6.1. Уровни энергий атома цезия в основном и первом возбужденном состоянии

Магнитные моменты связаны с ядерным и электронным спинами. Ядерный момент равен

$$\mu_I = g_I \mu_B \mathbf{I}, \qquad (6.1)$$

где μ_B – магнетон Бора и g_I – ядерный фактор Ландэ. Магнитный момент непарного электрона дается выражением

$$\mu_J = g\mu_B \mathbf{S}, \tag{6.2}$$

где g_J – фактор Ландэ для электрона, связанного с атомом. Величина g_J положительна и близка к 2. Величина g_I примерно в 1000 раз меньше. Она положительна для атомов и ионов, которые здесь рассматриваются. Например, для водорода $g_I = 3{,}042\ 01 \times 10^{-3}$.

Эти магнитные моменты дают в результате энергию магнитного взаимодействия. В нулевом магнитном поле магнитное взаимодействие расщепляет каждый энергетический уровень, рассмотренный выше, на столько же сверхтонких уровней, сколько допускается значений F или F'. Это расщепление иллюстрируется на рисунке 6.1.

Вычисление разности энергий между двумя сверхтонкими уровнями в нулевом внешнем поле приводит к выражению магнитного возмущения, создаваемого внешним электроном на ядро. В атоме водорода частота перехода ν_H между уровнями основного состояния $F = 0$ и $F = 1$ в первом приближении дается следующим выражением

$$\nu_H = \frac{4}{3}\frac{mc^2\alpha^4 g_I}{h}\left(1+\frac{m}{M_p}\right)^{-3}, \tag{6.3}$$

где m и M_p являются массами электрона и протона соответственно, c – скорость света и α –постоянная тонкой структуры, равная $e^2/4\pi\,\varepsilon_0\hbar c$, где e – заряд электрона и ε_0 – диэлектрическая постоянная вакуума. Эта приблизительная формула дает значение частоты сверхтонкого перехода с относительным отклонением 10^{-3} от экспериментально определенного значения. Фактически следует учесть большое число поправок, выраженных через постоянную тонкой структуры, однако сейчас этот вопрос мы обсуждать не будем. Для этого требуются современные сложные вычисления. В настоящее время теоретическое выражение для частоты сверхтонкого перехода в атоме водорода в основном состоянии известно с такой же точностью, что и фундаментальные физические константы.

Вычисление сверхтонкого расщепления в основном состоянии в щелочных атомах является еще более сложной задачей, чем вычисления для водорода. Действительно, даже в первом порядке приближения мы должны принять во внимание наличие электронов на внутренних оболочках.

Как общее правило, энергия магнитного взаимодействия между внешним электроном и ядром очень мала. Отсюда следует, что частота сверхтонкого перехода ν_0 в основном состоянии лежит в диапазоне сверхвысоких частот, как это можно увидеть из таблицы 6.1.

Статическая плотность магнитного потока[2] **B** определяет направление квантования и расщепляет сверхтонкие уровни на столько сверхтонких подуровней, сколько имеется возможных значений m_F, т.е. $2F+1$. Это явление называется *эффектом Зеемана*. Для щелочных атомов в основном состоянии изменение энергии $E(F, m_F)$ каждого из этих подуровней может быть точно представлено функциональной зависимостью от внешнего поля B. Оно дается формулой Брайта-Раби

$$E(F,m_F)=-\frac{h\nu_0}{2(2I+1)}-g_I\mu_B Bm_F+\varepsilon\frac{h\nu_0}{2}\left(1+\frac{4m_F}{2I+1}x+x^2\right)^{1/2}, \qquad (6.4)$$

где

$$x=\frac{(g_J+g_I)\mu_B B}{h\nu_0}. \qquad (6.5)$$

Если $F=I+1/2$, то $\varepsilon = +1$, а если $F=I-1/2$, то $\varepsilon = -1$. Эту формулу можно использовать для вычисления сдвига частоты любого перехода между двумя сверхтонкими подуровнями в зависимости от магнитного поля. Рисунки 6.2 и 6.3 показывают изменения в $E(F, m_F)$ для атома водорода и цезия соответственно.

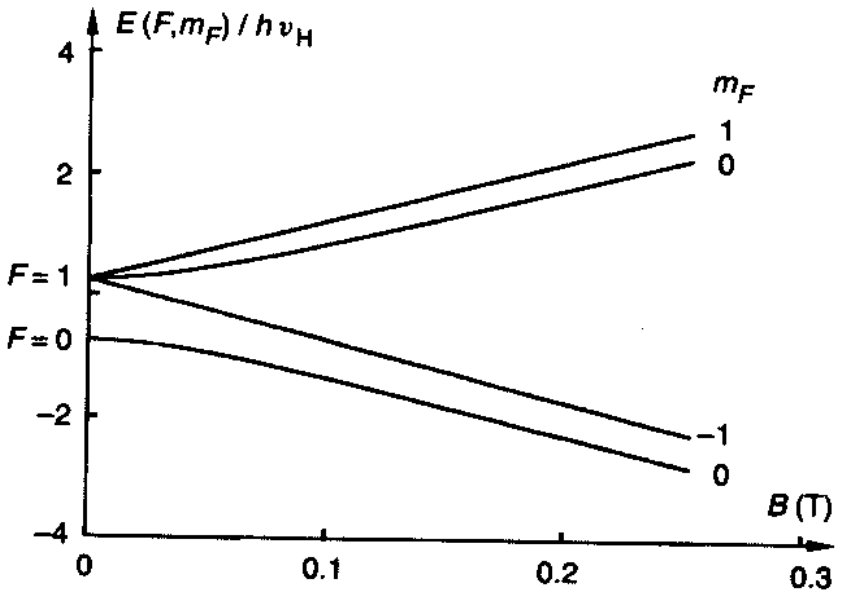


Рис. 6.2. Изменение энергии сверхтонких подуровней основного состояния атома водорода как функция внешнего магнитного поля

Уравнение (6.4) и рисунки 6.2 и 6.3 обнаруживают некоторые заметные свойства.

[2] В дальнейшем мы будем использовать более общий термин «магнитное поле».

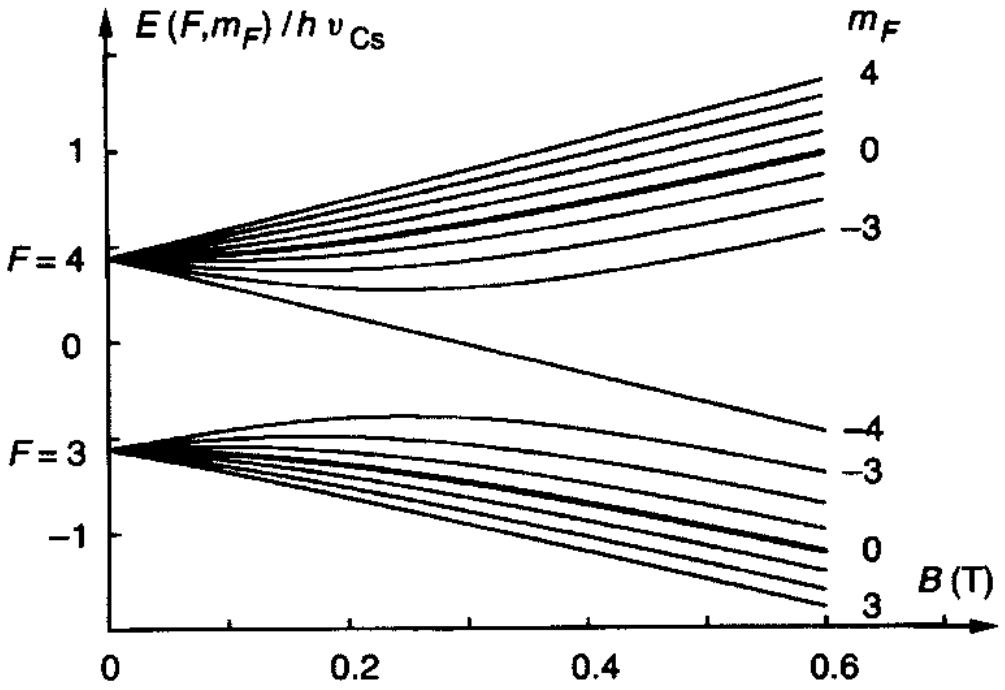


Рис. 6.3. Изменение энергии сверхтонких подуровней основного состояния атома цезия в зависимости от внешнего магнитного поля

(1) Энергия подуровней с $m_F = 0$ изменяется очень медленно при малых магнитных полях. Переход между этими уровнями иногда называют *часовым переходом*, т.е. переходом, использующимся в атомных часах. Изменение $\Delta\nu_Z$ в частоте перехода между двумя подуровнями с $m_F = 0$ зависит квадратично от поля B согласно выражению:

$$\Delta\nu_Z = K_0 B_0^2, \tag{6.6}$$

где коэффициент K_0 является функцией только атомных постоянных и B_0 обозначает величину магнитного поля для случая малых полей. Она обычно составляет 10^{-7} Т для водородного мазера и 6×10^{-6} Т для цезиевых пучковых часов. Соответствующие значения $\Delta\nu_Z$ равны тогда $2{,}77\times10^{-3}$ Гц и 1,5388 Гц соответственно.

(2) Для малых величин магнитного поля уровни энергий с $m_F \neq 0$ изменяются линейно или почти линейно. Следовательно, для тех же упоминавшихся выше величин B_0 частота перехода водорода

$$F = 1, m_F = 0 \leftrightarrow F = 1, m_F = \pm 1$$

равна 1,399 кГц, а частота переходы цезия

$$F = 3, m_F = 1 \leftrightarrow F = 4, m_F = 1$$

равна 42,050 кГц. Поэтому путем измерения частоты некоторых переходов между сверхтонкими подуровнями можно установить величину действующего магнитного поля. Тогда сдвиг частоты для часового перехода может быть выведен из (6.6).

(3) Для больших значений магнитного поля уровни энергий разделяются на две группы. Физическая причина этого явления состоит в том, что магнитная энергия, приобретенная электронами, превышает их энергию взаимодействия с ядром. Таким образом магнитное поведение атомов начинает становиться похожим на поведение внешнего электрона. Энергия последнего является возрастающей функцией B_0 если спин ориентируется антипараллельно направлению магнитного поля, и убывающей функцией магнитного поля B_0 если спин ориентируется параллельно вектору $\mathbf{B}_0$. Поэтому подуровни с m_F=0 принадлежат различным группам. Таким образом, атом может быть в состоянии либо с одним, либо с другим уровнями энергии, между которыми происходит часовой переход. Эта особенность лежит в основе *метода магнитной селекции* (см. раздел 6.2.2).

Возбужденные состояния также обладают сверхтонкой структурой если $I \neq 0$. Рисунок 6.1 показывает сверхтонкую структуру уровней для $^2P_{1/2}$ и $^2P_{3/2}$ состояний атомов цезия, а также число подуровней, соответствующих различным возможным значениям m_F. Разность энергий сверхтонких уровней меньше чем в основном состоянии.

Электрическое поле, действующее на атом, сдвигает его энергетические уровни. Это явление называется *Штарк-эффектом*. Этот эффект для рассматриваемых здесь уровней энергии имеет квадратичную зависимость от поля. На практике связанные с ним сдвиги частоты очень малы. Очень малая чувствительность к электрическому полю дает возможность удержания ионов, таких как ионы ртути, для того чтобы наблюдать их часовые переходы в оптимальных условиях.

Давайте только упомянем, что квантовое число I равно нулю для большинства изотопов с четной атомной массой. Поэтому у них отсутствует сверхтонкая структура.

6.1.3. Правила отбора

Взаимодействие между атомом и электромагнитным полем не происходит, если не соблюдаются законы сохранения энергии и углового момента.

Пусть E_1 и E_2 энергии двух атомных уровней ($E_2 > E_1$). Необходимым условием взаимодействия является то, что частота электромагнитной волны должна удовлетворять условию Бора:

$$E_2 - E_1 = h\nu_0, \tag{6.7}$$

выражающему закон сохранения энергии.

Обмен энергией может происходить через три различных процесса.

- *Поглощение.* Падающий фотон с энергией $h\nu_0$ поглощается атомом, находи вшимся в состоянии E_1. Атом переходит в состояние с энергией E_2, и фотон исчезает.

- *Спонтанное излучение.* Атом, предварительно переведенный на энергетический уровень E_2, спонтанно затухает в случайные моменты времени без внешних причин. Его энергия падает до E_1 и порождается фотон с энергией $h\nu_0$, который называется *флуоресцентным фотоном.*
- *Стимулированное* или *индуцированное излучение.* Этот тип взаимодействия, предсказанный Эйнштейном, является основой для *лазерного* или *мазерного* эффектов. Возбужденный атом с энергией E_2 в присутствии фотона с энергией $h\nu_0$ переходит в состояние с энергией E_1 под влиянием падающего электромагнитного излучения. Таким образом, создается фотон, но без исчезновения участвовавшего в процессе фотона. Ключевая особенность стимулированного излучения состоит в том, что электромагнитная волна, связанная с излученным фотоном находится в фазе с падающей волной так, что последняя усиливается.

Для того чтобы происходили переходы между двумя атомными уровнями, условие (6.7) необходимо, но недостаточно. Должны также выполняться правила отбора. Они происходят из свойств атомного углового момента. Сохранение углового момента при взаимодействии с излучением требует, чтобы излучение имело определенную поляризацию.

Давайте рассмотрим переходы между сверхтонкими подуровнями в данном состоянии, т.е. в основном состоянии. Они включают в себя часовые переходы и переходы, используемые для измерения действующего магнитного поля B_0. Разрешено два типа переходов:

- переходы с $\Delta F = \pm 1$ и $\Delta m_F = 0$, для которых необходимо чтобы магнитное поле сверхвысокой частоты (СВЧ) было параллельно статическому магнитному полю;
- переходы с $\Delta F=0$ и $\Delta m_F=\pm 1$ или $\Delta F=\pm 1$ и $\Delta m_F=\pm 1$, для которых необходимо чтобы магнитное поле низкой частоты ($\Delta F=0$) или СВЧ ($\Delta F=\pm 1$) было перпендикулярно статическому магнитному полю.

Переходы также происходят между подуровнями сверхтонкой структуры, принадлежащим различным состояниям, таким как $^2S_{1/2}$, $^2P_{1/2}$ и $^2P_{3/2}$. Они появляются в оптической накачке атомов цезия, рубидия и ионов ртути, а также при оптическом детектировании часовых переходов этих атомов. Переходы между уровнями тонкой структуры $^2P_{1/2}$ и $^2P_{3/2}$ запрещены. Возможными переходами являются переходы между уровнями состояния $^2S_{1/2}$ и уровнями состояния $^2P_{1/2}$ с одной стороны, а также между уровнями состояния $^2S_{1/2}$ и уровнями состояния $^2P_{3/2}$ с другой стороны. Разрешенными являются те переходы, для которых или $F–F'=\pm 1$ и $m_F–m_F=0$, или $F–F'=0$ или ± 1, и $m_F–m_F=\pm 1$. Для упрощения мы не указываем на соответствующую поляризацию излучения.

Правила отбора позволяют нам сделать выбор между различными возможными переходами.

6.1.4. Ширина резонанса

На практике стандарты времени и частоты спроектированы так, что доплеровское уширение не действует на резонанс (см. раздел 6.2.3). В таком случае ширина резонанса связана с временем наблюдения Δt через гейзенберговское соотношение неопределенности между энергией и временем

$$\Delta\nu_0 \sim \frac{1}{\Delta t}. \tag{6.8}$$

Вероятность спонтанного излучения в единицу времени зависит в кубической степени от частоты перехода между соответствующим уровнем и основным состоянием. На практике это проявляется в том, что время жизни τ уровней тонкой структуры оказывается очень малым, порядка нескольких наносекунд например 33 нс для атома Cs и 2 нс для иона ртути Hg^+. Естественная ширина $1/2\pi\tau$ рассматриваемого резонанса, т.е. ширина, обусловленная только спонтанным излучением, близка к 5 МГц для Cs и 100 МГц для Hg^+.

Напротив, сверхтонкие уровни основного состояния с $F = I+1/2$ имеют много большее время жизни, порядка нескольких лет, потому что частоты переходов в основном состоянии очень малы по сравнению с частотами оптических переходов. Наиболее низкий подуровень энергии с $F = I-1/2$ имеет почти бесконечное время жизни. Поэтому можно сказать, что спонтанное излучение не оказывает практически никакого эффекта на ширину сверхтонкого резонанса в основном состоянии.

В атомных стандартах частоты типичное время наблюдения составляет 1 мс в рубидиевых и в некоторых цезиевых часах и несколько секунд в часах на ионах ртути Hg^+. Ширина атомного резонанса лежит в пределах от 10^3 до 10^{-1} Гц.

Отсюда следует, что для типовых значений действующего магнитного поля различные переходы хорошо разделены один от другого. В первом приближении, среди всех уровней сверхтонкой структуры только два фактически взаимодействуют с магнитным полем сверхвысокой частоты. Эволюция атомных свойств может таким образом трактоваться в упрощенном приближении к двухуровневой системе.

Добротность $Q_{at} = \nu_0/\Delta\nu_0$ атомного резонанса очень высока. Она находится между 10^7 и 10^{12}. Очень высокий уровень этого параметра является решающим фактором при рассмотрении стабильности атомных стандартов частоты. В некоторой степени он дает также вклад и в их точность.

6.2. Стадии работы атомных часов

Во всех типах атомных часов осуществляются похожие операции. Атомы рассматриваются в предельно разреженных условиях. Обеспечивается разность населенностей сверхтонких уровней. Часовой переход осуществляется в условиях, когда возможна спектроскопия сверхвысокого разрешения. Осуществляется детектирование перехода. Слабое магнитное поле, защищенное от внешнего возмущения, настраивается на ту область, где атомы взаимодействуют с электромагнитным излучением. Используются электронные средства для обеспечения потребителей практически приемлемыми сигналами.

6.2.1. Наблюдение в очень разреженных средах

Очевидно, что атомы должны быть по возможности слабо возмущенными из-за присутствия соседних атомов. Вот почему в атомных часах атомы всегда разрежены. Это позволяет минимизировать скорость, с которой атомы взаимодействуют между собой. Атомы наблюдаются в атомных пучках, в диффузных облаках ячеек низкого давления. В атомных пучках (например, в цезиевых часах) вероятность столкновений ниже, чем в ячейках при той же плотности атомов. Облака могут испытывать баллистическое движение, как в случае фонтана холодных атомов, или быть ограниченными в движении, как в случае ртутных ионных часов. Ячейки могут быть реализованы физически, как в случае водородного мазера или рубидиевых часов или могут быть осуществлены электрическими полями (например, ртутные ионные часы). Кроме рубидиевых часов с отпаянными ячейками во всех остальных случаях осуществляется эффективная откачка остаточных газов для поддержания высокого вакуума.

6.2.2. Подготовка и детектирование атомных состояний

Для эффективного наблюдения переходов между основными состояниями сверхтонких уровней требуется подготовка атомных систем.

Рассмотрим два подуровня сверхтонкой структуры, с энергиями E_1 иE_2 (E_2>E_1). Пусть n_1 есть населенность уровня E_1 в момент времени t, т.е. число атомов с энергией E_1 в этом момент времени и аналогично n_2 населенность уровня E_2 в этот же момент времени. Под действием электромагнитного излучения за промежуток времени dt поглощение стимулирует переходы с уровня E_1 на уровень E_2 так, что населенность E_2 возрастает на dn_2. Тогда

$$\mathrm{d}n_2 = p_{12}\, n_1 \mathrm{d}t, \tag{6.9}$$

где p_{12} – вероятность поглощения в единицу времени. Аналогично индуцированное излучение заставляет атомы осуществлять переходы с

уровня E_2 на уровень E_1, и населенность уровня E_2 изменяется, таким образом, на

$$dn_2 = -p_{21}n_2dt, \tag{6.10}$$

где p_{21} – вероятность индуцированного излучения в единицу времени.

Эйнштейн показал, что $p_{12}=p_{21}$. Общее изменение населенности энергетического уровня E_2 поэтому

$$dn_2=p_{12}(n_1-n_2)dt. \tag{6.11}$$

Вплоть до знака это также изменение в населенности уровня E_1, так как обмен ограничивается только этими двумя уровнями.

Уравнение (6.11) показывает, что населенность уровня изменяется больше в том случае, когда велика начальная разность населенностей n_1-n_2. Поэтому изменения в атомных свойствах будет соответственно проще наблюдать в том смысле, что отношение сигнал/шум будет больше, если увеличить разность населеностей между двумя уровнями.

При термодинамическом равновесии отношение населенностей двух невырожденных уровней, таких как те, которые мы здесь рассматривали в ненулевом магнитном поле, дается *законом Больцмана*. Следовательно,

$$\frac{n_1}{n_2} = \exp\frac{E_2 - E_1}{kT}, \tag{6.12}$$

где k – постоянная Больцмана и T – температура. Если мы предположим, что частота перехода между двумя уровнями равна 10 ГГц и температура 300 К, тогда

$$\frac{n_1 - n_2}{n_2} = 1{,}6\times10^{-3}.$$

При термодинамическом равновесии относительная величина разностей населенности для двух уровней, между которыми происходит часовой переход, очень мала.

Для того чтобы обеспечить условия уверенного детектирования сверхтонкого перехода, это равновесие необходимо нарушить с помощью методов, которые создают максимально возможную разность населенностей, а именно *магнитное отклонение* [6.6] или *оптическую накачку* [6.7]. Здесь нам приходится платить за то преимущество, что мы используем очень низкие частоты сверхтонких переходов.

(а) Магнитное отклонение

Как можно видеть на рисунках 6.2 и 6.3, потенциальная энергия атомов водорода в состоянии F=0, m_F=0 и атомов цезия в состоянии F=3, m_F=0, уменьшается при возрастании магнитного поля. Поэтому они втягивают-

ся в области сильного магнитного поля, так как каждая физическая система стремится к положению с минимальной потенциальной энергией. Напротив, атомы водорода в состоянии F=1, m_F=0 и атомы цезия в состоянии F=4, m_F=0, выталкиваются из области с сильным магнитным полем. Можно показать, что сила **F**, действующая на атом равна

$$\mathbf{F}(F,m_F) = -\frac{\partial E(F,m_F)}{\partial B}\nabla B, \qquad (6.13)$$

где E (F, m_F) дается выражением (6.4) и ∇B является градиентом индукции магнитного поля. Абсолютная величина $-\partial E(F, m_F)/\partial B$ максимальна в сильных полях, где она равна магнетону Бора. На практике используются поля с магнитной индукцией около 1 Тесла и градиентом порядка несколько тесла на сантиметр.

В сильном неоднородном поле два состояния с m_F=0 подвергаются действию противоположных сил. В атомном пучке атомы в одном из этих состояний могут быть направлены в зону, где они используются для работы часов, в то время как атомы в другом состоянии переводятся в зону, где они не играют никакой роли.

(b) Оптическая накачка

Оптическая подготовка атомных состояний используется в цезиевых часах с оптической накачкой, рубидиевых часах и часах на ионах ртути.

Рассмотрим упрощенную диаграмму энергий на рис. 6.4а. Например, пусть энергетические уровни α и β являются уровнями сверхтонкой структуры с F = 3 и F = 4 атома цезия в основном состоянии, и уровень γ является уровнем сверхтонкой структуры с F' = 3 состояния $^2P_{3/2}$.

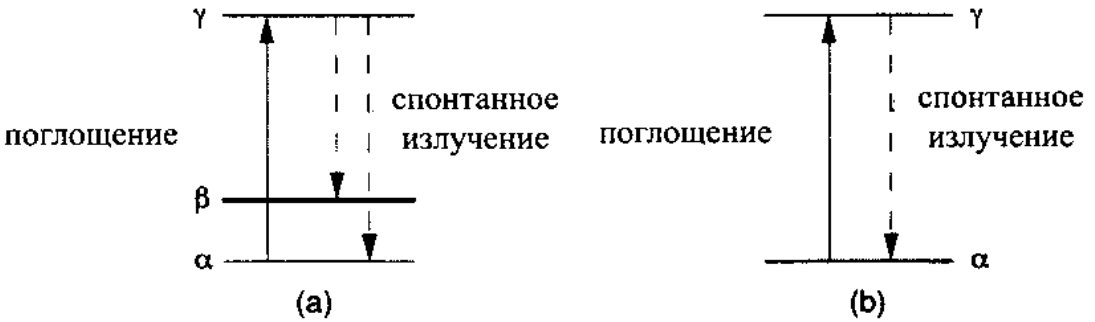


Рис. 6.4. Оптическая подготовка и детектирование состояний. (a) Переходы накачки, (b) циклические переходы

При термодинамическом равновесии уровни α и β населены примерно одинаково, в то время как уровень γ является практически пустым, так что для оптических переходов $\alpha \leftrightarrow \gamma$ мы имеем

$\frac{E_\gamma - E_\alpha}{kT} >> 1$ и, следовательно, $\frac{n_\gamma}{n_\alpha} << 1$,

из (6.12). Излучение, настроенное на этот переход, будет таким образом поглощаться, заставляя атомы переходить с уровня α на уровень γ. Затем через спонтанное излучение атом очень быстро возвращается назад в основное состояние либо на уровень α, либо на уровень β. Вероятность попадания на эти два состояния практически одинакова. После нескольких циклов поглощения и спонтанного излучения атомы таким образом переходят с уровня α на уровень β. Разность населенностей между двумя уровнями основного состояния создается тем более эффективно, чем больше фотонов возбуждает переход $\alpha \rightarrow \gamma$. Лазер, настроенный на оптический переход, является наиболее подходящим источником света для осуществления оптического метода накачки.

Переходы типа $F = 3 \rightarrow F' = 3$ в атоме цезия называются *переходами накачки* из-за правил отбора (раздел 6.1.3), которые позволяют атомам возвращаться в основное состояние, отличающееся от стартового состояния, и обеспечивают таким образом осуществление разности населенностей между этими уровнями.

(c) Оптическое детектирование

Оптические методы могут также использоваться для измерения населенности на данном уровне. Существует несколько возможностей.

(1) Когда уровень α на рис. 6.4а ненаселен, падающий свет не может быть поглощен, и атомная среда остается прозрачной. Она может быть сделана непрозрачной опять, если атомы на уровне β заставить перейти обратно на уровень α с помощью излучения на соответствующей сверхвысокой частоте. Эта возможность используется в рубидиевых часах.

(2) Вернемся опять к рис.6.4а, когда уровень α не населен и падающий свет больше не переводит атомы на уровень γ. Тогда прекращается спонтанное излучение и отсутствует свет флуоресценции. Если осуществить переход $\beta \rightarrow \alpha$ с помощью соответствующей сверхвысокой частоты, то возобновится флуоресценция. При достаточно интенсивном падающем свете среднее число фотонов флуоресценции, излучаемых каждый раз, когда атом возвращается на уровень α, составляет от одного до нескольких фотонов. Оно зависит от значений F и F', связанных с уровнями α и γ. Такой метод детектирования используется в цезиевых часах с оптической накачкой и в часах на ионах ртути.

(3) Давайте предположим, что падающий свет настроен на частоту перехода между уровнем $F = 3$ состояния $^2S_{1/2}$ и уровнем $F' = 2$ состояния

$^2P_{3/2}$ в атоме цезия. Спонтанное излучение может вернуть атом только назад на начальный уровень $F = 3$, потому что переход $F' = 2 \rightarrow F = 4$ запрещен (раздел 6.1.3). В таком случае мы имеем закрытую двухуровневую систему, как показано на рис.6.4b, и осуществляется *циклический переход*. Так как время жизни возбужденного состояния очень мало, атом очень скоро становится опять доступным для участия в следующем цикле. Число флуоресцентных фотонов пропорционально интенсивности падающего света и продолжительности взаимодействия. Если этот метод детектирования используется в цезиевых часах, то он позволяет получить несколько десятков флуоресцентных фотонов на один атом.

6.2.3. Устранение эффекта Доплера

Мы начнем с того, что если не предпринять соответствующих шагов, то эффект Доплера становится серьезным препятствием для спектроскопии сверхвысокого разрешения.

Если источник излучения и прибор, способный детектировать это излучение, находятся в относительном движении с ненулевой компонентой скорости по направлению, соединяющему источник и детектор, то частота сигнала будет казаться измененной. Этот так называемый эффект Доплера обычно наблюдается в распространении звуковых волн.

В случае распространения электромагнитных волн упрощенное выражение для сдвига частоты из-за эффекта Доплера, подразумеваемого в (3.6), следующее:

$$\nu - \nu_0 = \nu_0 \frac{\upsilon}{c} \cos\theta , \tag{6.14}$$

где ν и ν_0 представляют частоту излучения, испускаемого атомом в движении со скоростью υ и атома в покое соответственно и θ – угол между направлением движения атома и направлением излучения. Выражение для сдвига частоты то же самое как для излучающего, так и для поглощающего атома. Мы предполагаем, что $\upsilon/c << 1$. Это условие позволяет рассматривать эффект Доплера в приближении первого порядка по υ/c.

Если атомы собраны в коллимированный пучок, то сдвиг частоты перехода дается непосредственно формулой (6.14). Рассмотрим пучок атомов водорода, производимый источником при температуре 300 К. Если угол между пучком и направлением излучения отличается от перпендикулярного на 1°, то частота сверхтонкого перехода сдвигается примерно на 200 Гц, или в относительном изменении на $1{,}4 \times 10^{-7}$. Такой сдвиг неприемлем для рассматриваемых здесь применений.

В ячейке тепловое возбуждение заставляет вектор скорости атомов принимать случайные, в среднем изотропные, направления. В таком случае

резонанс не имеет сдвига Доплера, но зато он уширен. Форма резонансной кривой определяется максвелловским распределением атомных или молекулярных скоростей. Для типичных температур доплеровская ширина резонанса составляет примерно 10^{-6} от резонансной частоты. Точное значение зависит от массы атома и температуры газа. Другими словами, добротность резонанса не может быть больше 10^6 при наличии такого доплеровского эффекта. Такая добротность слишком мала, чтобы обеспечить точное измерение положения резонансного максимума.

В атомных стандартах частоты, представленных в этой книге, ограничения из-за эффекта Доплера снимаются путем пространственного ограничения области, в которой происходит взаимодействие между атомом и излучением.

(a) Цезиевые часы

Из-за эффекта Доплера первого порядка фаза бегущей волны изменяется в зависимости от положения, а для чистой стоячей электромагнитной волны в области пучности магнитного (или электрического) поля, фаза не зависит от положения. Все, что случается теперь с атомами, находящимися в этой области (в пучности поля), является результатом осцилляций в этом месте, и эффект Доплера первого порядка не проявляется. Такая конфигурация используется в цезиевых часах. Каждое взаимодействие между атомом и магнитным полем сверхвысокой частоты происходит в пучности этого поля, которая принадлежит стоячей волне, поддерживаемой в резонансной полости.

(b) Другие атомные часы

Решения, реализованные в других типах часов с беспорядочным движением атомов, опираются на предложения, сделанное Дике [6.8]. Для того чтобы выявить условия, которые должны выполняться, рассмотрим атом, который, если бы он был в покое, излучал бы волны с частотой ν_0 и длиной волны λ. Для упрощения сделаем следующие предположения:

(1) атом движется вперед и назад перпендикулярно двум стенкам, удаленным на расстояние d;

(2) абсолютное значение скорости атома одно и то же в обоих направлениях;

(3) атомные свойства не изменяются, когда атом вынужден поворачивать обратно от стенки.

Наблюдатель, расположенный по оси движения, принимает излучение с частотой, которая попеременно равна то $\nu_0(1+v/c)$, то $\nu_0(1-v/c)$ благодаря сдвигу из-за эффекта Доплера. В этом случае спектр принимаемого излу-

чения состоит из компоненты на частоте ν_0 и боковых компонент на частотах $\nu_0(1\pm nv/c)$, где n – целое число. Если просуммировать по всему набору атомов, испытывающих такие же движения, как рассмотрено выше, но с разными скоростями, то принимаемые спектральные компоненты складываются вместе на частоте ν_0 и складываются рядом друг с другом, если они принадлежат боковым линиям. Можно показать, что если выполняется условие Дике

$$d < \frac{\lambda}{2}, \tag{6.15}$$

то спектр излучения включает в себя не уширенную эффектом Доплера линию на частоте ν_0, возвышающуюся на пьедестале, сформированном набором боковых линий. Ширина пьедестала такого же порядка по величине, как и обычная доплеровская ширина. Однако она имеет малую амплитуду, так что практически выделяется только центральная линия. Таким образом оказывается возможным наблюдение атомного излучения, свободного от эффекта Доплера. Те же доводы применимы и для атомов, поглощающих излучение. Условие (6.15) для подавления доплеровского уширения справедливо и для бегущих и для стоячих волн. Результат также можно обобщить на трехмерное ограничение.

Для такой бездоплеровской спектроскопии может быть использовано несколько методов.

- *Буферный газ*. Атомы, у которых необходимо наблюдать сверхтонкий переход, растворяются в инертном газе, называемом буферным газом. Когда выполняется условие (6.15), где d есть средняя длина пробега между двумя столкновениями с буферным газом, то доплеровское уширение исключается. Такой метод используется в рубидиевых часах.
- *Ограничение атомов*. В водородном мазере атомы ограничиваются в колбе, где они отскакивают от стенок. В ртутных ионных часах электрическое поле удерживает ионы в ограниченном объеме пространства. В обоих случаях условие (6.15) учитывается в конструкции системы.

Уравнение (6.14) является упрощенным выражением для доплеровского сдвига частоты в приближении первого порядка по v/c. В атомных стандартах частоты второй порядок оказывает значительное влияние. Он происходит от релятивистского замедления времени для атомов, движущихся по отношению к измерительной аппаратуре (раздел 3.2.3). Нет способов подавить или компенсировать этот эффект кроме выбора медленно движущихся атомов или использования эффективного способа для замед-

ления их скорости (раздел 6.4.3). Соответствующий сдвиг частоты $\Delta\nu_0$ дается из

$$\Delta\nu_D = -\frac{1}{2}\left(\frac{\upsilon}{c}\right)^2\nu_0, \text{ или } \Delta\nu_D = -\frac{E_{kin}}{Mc^2}\nu_0, \qquad (6.16)$$

где E_{kin} – кинетическая энергия атома с массой M. Для атома цезия испускаемого из печи с температурой около 100°C, $\upsilon \approx 200 мс^{-1}$. Это означает, что $\Delta\nu_D/\nu_0 \approx -2\times10^{-13}$. Хотя эта величина и очень мала, ею нельзя пренебрегать.

6.2.4. Взаимодействие между атомами и электромагнитным полем

Взаимодействие между атомами и волнами со сверхвысокой частотой обычно происходит в резонансной электромагнитной полости в присутствие слабого однородного статического магнитного поля.

Осциллирующее и статическое магнитные поля параллельны в пределах объема, занимаемого атомами, что позволяет осуществлять селективное наблюдение часового перехода, для которого $\Delta F=\pm1$, $\Delta m_F=0$ (раздел 6.1.3).

Так как значение статического магнитного поля много меньше, чем магнитное поле Земли, то область взаимодействия должна быть заключена в магнитный экран. Он выполняет ещё и вторую роль, значительно ослабляя переменные вариации колебаний внешних магнитных полей.

Действительно, (6.6) показывает, что малые изменения δB_0 в действующим на атомы магнитном поле вызывают изменение резонансной частоты, равное $2K_0B_0\delta B_0$. Например, в цезиевых часах относительное изменение частоты перехода около 3×10^{-14} вызывается относительным изменением $\delta B_0/B_0\times10^{-4}$.

В водородных мазерах и цезиевых часах обычно реализуются коэффициенты экранирования от 10^4 до 10^5. Для того чтобы получить такую эффективность, магнитные экраны выполняются из вложенных в друг друга многослойных оболочек из материалов с очень высокой магнитной проницаемостью. Статическое магнитное поле создается током в катушках, помещенных внутри защитных экранов с целью получения как можно более однородного и стабильного магнитного поля.

6.2.5. Соответствующие электронные системы

Существует два типа атомных часов: *пассивные часы* и *активные часы*.

Пассивными являются цезиевые часы, рубидиевые часы и часы на ионах ртути. Переход возбуждается с помощью электромагнитного сигнала, и чем ближе частота возбуждения находится от резонансной часто-

ты, тем больше получается отклик. Поэтому необходимо создавать сигнал возбуждения на пробной (зондирующей) частоте. Это достигается с помощью кварцевых генераторов и методов синтеза частоты.[3] С помощью цепи обратной связи резонансный сигнал используется для автоматического контроля частоты кварцевого генератора по частоте атомного перехода.

Водородный мазер может использоваться и как пассивные и как активные часы. В последнем случае рабочие условия выбраны так, что он может генерировать излучение на частоте сверхтонкого перехода. Электронная система в таком случае проектируется как супергетеродиный приемник, для того чтобы принять сигналы, генерируемые мазером. К тому же она автоматически контролирует фазу и, следовательно, частоту кварцевого генератора по частоте генерации мазера с помощью цепей, использующихся в цепи синтеза частоты.

В обоих случаях частота атомного перехода сводится к величине, скажем, 10 МГц, которая может быть использована непосредственно потребителем. Делители частоты могут производить импульсы, например каждую секунду, которые используются для маркирования шкалы времени, производимой часами. Электронные приборы обрабатывают сигналы на сверхвысокой частоте и на радиочастоте. В этой области существует стабильный технологический прогресс. Обычно реализуются надежные высококачественные электронные системы для целей точного определения и использования частоты атомного резонанса.

Другие электронные цепи могут выполнять множество различных операций, таких как, например, генерация исключительно стабильного электрического тока, необходимого для создания статического магнитного поля с помощью соленоидов.

6.3. Реализация приближения изолированного атома в покое

Идеальный атомный стандарт частоты был бы таким, в котором опорная частота определялась бы переходом между двумя уровнями, хорошо отделенными от соседних уровней и этот переход происходил бы в одном изолированном атоме или ионе, находящемся в покое.

Такая идеальная ситуация на практике осуществляется только приближенно главным образом из-за технических ограничений. Любое отклонение от идеала вызывает сдвиг частоты перехода от ее невозмущен-

---

[3] Электрические цепи, известные как смесители частоты, умножители и делители используются для того, чтобы применить все четыре арифметические операции к значению частоты. Таким образом, можно произвести сигналы с отношением частот, равным любому рациональному числу.

ного значения. Этот сдвиг может иметь несколько составляющих [6.4, 6.9], например, эффект Зеемана второго порядка или эффект Доплера второго порядка. Однако эти сдвиги частоты очень малы, если их выразить в относительных величинах, и неопределенности в их оценке таковы, что неточности некоторых лабораторных цезиевых часов в 2000 году составляли около 10^{-15}.

Мы теперь только просуммируем физическое происхождение некоторых остаточных сдвигов частоты, указывая вкратце шаги, которые можно было бы предпринять, чтобы их уменьшить. Детальное обсуждение этих сдвигов и описание работ, выполненных на сегодня чтобы приблизить нас к идеальной ситуации, будет далее приведено для каждого типа стандарта частоты.

(1) Два уровня сверхтонкой структуры вырождаются в нулевом поле и можно подумать, что частота перехода может быть сделана нечувствительной к величине B_0 путем простого выполнения этого условия. Однако в случае цезия, например для того чтобы убедиться, что все возможные переходы между подуровнями с состояниями $I+1/2$ и $I-1/2$ не отличаются более чем на 1 Гц, B_0 должно быть меньше 5×10^{-11} Т во всей области, в которой может наблюдаться переход в сверхтонкой структуре. Интенсивности этих переходов должны также проявлять симметричные свойства относительно центрального перехода (4,0)↔(3,0), иначе наблюдаемый резонанс будет искажен, и в результате измерения резонансной частоты появятся практически неприемлемые ошибки. В приборах с разумными размерами и ценой все еще не известно как получить достаточно однородное поле и достаточную нечувствительность к внешним возмущениям при таких малых значениях B_0. Поэтому предпочтительнее расположить атомы в магнитном поле, чтобы отчетливо разделить различные возможные переходы.

Фактически, величина B_0 выбирается из компромисса между двумя противоположными требованиями. Она должна быть достаточно большой, чтобы убедиться, что крылья соседних переходов не могут исказить опорный резонанс до неприемлемой степени. Но она должна быть как можно меньше, чтобы ограничить чувствительность частоты часового перехода к флуктуациям δB_0 в B_0.

Легче уменьшить практически выбранную величину B_0, если соседние резонансы очень узкие и расположены симметрично относительно опорной частоты.

(2) Атомы подвергаются воздействию излучения черного тела. Оно происходит из-за теплового движения носителей зарядов в материалах, окружающих наблюдаемые атомы. Соответствующий сдвиг частоты, близкий к -2×10^{-14} при комнатной температуре, становится значитель-

ным из-за прогресса в современных атомных стандартах частоты. Единственным способом преодоления этого препятствия было бы охлаждение всех частей оборудования. Технические ограничения для решения этой задачи были бы непреодолимыми, по крайней мере, для цезиевых пучковых стандартов.

(3) Атомы находятся в движении. Несмотря на предосторожности, предпринятые для подавления эффекта Доплера первого порядка, он не может быть исключен полностью. Более того, эффект Доплера второго порядка становится неизбежным. Поэтому скорости атомов или ионов должны быть уменьшены. Современный и будущий прогресс в атомных стандартах частоты зависит от нахождения эффективных путей замедления атомов и ионов.

(4) Столкновения между щелочными атомами, или между ними и буферным газом, или между ними и стенками ячейки сдвигают энергии уровней сверхтонкой структуры. Благоприятная конфигурация обеспечивается в атомном пучке, или в пучке ионов низкой плотности, или в атомном облаке. Идеальная ситуация, обсуждаемая в конце этой главы, состоит в том, что можно наблюдать единственный ион.

(5) Селекция состояний с помощью оптической накачки вызывает сдвиг уровней сверхтонкой структуры, который называется световым сдвигом. Этот сдвиг очень существен в рубидиевых часах. Проблема может быть решена путем пространственного и временного разделения взаимодействия с оптической и сверхвысокой частотой, как это делается в цезиевых часах с оптической накачкой, в цезиевом фонтане и в часах на ионах ртути.

(6) Атомы связываются с полем сверхвысокой частоты обычно в резонансной полости Реальная частота атомного перехода имеет небольшую зависимость от настройки частоты резонатора (затягивание частоты резонатором). Соответствующая отстройка частоты пассивных часов значительно меньше чем активных часов (если не приближаться к условиям генерации как в пассивном водородном мазере). Кроме того, затягивание частоты становится меньше, если атомный резонанс сужается.

(7) В пассивных часах определение частоты атомного резонанса чувствительно к спектральной симметрии частотно-модулированного сигнала возбуждения. Должны быть приняты все меры, чтобы убедиться в его симметрии. Более того, в цезиевых и рубидиевых часах некоторые сдвиги частот зависят от амплитуды осциллирующего поля, значение которой по этой причине должно быть стабилизированным.

6.4. Цезиевые пучковые часы

Первый работающий пучковый цезиевый атомный стандарт частоты был построен в 1955 году Эссеном и Пэри в национальной физической лаборатории в Великобритании [6.10]. С тех пор изучением и конструированием цезиевых часов с различной степенью успеха занимались во многих странах, включая Канаду, Китай, Францию, Италию, Японию, Корею, Польшу, Восточную и Западную Германии, Швейцарию, СССР и США.

Цель этих лабораторных приборов состоит в том, чтобы как можно лучше осуществлять физическую реализацию определения секунды. Эти приборы называются *первичными стандартами времени и частоты*. Например, первичные стандарты PTB в Брауншвейге Германия, воспроизводили секунду с неточностью около10^{-14} в течение многих лет [6.11, 6.12, 6.13, 6.14]. Более того, цезиевые часы производятся промышленным способом, и несколько тысяч таких приборов находится в эксплуатации.

До 1990 года все цезиевые часы были построены во многом таким же способом, как и первый прибор Эссена и Пэри, с использованием магнитного отклонения атомов. Технологические совершенствования довели их до уровня очень надежных приборов с очень высокими характеристиками. Наиболее важные инновации за это время были введены в PTB.

Развитие и разработка полупроводниковых лазеров сделали возможным применение оптических методов к пучковым цезиевым часам путем замены магнитной селекции по состояниям на оптическую накачку и детектирование. Первая демонстрация таких часов была проведена в институте фундаментальной электроники в Орсэ, Франция [6.15]. Затем цезиевые часы с оптической накачкой последовательно изучались в нескольких лабораториях, включая Национальное Бюро Стандартов, переименованное теперь в Национальный Институт Стандартов и Технологии (NIST), в Боулдере, Колорадо (США), а также в лаборатории атомного времени в Орсэ, Франция [6.16, 6.17]. В 2002 году несколько лабораторных стандартов с оптической накачкой и детектированием на пучках атомов цезия достигли качества первичных стандартов. Два наиболее точных стандарта находятся в лаборатории первичного атомного времени и частоты (LPTF) в Париже и в NIST [6.18, 6.19].

За последние примерно десять лет во многих лабораториях мира, включая высшую техническую школу (ENS) в Париже, изучались и развивались методы лазерного охлаждения [6.20]. Они были впервые применены в LPTF в экспериментальном первичном стандарте, использующем фонтан медленных атомов цезия. Была достигнута настолько низкая неточность, что она составила примерно 10^{-15}, и это является лучшим результатом, который когда-либо был получен, [6.21, 6.22, 6.23]. Ясно, что в будущем лучшие цезиевые часы будут использовать атомы, охлажденные лазером.

В рамках этой книги невозможно описать ни все возможные варианты цезиевых пучковых стандартов частоты, ни все идеи, которые использовались для их дальнейшего совершенствования. Поэтому мы обсудим принципы работы и главные особенности по отношению к промышленным моделям, т.е. к коммерческим приборам с магнитным отклонением. Мы укажем характерные параметры и характеристики лабораторных приборов, основной задачей которых является воспроизведение физической реализации секунды настолько точно, насколько это возможно. Мы опишем цезиевые часы с оптической накачкой и сравним их с приборами с магнитным отклонением. Наконец, мы опишем способы, как можно охладить атомы через лазерное переизлучение, отмечая, что преимущество использования холодных атомов приводит к весьма впечатляющим характеристикам атомных стандартов частоты.

6.4.1. Цезиевые часы с магнитным отклонением пучка

Рис. 6.5 показывает схематически структуру цезиевой лучевой трубки с магнитным отклонением атомных траекторий. Для упрощения мы выбрали обычнейшую и наиболее основную модификацию, характерную для промышленных приборов, используемых в настоящее время для земных и космических применений.

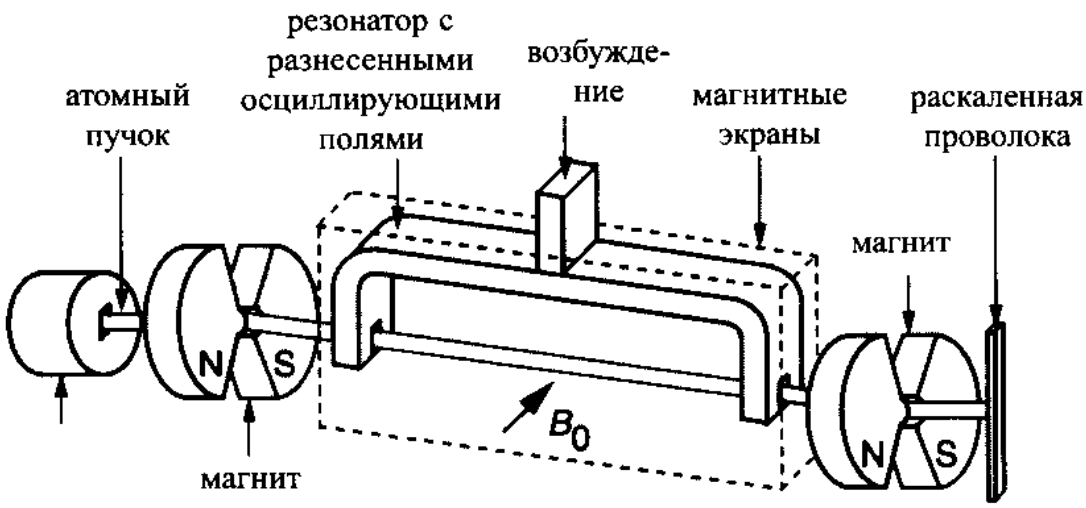


Рис. 6.5.Схематическое изображение цезиевой трубки с магнитным отклонением пучка. Масс-спектрометр, электронный умножитель и некоторые другие компоненты не показаны. Угол отклонения сильно преувеличен

Печь, содержащая несколько граммов Cs 133, единственного стабильного изотопа этого элемента, нагревается примерно до 100°C. Система капилляров формирует направленный пучок атомов цезия, в котором имеется шестнадцать равнонаселенных подуровней сверхтонкой структуры. Пучок проходит *первый магнит*, производящий сильное неодно-

родное магнитное поле. При прохождении этого поля атомы подвергаются действию сил, перпендикулярных плоскости рисунка (см. раздел 6.2.2). Часть атомов отклоняется в направлении оси прибора. Эта часть включает в себя одно из $m_F = 0$ состояний, обычно состояние $F = 3$, $m_F = 0$. Угол отклонения составляет около 1°. Траектории остальных атомов, включая атомы в состоянии $F = 4$, $m_F = 0$, наклонены по отношению к этой оси. Это не показано на рис.6.5. Соответствующие атомы, в конечном счете, ударяются в стенку, которая поглощает цезий.

Пучок затем пересекает *сверхвысокочастотную область взаимодействия*, которая включает в себя резонатор с двумя плечами. Атомы последовательно подвергаются действию двух магнитных полей на сверхвысоких частотах (СВЧ), осциллирующих на частоте 9,192...ГГц, и разделенных областью с отсутствующей осцилляцией. Эта конфигурация была предложена Рэмси [6.24]. Расстояние L между двумя осциллирующими полями может выбираться свободно и оно определяет эффективное время взаимодействия Δt в (6.8). L составляет около одного метра в лабораторных первичных стандартах, в то время как для промышленных приборов его величина лежит вблизи 15 см. Вся область взаимодействия имеет магнитные экраны для того, чтобы защитить ее от внешних магнитных полей и их изменений. В этой области создается однородное и стабильное *магнитное поле*. Его индукция составляет около одной десятой от магнитной индукции земного поля. (Это магнитное поле иногда называется С-поле, в то время как первое и второе поля для селекции состояний относятся к магниту А и магниту В соответственно.) Если частота вводимой в резонатор электромагнитной волны находится в резонансе с часовым переходом, то при пролете через область взаимодействия, атомы меняют сверхтонкий подуровень и переходят с подуровня $F = 3$, $m_F = 0$ на подуровень $F = 4$, $m_F = 0$.

На рис.6.6 показана схематически структура каждой области осциллирующего поля в точке, где атомный пучок проходит плечо резонансной полости. Плечо закрывается короткозамкнутой цепью, и магнитное поле сверхвысокой частоты действует на атомы в области одной пучности стоячей волны, поддерживаемой в резонаторе. Таким образом подавляется эффект Доплера первого порядка (до малой остаточной величины, раздел 6.2.3, раздел 6.4.1e). В области, общей для атомного пучка и плеча резонатора, осциллирующее магнитное поле направлено параллельно статическому магнитному полю, чтобы удовлетворить правилам отбора для переходов $\Delta F = \pm 1$, $\Delta m_F = 0$. По оси атомного пучка на концах каждого плеча резонатора установлены волноводы с малым сечением (запредельные волноводы). Они предотвращают проникновение электромагнитного излучения через отверстия в резонаторе, чтобы не создавалась бегущая волна, параллельная атомному пучку.

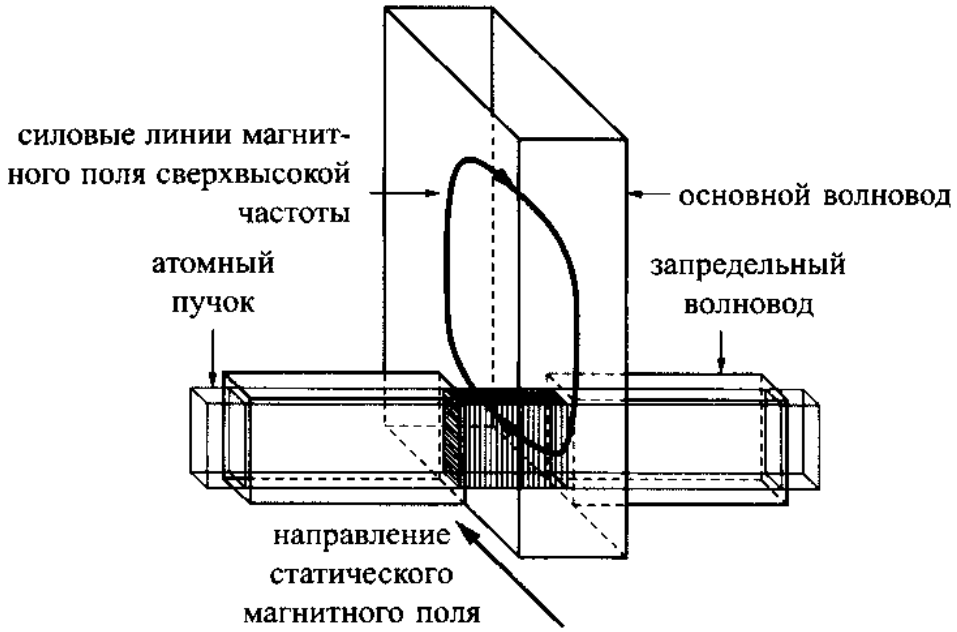


Рис. 6.6. Структура в области осциллирующего поля. Взаимодействие с атомами происходит в затененной области

Второй селектирующий магнит, похожий на первый, отклоняет атомы, совершившие сверхтонкий переход в направлении *детектора на поверхностной ионизации*. Он состоит из металлической проволоки, имеющей работу выхода больше, чем энергия ионизации цезия. Когда проволока нагревается примерно до 1000°C, падающие на нее атомы цезия испаряются затем с нее в виде ионов цезия Cs^+. В лабораторных цезиевых часах ионный ток имеет величину порядка 1 – 10 пА и может быть измерен непосредственно электрометром. В промышленных приборах значение ионного тока в десятки раз меньше. Ионный поток затем фильтруется с помощью масс-спектрометра, для того чтобы исключить нежелательные ионы, излучаемые горячей проволокой. Ионы цезия Cs^+ далее попадают на динод, где они вызывают вторичную эмиссию электронов. Поток этих электронов усиливается далее электронным умножителем.

Полученный таким образом электрический ток составляет выходной сигнал цезиевой атомно-лучевой трубки. Он пропорционален вероятности того, что атомы совершили переход $F = 3, m_F = 0 \rightarrow F = 4, m_F = 0$.

(b) Частотный отклик цезиевой лучевой трубки

Цезиевая лучевая трубка ведет себя как резонатор, в котором выходной ток зависит от частоты действующего сигнала. Общая форма резонансной кривой показана на рис.6.7. Мы видим, что она имеет узкую центральную часть, называемую интерференционными *осцилляциями Рэмси* (*Ramsey fringe*) и расположенную в центре более широкой структуры, известной как *пьедестал Раби*. (Мы рассматриваем здесь только наиболее общую ситуацию, когда два осциллирующих поля находятся в фазе. Рэмси резонатор может

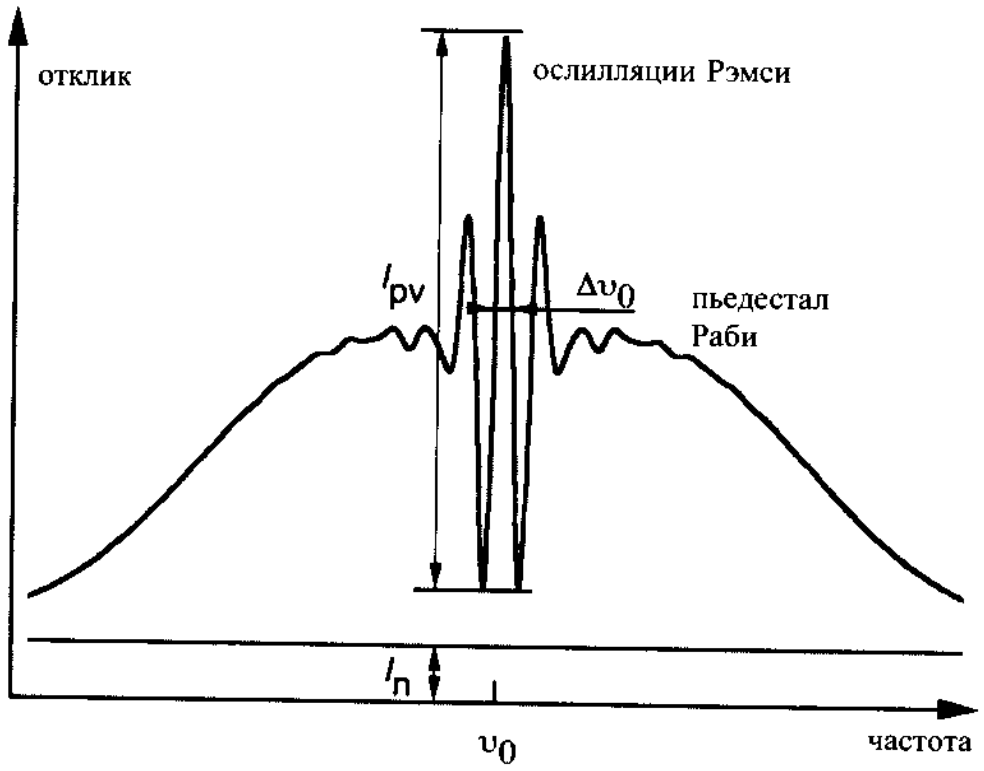


Рис. 6.7. Форма резонансной кривой для цезиевой лучевой трубки

быть сделан так, что два осциллирующих поля находятся совершенно не в фазе. В этом случае центральная часть инвертируется и имеет на резонансе минимум. Общая форма пьедестала Раби не изменяется).

Центральный пик является полезной частью резонансной кривой для автоматической подстроки кварцевого генератора по частоте атомного перехода. Его ширина определяется временем пролета атомов между двумя осциллирующими полями. Она составляет около 500 Гц для промышленно изготовленных цезиевых лучевых трубок. Она может быть настолько малой, что составляет 1 Гц для лабораторных установок, использующих очень медленные атомы с лазерным охлаждением (раздел 6.4.3b). Ширина на половине высоты для пьедестала Раби зависит от времени пересечения осциллирующего поля. Она близка к 20 или 25 кГц для упомянутых выше модификаций, в которых атомы испускаются печью.

Давайте начнем с приведения упрощенных алгебраических выражений для узкой части резонансной кривой. Мы подтвердим это упрощенное выражение теоретически и установим полную формулу, в которой принимается во внимание и центральная часть и крылья резонансной кривой. Вывод предполагает знание основ квантовой механики и при первом чтении может быть опущен.

Формула для центральной части резонансной кривой. Как мы покажем ниже, выражение для вероятности $P(\tau)$, того что произошел часовой

переход, может быть упрощено, если мы будем интересоваться только областью в непосредственной близости от резонансной частоты. Эта вероятность дается следующим выражением:

$$P(\tau) = \frac{1}{2}\sin^2 b\tau[1 + \cos(\omega - \omega_0)T], \tag{6.17}$$

где τ – время прохождения атомом через каждое осциллирующее поле и T – время пролета между этими полями. Величины ω и ω_0 представляют собой угловые частоты осциллирующего поля и атомного перехода соответственно. Вероятность $P(\tau)$ зависит от амплитуды осциллирующего поля через параметр b. Этот параметр связан с магнитным полем сверхвысокой частоты B' через

$$b = \frac{\mu_B B'}{\hbar}, \tag{6.18}$$

где μ_B – магнетон Бора. Величина b, часто называемая частотой Раби, является угловой частотой. Она характеризует скорость, с которой осциллирующее поле изменяет квантовые состояния атома.

Уравнение (6.17) показывает, что вероятность сверхтонкого перехода, происходящего в цезиевой лучевой трубке, имеет максимум, когда $\omega = \omega_0$ и когда ширина на полувысоте резонансной кривой для моноскоростного пучка в соответствии с соотношениями неопределенностей (6.8) равна

$$\Delta\nu_0 = \frac{1}{2T}. \tag{6.19}$$

Более того, член $\sin^2 b\tau$ показывает, что эта вероятность оптимальна, когда амплитуда осциллирующего поля удовлетворяет условию $b\tau = \pi/2$ в предположении, что все атомы в пучке имеют одинаковые скорости.

Упрощенное выражение (6.17) справедливо при условии, что выполняется неравенство

$$|\omega - \omega_0| << b. \tag{6.20}$$

Оно может быть написано в виде

$$\frac{|\nu - \nu_0|}{\Delta\nu_0} << \frac{L}{l}\frac{b\tau}{\pi}, \tag{6.21}$$

где l и L обозначают длины каждого из осциллирующих полей и расстоянием между ними соответственно. Так как на практике обычно мы имеем $b\tau \sim \pi/2$ и $L/l > 10$, то приближение адекватно тому, чтобы описать центральную часть резонансной кривой. Более того, оно полезно для упрощенного анализа свойств цезиевого пучкового резонатора.

В реальности скорости атомов распределены вокруг средней скорости. В случае цезиевой лучевой трубки с магнитным отклонением пучка распределение по скоростям отличается от того распределения, которое производится цезиевой печью. Это происходит из-за того, что угол отклонения в двух селектирующих магнитах сильно зависит от скорости. Пусть $f(\tau)$ есть нормированное распределение времени взаимодействия. Упрощенное выражение для средней вероятности, того что атом испытает переход, равно

$$P = \frac{1}{2}\int_0^\infty f(\tau)\sin^2 b\tau[1+\cos(\omega-\omega_0)T]\,\mathrm{d}\tau\,. \qquad (6.22)$$

Распределение скоростей приводит к дисперсии в положении различных максимумов функции $\cos(\omega-\omega_0)T$, кроме случая когда $\omega=\omega_0$. Дисперсия возрастает, если мы удаляемся от резонансной частоты, и это уменьшает амплитуду боковых линий. Для $|\nu-\nu_0|$ много больше, чем ширина центральной линии, отклик цезиевой атомно-лучевой трубки прописывается монотонно меняющимся профилем, симметричным относительно ν_0. Для того чтобы интерпретировать форму пьедестала Раби мы должны рассматривать неупрощенное выражение для $P(\tau)$, которое будет установлено в следующем разделе.

Выходной ток трубки I является линейной функцией от P. Он может также содержать вклад I_n например от нежелательных атомов цезия (фон). Таким образом, вблизи резонанса мы имеем

$$I = I_n + \frac{I_0}{2}\int_0^\infty f(\tau)\sin^2 b\tau[1+\cos(\omega-\omega_0)T]\,\mathrm{d}\tau\,, \qquad (6.23)$$

где I_0 является константой.

Вычисление кривой отклика цезиевой лучевой трубки. Взаимодействие между атомами и электромагнитным полем может быть истолковано строго только в рамках квантовой механики [6.25]. Мы будем использовать основные элементы этой теории, чтобы установить главные шаги в промежуточных вычислениях.

В обозначениях Дирака квантовое состояние атома характеризуется вектором состояний $|\Psi\rangle$, принадлежащим пространству состояний системы. Пусть $|\Psi_1\rangle$ и $|\Psi_2\rangle$ являются векторами состояний, образующих ортогональный базис для этого пространства. Они позволяют нам представить квантовое состояние атома и его развитие во времени. Положив

$$|\Psi\rangle = c_1|\Psi_1\rangle + c_2|\Psi_2\rangle\,, \qquad (6.24)$$

где c_1 и c_2 – амплитуды вероятностей перехода в состояния $|\Psi_1\rangle$ и $|\Psi_2\rangle$ соответственно. Обе амплитуды в общем зависят от времени.

Базисные векторы $|\Psi_1\rangle$ и $|\Psi_2\rangle$ выбираются так, чтобы они были собственными векторами невозмущенного гамильтониана. Энергии, соответствующие собственным векторам $|\Psi_1\rangle$ и $|\Psi_2\rangle$, есть E_1 и E_2 соответственно, и $E_2 > E_1$. Эти значения энергий являются собственными значениями невозмущенного гамильтониана, который мы рассматриваем. Поэтому матрица $\mathcal{H}_0$, соответствующая этому оператору, представлена в этом базисе с матричными элементами E_1 и E_2 по диагонали (и остальные элементы равны нулю). Так как энергия определяется только до произвольной постоянной, то мы можем добавить величину $-(E_1+E_2)/2$ такую, чтобы

$$\mathcal{H}_0 = \frac{\hbar}{2}\begin{pmatrix} -\omega_0 & 0 \\ 0 & \omega_0 \end{pmatrix}, \tag{6.25}$$

где $\omega_0 = (E_2 - E_1)/\hbar$ является угловой частотой перехода.

Можно показать, что для рассматриваемого сверхтонкого перехода матрица $\mathcal{H}_1$, представляющая возмущение к гамильтониану, дается

$$\mathcal{H}_1 = \frac{\hbar}{2}\begin{pmatrix} 0 & 2b\cos\omega t \\ 2b\cos\omega t & 0 \end{pmatrix}, \tag{6.26}$$

где параметр b определяется из (6.18). Он предполагается постоянным вдоль атомной траектории, т.е. в каждом плече резонансной полости, и, следовательно, на периоде τ взаимодействия с каждым из двух осциллирующих полей.

Матрица $\mathcal{H}$, представляющая общий гамильтониан, находится сразу же путем сложения вместе (6.25) и (6.26). Уравнение Шредингера

$$\mathrm{i}\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\Psi\rangle = \mathcal{H}|\Psi\rangle, \tag{6.27}$$

ведет к

$$\left.\begin{aligned} \mathrm{i}\frac{\mathrm{d}c_1}{\mathrm{d}t} &= -\frac{\omega_0}{2}c_1 + \frac{b}{2}[\exp(\mathrm{i}\omega t) + \exp(-\mathrm{i}\omega t)]c_2 \\ \mathrm{i}\frac{\mathrm{d}c_2}{\mathrm{d}t} &= \frac{b}{2}[\exp(-\mathrm{i}\omega t) + \exp(\mathrm{i}\omega t)]c_1 + \frac{\omega_0}{2}c_2 \end{aligned}\right\}. \tag{6.28}$$

Атомные состояния зависят от времени наблюдения t и от времени θ, пройденного со времени приложения первого возмущения. Следовательно, полная производная по времени есть

$$\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\partial}{\partial t} + \frac{\partial}{\partial\theta},$$

и мы ищем решение в таком виде, в котором эффекты от двух переменных разделяются:

$$\left.\begin{aligned} c_1(t,\theta) &= \gamma_1(\theta)\exp\frac{\mathrm{i}\omega t}{2} \\ c_2(t,\theta) &= \gamma_2(\theta)\exp\frac{\mathrm{i}\omega t}{2} \end{aligned}\right\}. \tag{6.29}$$

Это приводит к

$$\left.\begin{aligned} \mathrm{i}\frac{\partial\gamma_1}{\partial\theta} &= \frac{\Omega_0}{2}\gamma_1 + \frac{b}{2}\gamma_2 \\ \mathrm{i}\frac{\partial\gamma_2}{\partial\theta} &= \frac{b}{2}\gamma_1 - \frac{\Omega_0}{2}\gamma_2 \end{aligned}\right\}, \tag{6.30}$$

где

$$\Omega_0 = \omega - \omega_0. \tag{6.31}$$

В (6.30), используя стандартное приближение, мы опустили члены с угловой частотой 2ω. Действительно, можно показать, что они приводят к появлению в решении очень быстро меняющихся компонент с очень малыми амплитудами. Поэтому уравнения (6.30) образуют систему линейных уравнений с постоянными коэффициентами. При пролете через каждую зону, которую мы должны рассмотреть, изменение атомного может быть представлено в матричной форме состояния следующим образом:

$$\begin{pmatrix} \gamma_1(\textit{конечный}) \\ \gamma_2(\textit{конечный}) \end{pmatrix} = M^k \begin{pmatrix} \gamma_1(\textit{начальный}) \\ \gamma_2(\textit{начальный}) \end{pmatrix}, \tag{6.32}$$

где k – номера зон в том порядке, в котором они пересекаются атомом. Первая и третья зоны являются первым и вторым осциллирующим полями, пересекаемыми атомами соответственно. Вторая зона – это промежуточное пространство без осциллирующего поля (пролетное пространство).

Рассмотрим, например, соотношения между значениями c_2 и γ_2 на выходе третьей зоны. Если t_0 есть время входа в первую зону, то атом покидает третью зону во время $t_0+2\tau+T$. Пусть $\gamma_2(\tau, T, \tau)$ является величиной γ_2, полученной после последовательного прохождения через три зоны. Можно сразу же показать из (6.29) и (6.32), что

$$c_2(t_0+2\tau+T, 2\tau+T) = \gamma_2(\tau,T,\tau)\exp-\frac{\mathrm{i}\omega(t_0+2\tau+T)}{2}. \tag{6.33}$$

Населенность энергетического уровня E_2 является квадратом модуля c_2. На выходе из третьей зоны она равна $|\gamma_2(\tau, T, \tau)|^2$. Следовательно, нам необходимо знать только значения γ_2. То же самое относиться и к γ_1.

Давайте покажем интерференционное явление, которое приводит к периодической зависимости отклика моноскоростной атомно-лучевой цезиевой трубки от $\omega - \omega_0$. Матрица $M^{(2)}$, полученная при b=0 в (6.30), является простой.

$$M^{(2)} = \begin{pmatrix} \exp -\dfrac{i\Omega_0 T}{2} & 0 \\ 0 & \exp \dfrac{i\Omega_0 T}{2} \end{pmatrix}. \tag{6.34}$$

Зоны 1 и 3 идентичны и соответствующие матрицы равны. На момент поинтересуемся их общей формой. Для k = 1, 3,

$$M^{(k)} = \begin{pmatrix} \Gamma_{11}^{(k)} & \Gamma_{12}^{(k)} \\ \Gamma_{21}^{(k)} & \Gamma_{22}^{(k)} \end{pmatrix}. \tag{6.35}$$

Давайте предположим, что атом находится в состоянии с энергией E_1, когда он входит в первое осциллирующее поле. Тогда мы имеем $\gamma_1(0) = 1$ и $\gamma_2(0) = 0$. Квантовое состояние атома на выходе третьей зоны дается

$$\begin{pmatrix} \gamma_1(\tau, T, \tau) \\ \gamma_2(\tau, T, \tau) \end{pmatrix} = M^{(3)} M^{(2)} M^{(1)} \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{6.36}$$

В этом случае модуль квадрата $\gamma_2(\tau, T, \tau)$ равен вероятности того, что переход произошел. Так как пролетное время T пропорционально τ, то мы примем

$$P(\tau) = \left|\gamma_2(\tau, T, \tau)\right|^2, \tag{6.37}$$

что дает

$$P(\tau) = \left|\Gamma_{11}^{(1)}\Gamma_{21}^{(3)} \exp -\frac{i\Omega_0 T}{2} + \Gamma_{22}^{(3)}\Gamma_{21}^{(1)} \exp \frac{i\Omega_0 T}{2}\right|^2. \tag{6.38}$$

В первом члене с правой стороны $\Gamma_{11}^{(1)}$ и $\exp(-i\Omega_0 T/2)$ определяют вариацию амплитуды вероятности для занятия энергетического уровня E_1 в зонах 1 и 2 соответственно. Коэффициент $\Gamma_{21}^{(3)}$ представляет амплитуду вероятности перехода с уровня E_1 на уровень E_2 в зоне 3. Первый член таким образом описывает изменение квантовых свойств атома, стартовавшего на уровне E_1 оставшегося на этом уровне в зонах 1 и 2, и затем совершившего переход на уровень E_2 в зоне 3. Таким же образом второй член представляет изменение в квантовых свойствах атома, стартовавшего на уровне E_1, совершившего переход с уровня E_1 на уровень E_2 в зоне 1, и оставшегося на этом уровне в зоне 2 и 3.

Эти комментарии показывают, что существуют два пути движения из начального к конечному состоянию. Амплитуда вероятности занятия конечного состояния равна сумме частных амплитуд вероятностей, связанных с этими двумя «путями». Так как фазы двух амплитуд вероятности как функции от Ω_0 меняются разным способом вдоль этих двух «путей», то между ними происходит интерференция. Она представляется перекрестным членом, когда мы вычисляем квадраты модулей суммы двух частных амплитуд. Отсюда происходят периодические вариации $P(\tau)$ в Ω_0.

Давайте теперь выпишем элементы матриц $\mathcal{M}^{(1)}$ и $\mathcal{M}^{(3)}$. Когда выполняется предположение (6.20), мы можем положить $\Omega_0=0$ в (6.30), что дает

$$\mathcal{M}^{(1)} = \mathcal{M}^{(3)} = \begin{pmatrix} \cos\dfrac{b\tau}{2} & -\mathrm{i}\sin\dfrac{b\tau}{2} \\ -\mathrm{i}\sin\dfrac{b\tau}{2} & \cos\dfrac{b\tau}{2} \end{pmatrix}. \tag{6.39}$$

При введении вышеуказанных матриц было показано, что величина $b\tau$, выражаемая в единицах угла, характеризует изменение амплитуды вероятности занятия уровней в каждом из осциллирующих полей.

Упрощенная форма для вероятности того, что совершится переход в моноскоростном пучке атомов цезиевой трубки, получается из (6.37- 6.39). В результате имеем

$$P(\tau) = \frac{1}{2}\sin^2 b\tau(1+\cos\Omega_0 T)\cdot \tag{6.40}$$

Наконец, для общего случая произвольной Ω_0 по отношению к b, решение системы дифференциальных уравнений (6.30) дает

$$\mathcal{M}^{(1)} = \mathcal{M}^{(3)} = \begin{pmatrix} \cos\dfrac{\Omega\tau}{2} - \mathrm{i}\dfrac{\Omega_0}{\Omega}\sin\dfrac{\Omega\tau}{2} & -\dfrac{\mathrm{i}b}{\Omega}\sin\dfrac{\Omega\tau}{2} \\ -\dfrac{\mathrm{i}b}{\Omega}\sin\dfrac{\Omega\tau}{2} & \cos\dfrac{\Omega\tau}{2} + \mathrm{i}\dfrac{\Omega_0}{\Omega}\sin\dfrac{\Omega\tau}{2} \end{pmatrix}, \tag{6.41}$$

где

$$\Omega^2 = \Omega_0^2 + b^2. \tag{6.42}$$

Для общего выражения вероятности $P(\tau)$ получим тогда [6.6]

$$P(\tau) = 4\frac{b^2}{\Omega^2}\sin^2\frac{\Omega\tau}{2}\left(\cos\frac{\Omega\tau}{2}\cos\frac{\Omega_0 T}{2} - \frac{\Omega_0}{\Omega}\sin\frac{\Omega\tau}{2}\sin\frac{\Omega_0 T}{2}\right)^2. \tag{6.43}$$

Для данного распределения атомов по скоростям форма резонансной кривой находится путем умножения $P(\tau)$ на функцию распределения времени взаимодействия $f(\tau)$ и последующим интегрированием по всем возможным значениям τ.

(c) Управление частотой кварцевого генератора с помощью атомного резонанса

Коль скоро цезиевая лучевая трубка выполняет роль резонатора, то цезиевые пучковые часы выполняют роль пассивного атомного стандарта частоты. Поэтому необходимо обеспечить присутствие сигнала, возбуждающего атомный резонанс. В дополнение следует заметить, что так как резонансная кривая имеет экстремум на резонансной частоте, то трудно указать точно на эту частоту. Мы должны получить сигнал, пропорциональный расстройке между возбуждающей частотой и частотой резонанса. Именно этот сигнал делает возможным захват частоты кварцевого генератора по атомному резонансу.

Сигнал сверхвысокой частоты 9,192...ГГц, питающий резонансную полость, генерируется системой синтеза из частоты кварцевого генератора, которая часто выбирается равной 10 МГц. Синтезатор может модулировать частоту, генерируя значение частоты около $\nu_0 = \nu_{Cs} + \Delta\nu_Z$, где ν_{Cs} – частота сверхтонкого перехода атома цезия в основном состоянии, т.е. 9 192 631 770 Гц, и $\Delta\nu_Z$ – сдвиг частоты из-за эффекта Зеемана второго порядка, близкий к 1,5 Гц. Значение $\Delta\nu_Z$ определяется из (6.6) величиной магнитного поля, действующего на атомы. В лабораторных первичных стандартах сигнал опроса частоты модулируется примерно до величины $\nu_0 = \nu_{Cs} + \Delta\nu_Z + \Delta\nu'$, где $\Delta\nu'$ – алгебраическая сумма сдвигов частоты, иных, чем сдвиг из-за эффекта Зеемана второго порядка (раздел 6.4.1e).

Резонансная частота находится совершенно стандартным методом, как показано схематически на рис.6.8. Сигнал сверхвысокой частоты периодически модулируется согласно

$$\nu(t) = \nu_i + \nu_m g(t), \tag{6.44}$$

где глубина модуляции представлена значением ν_m, а среднее значение частоты ν_1. Функция $g(t)$ представляет форму модулирующего сигнала и является периодической функцией с частотой ν_M и с нулевым средним значением,. Пусть $g(t)$ является меандром со значением +1 в течение одного полупериода модуляции и –1 в течение второго.

Переходные процессы, происходящие при изменении значения действующей частоты, не будут приниматься во внимание. Они кратковременны, происходят на основной частоте $2\nu_M$ и их влияние может быть исключено при обработке отклика цезиевой лучевой трубки. Рис. 6.9 иллюстрирует отклик резонатора на такую модуляцию частоты, когда $\nu_i \neq \nu_0$.

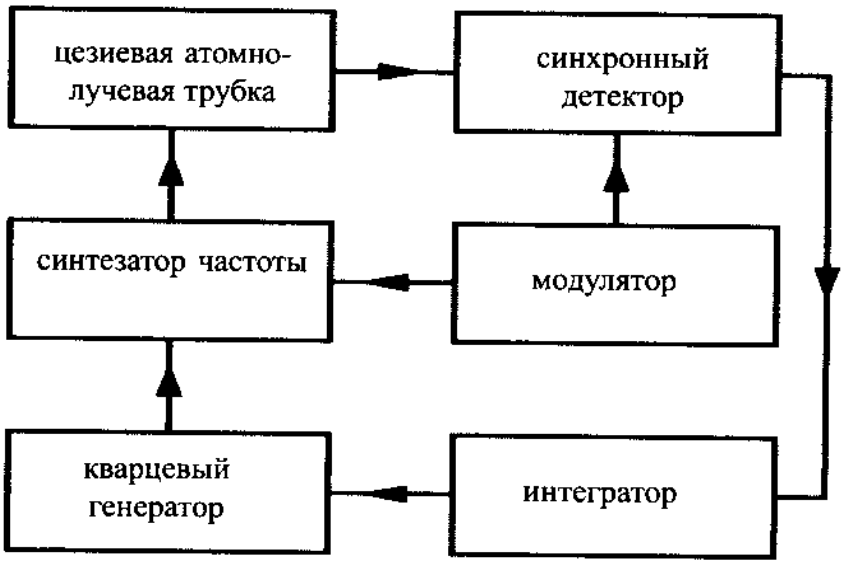


Рис.6.8. Блок-схема электронной системы автоматической подстройки кварцевого генератора по частоте атомного резонанса.

Переменная составляющая в отклике цезиевой лучевой трубки на модуляцию пропорциональна $\nu_i - \nu_0$, когда эта расстройка мала по сравнению с шириной резонанса. Она исчезает при $\nu_i - \nu_0$ равной нулю, и ее фаза изменяется на π, когда расстройка меняет знак. Эта компонента усиливается и демодулируется синхронно с сигналом модуляции. Затем она умножается на функцию $g(t)$. Среднее значение полученного таким образом сигнала составляет *сигнал ошибки E*. Используя упрощенное выражение (6.23) для формы резонансной кривой и полагая что $\nu_i - \nu_0$ много меньше, чем ширина резонансной кривой, мы получим

$$E = KI_0(\nu_i - \nu_0)\int_0^\infty Tf(\tau)\sin^2 b\tau \sin(2\pi\nu_m T)\mathrm{d}\tau, \qquad (6.45)$$

где K – константа. Сигнал ошибки равен нулю, когда $\nu_i = \nu_0$, и он пропорционален отстройке частоты возбуждения ν_i от резонансной частоты ν_0, если эта отстройка достаточно мала. Наклон сигнала ошибки пропорционален производной $\partial P/\partial\omega$ от вероятности P, определяемой из (6.22) для $\omega - \omega_0 = \pm\omega_m$. Он зависит от пролетного времени T и возрастает, если ширина резонанса убывает, или другими словами, возрастает добротность. Наклон также зависит от распределения времен взаимодействия $f(\tau)$, от амплитуды сигнала сверхвысокой частоты, представляемой через b, и от глубины модуляции частоты ν_m. Оптимальные значения для этих параметров следующие: $b\tau \sim p/2$ и $\nu_m \sim 1/4T$. Последнее условие предполагает, что глубина модуляции должна быть выбрана близкой к полуширине резонансной кривой. В этом случае абсолютное значение наклона производной от резонанса $\delta P/\delta\omega$ достигает максимума.

Частотный дискриминатор настраивается по модуляции частоты возбуждения и синхронного детектирования отклика от цезиевой лучевой

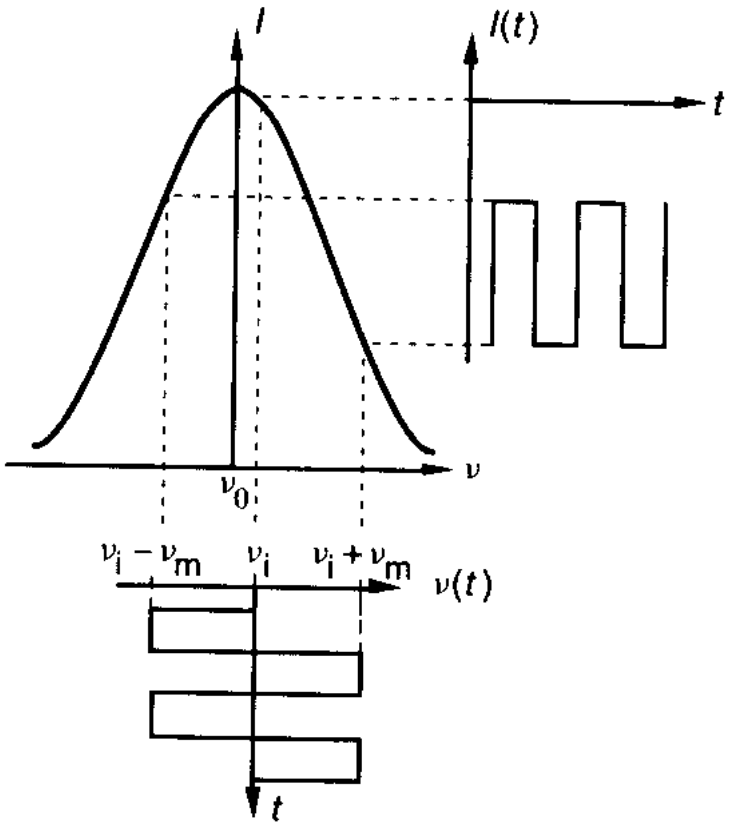


Рис. 6.9. Отклик резонатора на возбуждение, частота которого модулируется меандром. Предполагается, что $\nu_i \neq \nu_0$

трубки. Затем непосредственно осуществляется автоматическая подстройка кварцевого генератора таким путем, чтобы выполнялось условие $\nu_i = \nu_0$. Его частота связана с частотой атомного резонанса в постоянной пропорции. Эта пропорция обозначается через $1/K_s$ и является некоторой рациональной дробью, зависящей только от архитектуры синтезатора частоты, входящего в цепь автоматической подстройки.

Отклик атомного резонатора обрабатывается с помощью цифровых электронных методов. Однако основная особенность частотной автоподстройки должна быть описана с помощью эквивалентной аналоговой функциональной диаграммы, показанной на рис.6.10. Обозначения ΔV или $\Delta V(p)$ например, относятся к преобразованиям Лапласа для отклонения величины $v(t)$ от ее равновесного значения. Сигнал ошибки, производимый синхронным детектором, может быть записан в виде:

$$\Delta E(p) = K_d[\Delta\Omega_r(p) - K_s\Delta\Omega_{sc}(p)], \tag{6.46}$$

где K_d – константа, $\Delta\Omega_r(p)$ представляет изменение в угловой частоте опорного стандарта, а $\Delta\Omega_{sc}(p)$ представляет то же самое для кварцевого генератора, когда он находится в режиме автоматической подстройки. Если $G(p)$ является коэффициентом усиления операционного фильтра, то из-

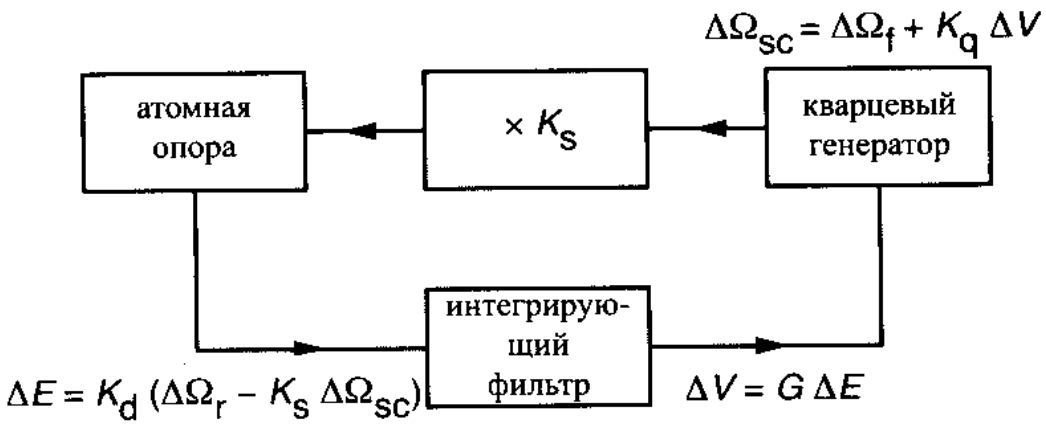


Рис. 6.10. Блок-схема автоподстройки частоты кварцевого генератора по пассивной атомной опоре

менения на его выходе приведут к выходному напряжению, определяемому из:

$$\Delta V(p) = G(p)\Delta E(p). \tag{6.47}$$

Это напряжение подается на варактор, используемый для настройки частоты кварцевого генератора таки образом, что мы имеем:

$$\Delta\Omega_{sc}(p) = \Delta\Omega_f(p) + K_q \Delta V(p), \tag{6.48}$$

где K_q – константа и $\Delta\Omega_f$ – представляет изменения частоты свободного кварцевого генератора, т.е. кварцевого генератора с разомкнутой петлей обратной связи.

Ограничиваясь простым случаем, когда операционный фильтр объединен с усилителем за которым следует идеальный интегратор, мы имеем

$$G(p) = \frac{A}{RCp}, \tag{6.49}$$

где A – константа и RC – постоянная времени интегрирующей цепи. Тогда из вышеприведенных уравнений мы получим:

$$\Delta\Omega_{sc}(p) = \frac{\mathcal{T}p}{1+\mathcal{T}p}\Delta\Omega_f(p) + \frac{1}{1+\mathcal{T}p}\frac{\Delta\Omega_r(p)}{K_s}, \tag{6.50}$$

где $\mathcal{T}$ – постоянная времени отклика петли обратной связи на возмущение. Она равна

$$\mathcal{T} = \frac{RC}{AK_d K_s K_q}. \tag{6.51}$$

На практике ее значение составляет от одной до нескольких секунд.

Если мы рассматриваем случайные частотные флуктуации, то спектральная плотность S_y этих относительных флуктуаций (раздел 5.2.4) дается

$$S_{y,sc}(f) = \frac{(2\pi f \mathcal{T})^2}{1+(2\pi f \mathcal{T})^2} S_{y,f}(f) + \frac{1}{1+(2\pi f \mathcal{T})^2} S_{y,r}(f), \qquad (6.52)$$

где f – частотная фурье-компонента и индексы sc и f обозначают то же самое, что и прежде.

Передаточные функции, появляющиеся с правой стороны в (6.50) ведут себя: – первая как фильтр верхних частот; и вторая – как фильтр нижних частот. Частота среза этих фильтров равна $1/2\pi\mathcal{T}$. Сущность передаточных функций сохраняется, если структура операционного фильтра будет более изощренная.

Как можно видеть из (6.50) и (6.52) вариации частоты генератора, охваченного цепью обратной связи, состоят из двух частей: – высокочастотные компоненты частотных вариаций свободного кварцевого генератора и низкочастотные компоненты вариаций частоты опорного источника. Это означает что:

- за краткие времена наблюдений, когда τ короче чем $\mathcal{T}$, управляемый генератор имеет такие же частотные флуктуации, как если бы он был свободный;
- за средние и большие времена для $\tau > \mathcal{T}$ кварцевый генератор воспроизводит вариации частоты опорного генератора с точностью до константы K_s.

Из этих наблюдений следует, что если цезиевые часы должны обладать хорошей стабильностью, то должны соблюдаться два условия. Первое, кварцевый генератор должен быть самого лучшего качества. (Существуют и другие причины для такого требования, в которые мы пока не будем вдаваться). Второе, очевидное, вариации атомной частоты должны быть как можно меньше.

(d) Стабильность частоты за средние времена

В дальнейшем мы будем в частности иметь дело с вариациями частоты, связанными со свойствами атомно-лучевой трубки. В этом разделе мы рассмотрим шум, связанный с детектированием атомного пучка. Он определяет нестабильность частоты стандартов на временах выборки τ от нескольких секунд до нескольких дней.

Атомы в пучке и носители зарядов в электрических цепях двигаются через поперечное сечение пучка или проводника соответственно со случайно распределенными временами. Соответствующее распределение является пуассоновским распределением. Из него следует, что выходной ток

цезиевой лучевой трубки флуктуирует около своего среднего значения. Эти флуктуации называются *дробовым шумом*. Для постоянного значения ν_i частоты, определенной из (6.44), петля обратной связи воспринимает эти флуктуации, как если бы они происходили с резонансной частотой цезиевой лучевой трубки и, следовательно, с опорной частотой.

Для того чтобы определить ключевые параметры, воздействующие на эти флуктуации, давайте упростим дальше выражение (6.23) для выходного тока. На практике форма центральной интерференционной полосы близка к синусоидальной, и мы можем написать с достаточной для наших целей степенью приближения

$$I = I_n + \frac{1}{2} I_{pv}\left[1 + \frac{\cos\pi(\nu - \nu_0)}{\Delta\nu_0}\right], \tag{6.53}$$

где I_{pv} и $\Delta\nu_0$ представляют высоту резонансной кривой от пика до долины и ее ширину на половине максимума, как показано на рис.6.7.

Рассмотрим полупериод модуляции, во время которого частота, действующая на атомы равна $\nu = \nu_0 + \nu_m$. Мы получили бы тот же результат, если бы рассматривали другой полупериод модуляции. Флуктуации тока δI, происходящие из-за малых флуктуаций δI_n и δI_{pv} в I_n и I_{pv}, даются следующим выражением

$$\delta I = \delta I_n + \frac{1}{2}\delta I_{pv}\left[1 + \cos\frac{\pi n_m}{\Delta n_0}\right]. \tag{6.54}$$

Если токи I_n и I_{pv} были бы постоянны, а частота резонанса изменялась в пределах $\nu_0 + \delta\nu_0$ с $\delta\nu_0 << \Delta\nu_0$, мы бы имели

$$\delta I = \frac{1}{2} I_{pv}\left[\cos\frac{\pi n_m}{\Delta n_0} + \frac{\pi\delta n_0}{\Delta n_0}\sin\frac{\pi n_m}{\Delta n_0}\right]. \tag{6.55}$$

Приравнивая два выражения для δI и полагая, что глубина модуляции равна полуширине линии, т.е. что $\nu_m = \Delta\nu_0/2$, мы получим следующую формулу для кажущейся относительной вариации резонансной частоты, индуцированной флуктуациями в I_n и I_{pv}:

$$\frac{\delta\nu_0}{\nu_0} = \frac{1}{\pi}\frac{2\delta I_n + \delta I_{pv}}{I_{pv} Q_{at}}, \tag{6.56}$$

где $Q_{at} = \nu_0/\Delta\nu_0$ – добротность атомного резонанса.

Когда поток частиц со средним значением λ удовлетворяет распределению Пуассона, то стандартное отклонение $\sigma'_{\delta\lambda}$ флуктуации $\delta\lambda$ в этом потоке таково, что

$$\sigma'_{\delta\lambda} = \left(\frac{\lambda}{\tau}\right)^{1/2}, \tag{6.57}$$

где τ – время наблюдения флуктуаций. Стандартное отклонение в обычном смысле в относительных флуктуациях частоты $\delta\nu_0/\nu_0$ таким образом равно

$$\sigma'_{\delta\nu_0/\nu_0} = \frac{1}{\pi}\frac{(2I_n + I_{pv})^{1/2}}{I_{pv}Q_{at}}\tau^{-1/2}. \tag{6.58}$$

Так как относительные флуктуации частоты обычно характеризуются вариацией Аллана, то по причинам, изложенным в разделах 5.2.5 и 5.2.6, полезно начать с установления выражения для их плотности спектральной мощности. Дробовый шум попадает в категорию белого шума со спектральной плотностью мощности, равной удвоенному среднему значению потока частиц. Так как два источника дробового шума независимы, то спектральная плотность мощности относительных флуктуаций частоты $y = \delta\nu_0/\nu_0$ равна

$$S_y = \frac{2}{\pi^2}\frac{2I_n + I_{pv}}{I_{pv}^2 Q_{at}^2}. \tag{6.59}$$

Отсюда получается немедленно (раздел 5.2.6) вариация Аллана для относительных флуктуаций частоты в цезиевых пучковых часах:

$$\sigma_y(\tau) = \frac{1}{\pi}\frac{(2I_n + I_{pv})^{1/2}}{I_{pv}Q_{at}}\tau^{-1/2}. \tag{6.60}$$

В частном случае белого частотного шума это стандартное отклонение оказывается равным обычному стандартному отклонению, получаемому из (6.58). Отметим, что числовой коэффициент $1/\pi$ не имеет реального смысла, так как апроксимация получена как попутный результат.

Уравнение (6.60) измеряет вредные эффекты от нежелательного атомного потока I_n. Все усилия при разработке этого типа часов были направлены для его уменьшения до пренебрежимо малой величины.

Отметим, что если I_n пренебрежимо мал по сравнению с I_{pv} для данного τ, то нестабильность частоты $\sigma_y(\tau)$ изменяется как величина

$$q = I_{pv}^{-1/2} Q_{at}^{-1}.$$

Поток атомов, достигающий детектора, и, следовательно, I_{pv} изменяется как D^{-2}, где D – расстояние между детектором и печью, в то время как добротность атомного резонанса меняется пропорционально пролетному рас-

стоянию L между двумя осциллирующими полями. Величина q меняется, таким образом, как D/L. Обычно величина этого отношения больше чем 1, но меньше 2, и приблизительно одинаковая во всех конструкциях цезиевых лучевых трубок, независимо от того являются ли они лабораторными инструментами, или компактными коммерческими приборами. Отсюда следует, что длина цезиевой лучевой трубки не играет ключевой роли в нестабильности частоты прибора, если эта нестабильность определяется дробовым шумом в потоке.

Уравнения (6.59) и (6.60) могут быть написаны в более общей форме. Действительно, так как $I_n + I_{pv}/2$ является выходным током в присутствие частотной модуляции, то величина $2I_n + I_{pv}$ представляет мощность дробового шума P_N в шумовой полосе 1 Гц. Мощность полезного сигнала равна $P_S = I_{pv}^2/4$. Отношение сигнал/шум по мощности для установленной полосы пропускания, измеряемой на боку интерференционной полосы, т.е. на частоте $\nu_0 \pm \nu_m$ равно

$$\left.\frac{P_S}{P_N}\right|_{1\text{Hz}} = \frac{I_{pv}^2}{4(2I_n + I_{pv})}. \tag{6.61}$$

Мы можем затем написать

$$S_y = \frac{1}{2\pi^2}\frac{1}{Q_{at}^2 (P_S/P_N)_{1\text{Hz}}}. \tag{6.62}$$

Вариация Аллана становится

$$\sigma_y(\tau) = \frac{1}{2\pi}\frac{1}{Q_{at}(P_S/P_N)_{1\text{Hz}}^{1/2}}\tau^{-1/2}. \tag{6.63}$$

Последние два соотношения показывают, что относительная нестабильность частоты зависит от относительной точности, с которой отношение сигнал/шум позволяет нам приблизиться к атомному резонансу. Для данного τ нестабильность частоты уменьшается, если добротность и отношение сигнал/шум увеличиваются. Это заключение справедливо для всех пассивных часов.

В промышленных конструкциях ширина атомного резонанса близка к 500 Гц и отношение сигнал/шум, определенное выше, составляет по порядку от 10^5 до 10^6. Таким образом, мы получаем $\sigma_y(\tau) \sim 10^{-11}\tau^{-1/2}$, точное значение зависит от рассматриваемой модели. Промышленные модели с лучшими характеристиками имеют нестабильность частоты около $5\times10^{-12}\tau^{-1/2}$, что совершенно сопоставимо с лучшими пучковыми лабораторными первичными стандартами с магнитным отклонением пучка. (Но неточность первичных стандартов примерно в сто раз меньше).

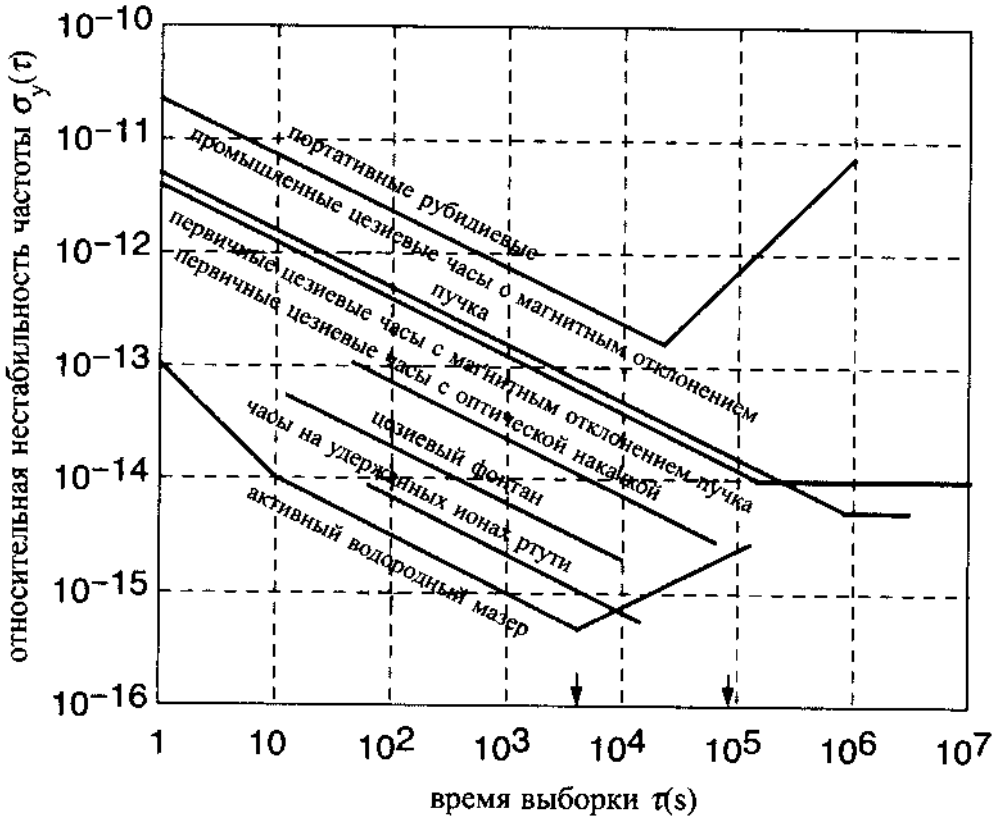


Рис. 6.11. Относительная нестабильность частоты, характеризуемая вариацией Аллана $\sigma_y(\tau)$ для различных стандартов частоты. Кривые дают асимптотические значения. Рисунок показывает нестабильности таковыми, каковыми они наблюдались в 1997 году. К 2000 году нестабильности двух, представленных здесь стандартов частоты, уменьшились примерно в 2,5 раза для цезиевого пучкового стандарта частоты с оптической накачкой и примерно в 3,5 раза для цезиевого фонтана.

Рис. 6.11 показывает типичные нестабильности цезиевых часов и других часов.

В долговременной области для τ большим, чем несколько дней, величина $\sigma_y(\tau)$ уже не уменьшается со временем. Нестабильность ограничивается теперь не дробовым шумом, а вариациями сдвигов частоты в атомных стандартах.

(е) Остаточные сдвиги в частоте часов

На цезиевых атомно-лучевых трубках некоторые физические эффекты и инструментальные неточности приводят к слабому изменению частоты сверхтонких переходов основного состояния атома цезия [6.4]. Мы проанализируем происхождение этих наиболее значительных сдвигов частоты и укажем порядок их величины в относительном выражении. В обсуждении мы ограничимся сдвигами частоты, специфическими для рассматриваемых стандартов частоты. Эффекты гравитационного потенциала и релятивистские эффекты определены в главе 3.

Любые изменения в сдвигах частоты могут привести к долговременной нестабильности частоты. К тому же, неопределенность их значений определяет неточность часов.

Прежде чем пытаться сделать сдвиги минимальными и понять, к чему надо стремиться, важно иметь ясное представление о происхождении различных сдвигов.

(1) Сдвиги частоты физического происхождения

Эффект Зеемана второго порядка. Как можно увидеть из измерений, представленных на рис.6.12 [6.26], в случае, если магнитное поле сверхвысокой частоты параллельно статическому магнитному полю B_0, то можно наблюдать семь переходов $\Delta F = \pm 1$, $\Delta m_F = 0$. Центральный резонанс, соответствующий переходу

$$F = 3,\ m_F = 0 \rightarrow F = 4,\ m_F = 0$$

согласно (6.6) испытывает квадратичный сдвиг частоты $\Delta\nu_Z$. Для типичных значений $B_0 = 6 \times 10^{-6}$ Т мы получим сдвиг $\Delta\nu_Z = 1{,}5388$ Гц, или в относительном выражении около $1{,}7 \times 10^{-10}$. Это наибольший частотный сдвиг, влияющий на опорный переход. Конечная точность стандарта сильно зависит от способа, каким определяется величина этого сдвига.

На практике мы используем атомные свойства сами по себе, и значение B_0 выводится из сдвига частоты в одном или в нескольких из шести переходов $(3, m_F) \leftrightarrow (4, m_F)$, для которых $\Delta m_F = 0$. Чтобы выразить эти сдвиги в функциональной зависимости от B_0 и от атомных констант можно использовать уравнение Брайта–Раби (6.4). Раскрывая это выражение, находим, что первый член пропорционален B_0. Для современного уровня точности в промышленных часах достаточно знать этот первый член. Для первичных пучковых цезиевых стандартов необходимо учитывать квадратичный член этого выражения.

В современных промышленных приборах величина B_0 поддерживается активной системой стабилизации. Для этого среднее значение частоты опроса ν_i, (6.44), переключается на доли секунды от значения $\nu_{Cs} + \Delta\nu_Z$ на какое-то другое предварительно определенной значение $\nu_{Cs} + \Delta\nu_Z'$, где ν_{Cs} – частота сверхтонкого перехода атома цезия. Затем B_0 настраивается таким образом, чтобы переход

$$F = 3,\ m_F = 1 \leftrightarrow F = 4,\ m_F = 1$$

попадал в резонанс с частотой $\nu_{Cs} + \Delta\nu_Z'$. Тогда мы, с достаточной точностью *приближения получим* $B_0 = \Delta\nu_Z'/K$, *где* K *коэффициент, зависящий только* от атомных констант. Он имеет значение, близкое к $7{\times}10^9$ Гц T^{-1}. Такая настройка производится автоматически один или несколько раз в минуту. Значение приложенного магнитного поля автоматически поддер-

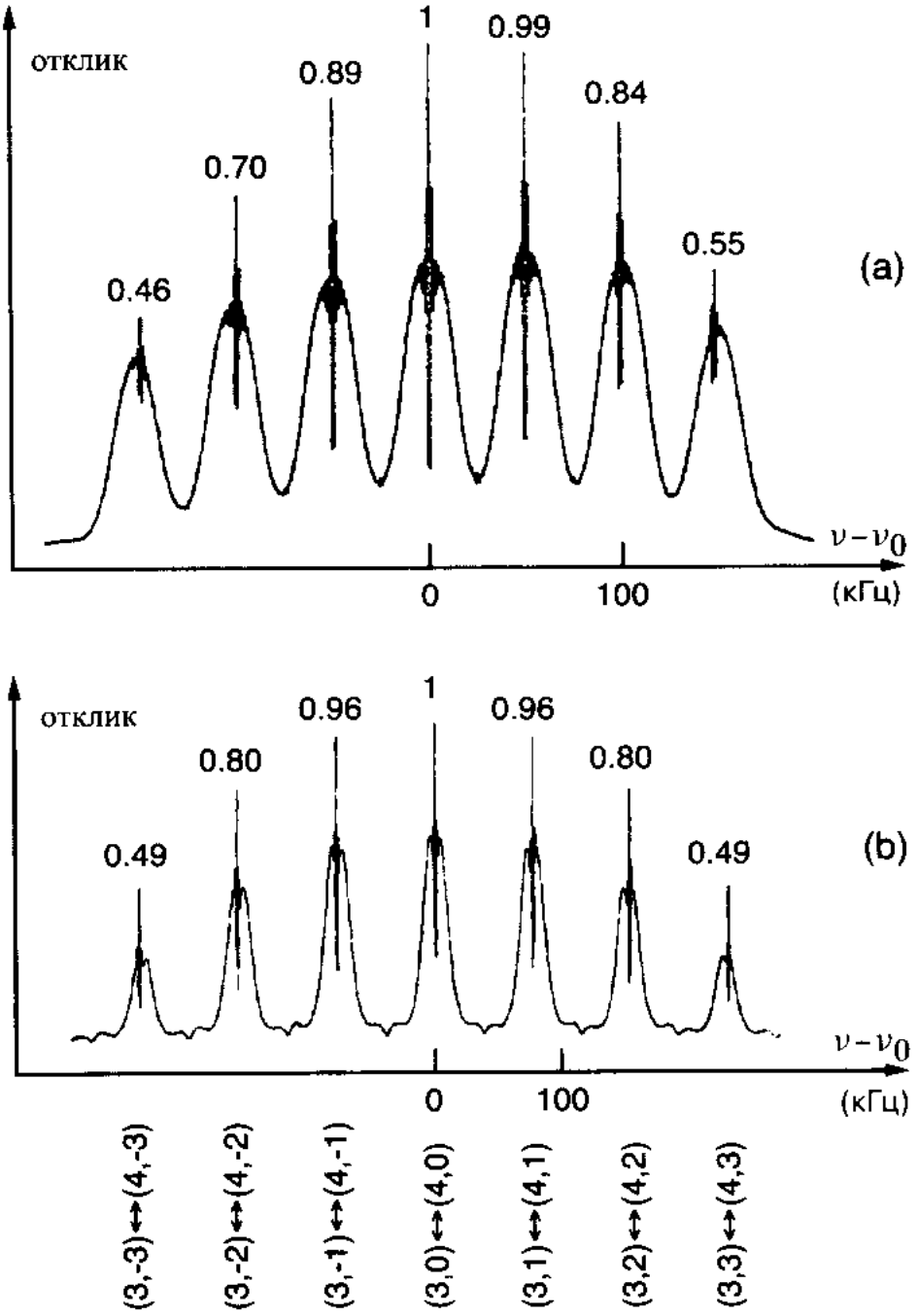


Рис. 6.12. СВЧ резонансы в цезиевой атомно-лучевой трубке. (a) Трубка с магнитным отклонением, (b) трубка с оптической накачкой. Амплитуды боковых линий приведены по отношению к амплитуде центральной линии. Два резонатора имеют разные конструкции и разные значения B_0. (Приведено из [6.26]).

живается на заданной величине [6.27]. В этом случае эффекты от изменяющихся внешних магнитных полей исключаются.

Если магнитное поле неоднородно, то измерения дают среднее значение $\langle B_0 \rangle$ вдоль направления пучка, а сдвиг частоты из-за эффекта Зеемана второго порядка пропорционален $\langle B_0^2 \rangle$, от среднего значения B_0^2. Вообще говоря, $\langle B_0 \rangle^2 \neq \langle B_0^2 \rangle$. Поэтому здесь появляется источник ошибок, который должен быть уменьшен до минимума путем создания по возможности однородного магнитного поля, экранированного от внешних полей.

Эффект Штарка второго порядка. Этот сдвиг частоты в сверхтонкой структуре основного состояния атомов щелочных металлов имеет квадратичную зависимость от электрического поля. В цезиевой атомно-лучевой трубке это поле обычно очень слабое из-за наличия магнитных экранов, действующих также как и фарадеевский цилиндр. Значение напряженности поля 1 В $см^{-1}$, которое на самом деле слишком велико для трубки, давало бы относительный сдвиг частоты около $2{,}5\times10^{-16}$. Это пренебрежимо малая величина.

Эффект излучения черного тела. Внутри замкнутого объема электромагнитное поле, находящееся в равновесии с черным телом при комнатной температуре, занимает очень широкий спектр с максимумом вблизи 30 ТГц (λ = 10 мкм). При температуре 300 К среднеквадратичное значение электрического и магнитного полей равно 832 В $м^{-1}$ и $2{,}77\times10^{-6}$ Т соответственно.

Переменное электрическое поле оказывает такой же эффект, как и постоянное электрическое поле. Для атома цезия относительное значение соответствующего сдвига частоты $\Delta\nu_c$ дается следующим выражением:

$$\frac{\Delta\nu_c}{\nu_{C_s}} = -1{,}69\times10^{-14}\left(\frac{T}{300}\right)^4, \tag{6.64}$$

где T – термодинамическая температура. Сдвиг частоты должен учитываться в лабораторных цезиевых пучковых стандартах.

Переменное магнитное поле также вызывает сдвиг частоты перехода сверхтонкой структуры основного состояния. Можно показать, что этот сдвиг имеет одно и то же значение для всех атомов, которые здесь рассматриваются. Он близок к 10^{-17} в относительном выражении, и в настоящее время им можно пренебречь.

Эффект спин-обменных столкновений. Когда два щелочных атома сталкиваются, их внешние электроны могут поменяться местами. Такое столкновение называется спин-обменным столкновением. Оно вносит сдвиг в частоту сверхтонкого перехода в основном состоянии. Этот сдвиг пропорционален плотности атомов, их средней относительной скорости и эффективному сечению спин-обменного столкновения. В 2000 году все еще не было теоретических оценок для сечения рассеяния в таком цезиевом пучке, в котором скорость атомов имеет тот же порядок величины, что и скорость теплового возбуждения при комнатной температуре. Вычисление усложняется очень большим количеством сверхтонких подуровней. В настоящее время нет экспериментальных доказательств наличия такого сдвига частоты в цезиевых пучковых часах при температуре источника пучка 100°С. Однако в це-

зиевых часах на холодных атомах сдвиг измерим и должен быть принят во внимание (раздел 6.4.3b).

Эффект Доплера второго порядка. Атомы находятся в движении по отношению к прибору, зондирующему их резонансную частоту. Поэтому здесь сказывается релятивистское замедление времени (глава 3), и это приводит к сдвигу частоты перехода из-за эффекта Доплера второго порядка. Если все атомы имеют одну и ту же скорость v, то сдвиг частоты дается выражением (6.16).

Форма резонансной кривой в таком случае описывается следующим выражением, выведенным из (6.23):

$$I = \frac{1}{2} I_0 \int_0^\infty f(\tau) \sin^2 b\tau \left\{ 1 + \cos\left[\omega - \omega_0 \left(1 - \frac{v^2}{2c^2} \right) \right] T \right\} \mathrm{d}\tau , \quad (6.65)$$

где член I_n опущен в предположении его независимости от частоты.

Наблюдаемая резонансная кривая находится путем наложения элементарных резонансных кривых, связанных с соответствующими значениями скорости v. При возрастании скорости кривые сдвигаются в сторону низких частот. Их вес зависит от скорости через коэффициент $f(\tau) \sin^2 b\tau$, так как τ обратно пропорционально скорости. Отсюда следует, что общая резонансная кривая становится несколько асимметричной.

В присутствии прямоугольной частотной модуляции с глубиной ν_m можно показать, что кажущаяся резонансная частота, т.е. частота, на которой сигнал ошибки обращается в нуль (раздел 6.4.1c), дается следующим выражением:

$$\frac{\Delta\nu_D}{\nu_0} = -\frac{\int_0^\infty v f(\tau) \sin^2 b\tau \sin(2\pi\nu_m T) \mathrm{d}t}{2c^2 \int_0^\infty \frac{1}{v} f(\tau) \sin^2 b\tau \sin^2(2\pi\nu_m T) \mathrm{d}t} . \quad (6.66)$$

Это уравнение показывает, что сдвиг частоты зависит от:

- ν_m, вследствие асимметрии в резонансной кривой;
- b, и, следовательно, от амплитуды СВЧ поля возбуждения перехода.

Величина ν_m устанавливается при проектировании соответствующей электроники и она нечувствительна к окружающим условиям. Напротив, значение b может меняться, если меняются внешние условия. В современных цезиевых часах значение b поддерживается постоянным с помощью соответствующей системой обратной связи.

Для рассматриваемых здесь часов сдвиг частоты из-за эффекта Доплера второго порядка составляет около10^{-13}. В общем, среди других причин, его точная величина зависит от распределения времен взаимодей-

ствия τ, и, следовательно, от распределения атомов по скоростям, а так же и от величины b. Однако, если распределение по скоростям узкое, то (6.66) стремится к $-v^2/2c^2$, и сдвиг частоты становится нечувствительным к значениям b и ν_m. Поэтому предпочтительнее по возможности использовать атомы с малым разбросом скоростей и, более того, с настолько малыми скоростями, насколько это возможно. Мы вернемся к этому вопросу в разделе 6.4.3.

(2) Сдвиги частоты инструментального происхождения

Эффект разности фаз между осциллирующими полями. В чисто стоячей электромагнитной волне фаза магнитного поля меняется скачком на π при пересечении каждого узла и остается постоянной между ними. Но существуют потери в резонансной полости из-за конечной проводимости ее стенок. Поэтому появляется бегущая волна, восполняющая потери энергии из-за эффекта Джоуля. Это означает, что стоячая волна уже не является чисто стоячей. Фаза магнитного поля все еще быстро изменяется на величину, близкую к π вблизи узла, но она изменяется почти линейным образом между узлами.

Рис. 6.13 схематически показывает, как меняется фаза в резонансной полости цезиевых часов между точками, где соединяются два плеча и между их короткозамкнутыми концами. Предполагается, что два плеча могут иметь некоторые малые отличия в длине. Так как фаза непрерывна на соединении, то мы наблюдаем разность фаз ϕ между двумя осциллирующими магнитными полями. Эта разность фаз исключительно мала, около десяти микрорадиан. Она появляется из-за конечной проводимости используемого материала в пределах расхождений в конструкции резонансной полости.

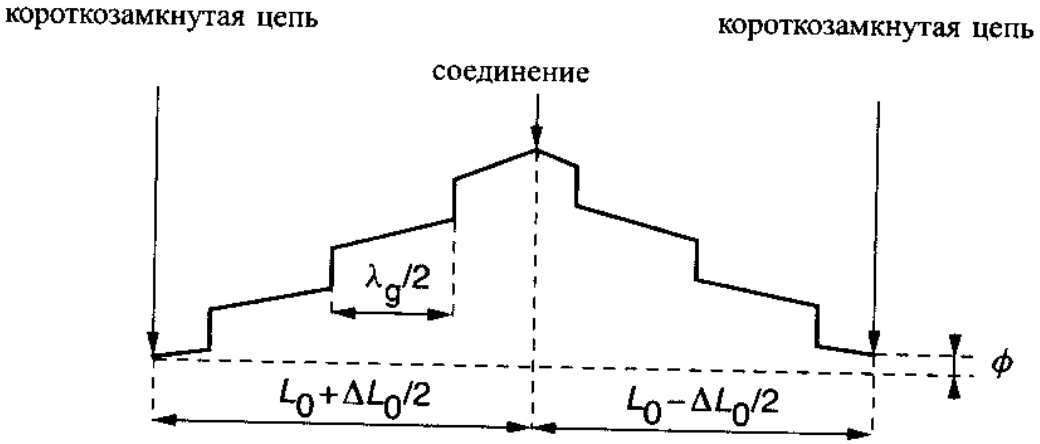


Рис. 6.13. Асимптотическое представление фазовых вариаций в двух плечах резонансной полости. Пучности осциллирующего поля расположены около короткозамкнутых цепей. Скачки фазы на π происходят при пересечении каждого узла. И разница в длине двух плеч и постепенное изменение фазы сильно преувеличены. λ_g – длина волны в волноводе.

Частотный отклик цезиевой лучевой трубки (раздел 6.4.1b) может быть пересчитан в предположении, что фаза второго осциллирующего поля отстает на ϕ от фазы первого осциллирующего поля. Упрощенное выражение (6.22) для вероятности перехода в сверхтонкой структуре для пучка, в котором все атомы имеют одну и ту же скорость, выглядит как

$$P(\tau) = \frac{1}{2}\sin^2 b\tau\{1 + \cos[(\omega - \omega_0)T + \phi]\}. \tag{6.67}$$

Резонансная частота сдвигается на $\Delta\nu_\phi$, где

$$\frac{\Delta\nu_\phi}{\nu_0} = -\frac{\phi}{2\pi\nu_0 T} = -\frac{\upsilon}{2\pi\nu_0 L}\phi, \tag{6.68}$$

где T – время пролета между двумя осциллирующими полями, разделенными расстоянием L. Если этот сдвиг частоты написать в терминах разбаланаса ΔL_0 электрической длины двух плеч резонатора, то мы получим

$$\frac{\Delta\nu_\phi}{\nu_0} = K_\phi \frac{\upsilon}{c}, \tag{6.69}$$

где c – скорость света и величина K_ϕ дается

$$K_\phi = -\frac{\lambda}{\lambda_g}\frac{L_0}{L}\alpha_g \Delta L_0, \tag{6.70}$$

где λ и λ_g – длины волн излучения в свободном пространстве и в волноводе соответственно, α_g – константа затухания СВЧ магнитного поля вдоль волновода, и L_0 –средняя длина плеч. Типичные значения следующие: $\lambda/\lambda_g = 0{,}7$, $L_0/L = 0{,}5$ и $\alpha_g = 2{,}6\times10^{-2}$ м$^{-1}$. Если мы предположим, что $\upsilon = 150$ мс$^{-1}$ и $\Delta L_0 = 10^{-4}$ м, то это даст $(\Delta\nu_\phi/\nu_0| \sim 5\times10^{-13}$. Поэтому любая разница в длине двух плеч резонансной полости должна быть уменьшена до минимума.

Уравнение (6.69) показывает, что эффект разности фаз между двумя осциллирующими полями может быть истолкован как остаточный эффект Доплера первого порядка, связанный с наличием бегущей волны в резонаторе. Постоянная затухания K_ϕ зависит от того насколько тщательно изготовлен резонатор. При вышеуказанных числовых значениях значение K_ϕ близко к 10^{-6}.

Так как сдвиг частоты зависит от скоростей атомов, то существование скоростного распределения искажает резонансную линию таким же образом, как и эффект Доплера второго порядка.

Если частота модулирована, сдвиг частоты дается выражением

$$\frac{\Delta\nu_\phi}{\nu_0} = -\frac{\phi}{2\pi\nu_0 L}\frac{\int_0^\infty f(\tau)\sin^2 b\tau\sin(2\pi\nu_m T)\mathrm{d}\tau}{\int_0^\infty \frac{1}{\upsilon} f(\tau)\sin^2 b\tau\sin^2(2\pi\nu_m T)\mathrm{d}\tau}. \qquad (6.71)$$

Здесь применимы общие комментарии, сделанные в отношении эффекта Доплера второго порядка. Если скорости атомов распределены в некотором диапазоне значений, то сдвиг частоты зависит от глубины модуляции ν_m, а также от амплитуды возбуждения на сверхвысокой частоте, представленной через b. Более того, хорошо видно, что сдвиг частоты уменьшается при уменьшении атомных скоростей.

Знак сдвига частоты изменяется при изменении направления атомного пучка. Это подразумевает, что этот сдвиг в принципе может быть измерен. Такая возможность заложена в большинстве лабораторных первичных пучковых цезиевых стандартов частоты.

Фаза магнитного поля сверхвысокой частоты зависит от его пространственного положения в области взаимодействия. Эти пространственные вариации также порождаются бегущей волной, восполняющей потери энергии в стенках резонатора. Хотя они могут быть сведены к минимуму путем специального выбора конфигурации концевых частей каждого плеча резонатора, они все еще остаются, даже если разность фаз между центрами двух областей взаимодействия обращается в нуль. Разность фаз ϕ, рассматриваемая выше, должна по этой причине рассматриваться как общая разность фаз между двумя осциллирующими полями, т.е как сумма разностей фаз, связанных с разной длиной двух плеч резонатора и распределенной разностью фаз, которая усредняется по траектории движения атомов в каждой области взаимодействия. В цезиевых часах с магнитным отклонением это усреднение может быть оценено довольно приблизительно. Действительно, отклонение атомов в селектирующих магнитах зависит от их скорости. Отсюда следует, что атомные скорости и распределение траекторий связаны в поперечном сечении пучка, и это доказывает трудность точного моделирования процесса.

В методе обращения пучка очень важно воспроизвести распределение траекторий в двух областях взаимодействия как можно более одинаковым способом для обоих направлений атомного пучка.

Эффект затягивания резонатором. Если частота резонатора ν_{cav} не равна в точности частоте атомного резонанса ν_0, то амплитуда магнитного поля сверхвысокой частоты изменяется асимметрично при модуляции частоты около ν_0. Следовательно, при прямоугольной модуляции значение b оказывается не одним и тем же для двух последовательных полупериодов. Средняя частота возбуждения в таком случае должна быть сдвинута так, чтобы, несмотря на этот разбаланс, отклик цезиевой трубки принимал два равных значения для двух последовательных полупериодов час-

тотной модуляции. Сигнал ошибки, управляющий частотой кварцевого генератора, в этом случае будет обращаться в нуль.

Давайте предположим, что расстройка резонатора $\nu_c - \nu_0$ мала по сравнению с шириной его собственного резонанса. Соответствующий этой расстройке сдвиг $\Delta\nu_{cav}$ в частоте атомного резонанса дается следующим выражением

$$\frac{\Delta\nu_{cav}}{\nu_0} = \frac{\nu_c - \nu_0}{\nu_0} \frac{2T_c^2 \int_0^\infty b\tau f(\tau) \sin b\tau \cos b\tau (1 + \cos\omega_m T) d\tau}{\int_0^\infty T^2 f(\tau) \sin^2 b\tau \frac{\sin\omega_m T}{\omega_m T} d\tau}, \qquad (6.72)$$

где $\omega_m = 2\pi\nu_m$ и ν_m – глубина модуляции. По вышеприведенной гипотезе амплитуда магнитного поля сверхвысокой частоты меняется только незначительно во время модуляции и b представляет одно из ее значений. T_c есть постоянная времени отклика резонатора, связанная с его добротностью через

$$T_c = \frac{Q_c}{\pi\nu_c}. \qquad (6.73)$$

Относительный сдвиг (6.72), постоянная отклика резонатора (6.73) и добротность атомного резонанса $Q_{at} \sim 2T\nu_0$, связаны между собой примерно следующим образом:

$$\frac{\Delta\nu_{cav}}{\nu_0} \sim \frac{\nu_c - \nu_0}{\nu_0} K_{cav} \frac{Q_c^2}{Q_{at}^2}, \qquad (6.74)$$

где K_{cav} – есть константа, зависящая от ν_m, от b и от распределения скоростей. Это уравнение говорит нам, что для данной расстройки частоты кажущийся сдвиг частоты атомного резонанса примерно пропорционален отношению квадрата добротности резонатора к квадрату добротности атомного резонанса. Очевидно, что очень желательно сделать это отношение по возможности минимальным.

Если частота возбуждения модулирована указанным выше способом, т.е. когда мы имеем $\omega - \omega_0 = \pm\omega_m$ в (6.23), то отклик цезиевой лучевой трубки дается следующим выражением:

$$I(b, \omega_m) = I_n + \frac{I_0}{2} \int_0^\infty f(\tau) \sin^2 b\tau (1 + \cos\omega_m T) d\tau. \qquad (6.75)$$

Отметим, что числитель в правой стороне (6.72) пропорционален производной по b от отклика трубки, а именно $\partial I(b, \omega_m)/\partial b$. Следовательно,

если резонансная полость расстроена, то сдвиг частоты исчезает, если выполняется условие

$$\frac{\partial I(b,\omega_m)}{\partial b}=0. \tag{6.76}$$

Это происходит для таких значений амплитуды сверхвысокой частоты возбуждения, при которых отклик цезиевой атомно-лучевой трубки имеет экстремум. Эта особенность является специфической для прямоугольной модуляции.

Рис.6.14 показывает характер изменений, наблюдаемых в $I(b,\omega_m)$ для изменяющейся амплитуды сверхвысокой частоты возбуждения. Рабочая точка выбирается там, где находится первый максимум отклика трубки.

В современных цезиевых пучковых стандартах электронная система автоматики следит за амплитудой СВЧ возбуждения, при которой отклик резонатора имеет максимум при наличии прямоугольной модуляции, т.е. когда частота равна или $\nu_0+\nu_m$ или $\nu_0-\nu_m$. Система автоматики поддерживает соответствующее значение амплитуды и, следовательно, соответствующее значение b [6.28, 6.29]. В этом способе использования свойств цезиевой лучевой трубки:

- сдвиг частоты, связанный с расстройкой резонансной полости, исключается;
- долговременное значение b стабилизируется вместе со всеми сдвигами частоты, зависимыми от b.

Эффект соседних переходов. Сверхтонкие подуровни, ближайшие к подуровням $F=3$, $m_F=0$ и $F=4$, $m_F=0$ (рис.6.2), могут возмущать часовой переход. Это возмущение обусловлено двумя источниками, ведущих к сдвигам частоты, и известным как *затягивание Раби* и *затягивание Рэмси*.

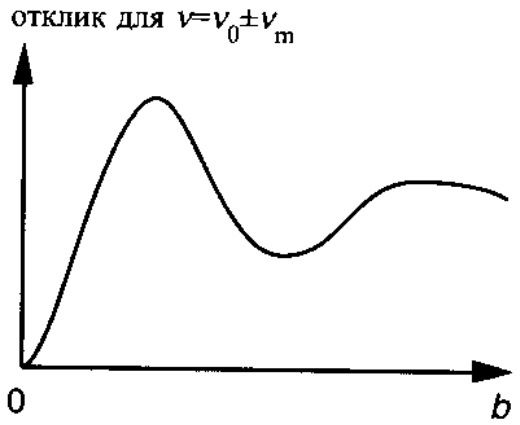


Рис. 6.14. Изменения отклика цезиевой атомно-дучевой трубки от амплитуды осциллирующего поля, частота которого равна либо $\nu_0+\nu_m$, либо $\nu_0-\nu_m$

- *Затягивание Раби.* Крылья пьедестала Раби для переходов

$$F = 3, m_F = 1 \leftrightarrow F = 4, m_F = 1$$

и

$$F = 3, m_F = -1 \leftrightarrow F = 4, m_F = -1$$

дают вклад, хотя и очень слабый, в отклик цезиевой атомно-лучевой трубки, если частота возбуждения равна частоте часового перехода ν_0. Боковые линии расположены симметрично относительно ν_0. Однако их амплитуды могут отличаться, как это можно увидеть на спектре из рисунка 6.12а, который типичен для цезиевых атомно-лучевых трубок с магнитным отклонением. В таких резонаторах магнитные поля, достигаемые в селектирующих магнитах, еще недостаточны для того, чтобы $\partial E(F, m_F)/\partial b$ в (6.13) имело одно и то же значение для всех сверхтонких подуровней. Отсюда следует, что угол отклонения зависит от рассматриваемого подуровня сверхтонкой структуры, и, следовательно, амплитуды боковых резонансов не симметричны с обеих сторон по отношению к центральному резонансу.
Когда существует такая асимметрия, то добавляется частотно-зависимый вклад в отклик трубки, описанный (6.23). Это несколько искажает опорную резонансную линию и возникает сдвиг в резонансной частоте. Он в основном зависит от мощности СВЧ сигнала, так как амплитуды крыльев пьедестала Раби пропорциональны этой мощности, а также значению статического магнитного поля B_0, определяющего разделение спектральных компонент, показанных на рис.6.12а [6.30].

- *Затягивание Рэмси.* До сих пор мы предполагали, что СВЧ магнитное поле параллельно статическому магнитному полю. Однако имеются также компоненты, перпендикулярные статическому полю, и обусловленные двумя причинами. Причина первая, линии СВЧ магнитного поля должны замыкаться в области, занимаемой каждой резонансной модой в резонансной полости. Это означает, что эти линии должны загибаться в пересекаемой атомами области, как показано на рис.6.6. Причина вторая, статическое магнитное поле может быть недостаточно идеальным. Это приводит к тому, что могут индуцироваться переходы $\Delta F = \pm 1$, $\Delta m_F = \pm 1$, хотя их вероятность и много меньше, чем вероятность переходов $\Delta F = \pm 1$, $\Delta m_F = 0$.

Четыре перехода

$$(3, 0) \leftrightarrow (4, \pm 1) \text{ и } (3, \pm 1) \leftrightarrow (4, 0)$$

имеют общий уровень с часовым переходом

$$(3, 0) \leftrightarrow (4, 0) .$$

Отсюда следует, что волновые функции для подуровней (3, 0) и (4, 0) возмущаются, когда осуществляются переходы $\Delta F = \pm 1$, $\Delta m_F = \pm 1$.

Было показано, что этот эффект добавляет член к вероятности часового перехода, и что это также искажает резонансную линию, если СВЧ спектр атома цезия не симметричен, как это показано на рис. 6.12а. Как и ранее соответствующий сдвиг частоты зависит от мощности СВЧ волн и от статического поля B_0 [6.31].

В конструкциях промышленных цезиевых атомно-лучевых трубок, использующих магнитное отклонение, относительный сдвиг частоты, происходящий из-за наличия соседних переходов, может достигать значения 10^{-12}. Он может быть уменьшен одним из следующих способов:

(1) такой конструкцией трубки, в которой амплитуды резонансов (3, 1 ↔ 4, 1) и резонансов (3, –1 ↔ 4, –1) были бы одинаковыми;

(2) ограничением степени пространственного взаимодействия с осциллирующим магнитным полем;

(3) настройкой направления статического и СВЧ полей с максимальной точностью.

Сдвиги частоты уменьшаются, если атомный резонанс становится уже. Обычно сдвиги малы в первичных стандартах частоты.

Эффект переходов Майорана. Если магнитное поле изменяется вдоль атомных траекторий, как это происходит возле магнита, селектирующего атомы по состояниям, то атомы подвергаются зависящим от времени возмущениям благодаря своему движению. Эти возмущения могут иметь спектральные компоненты, соответствующие частотному разделению между подуровнями состояний $F = 3$ и $F = 4$. В этом случае могут индуцироваться переходы $\Delta F = 0$, $\Delta m_F = \pm 1$, которые называются переходами Майорана. Если они связаны с переходом

$$(3, 0) \leftrightarrow (4, 0),$$

то частота часового перехода сдвигается. Для предотвращения таких переходов при разработке атомных часов предпринимаются все усилия для того, чтобы происходило достаточно плавное изменение магнитного поля.

Эффект нежелательных спектральных компонент. Спектр сигнала опроса может иметь нежелательные компоненты. Некоторые из них производятся цепями питания на частотах электрических систем и на кратных частотах. Другие порождаются нежелательными частотами, возникающими при синтезе частоты сигнала возбуждения.

Наличие таких компонент возмущает свойства квантовой системы из двух состояний $F = 3, m_F = 0$ и $F = 4, m_F = 0$. По этой причине изменяется разность энергий между этими уровнями и это приводит к изменению

частоты перехода. Эффект исчезает если интерферирующие спектральные компоненты симметричны относительно частоты атомного перехода по положению и по амплитуде. Другими словами сдвиг частоты наиболее чувствителен к присутствию мешающих линий, если эти линии лежат в спектральном интервале с центром на частоте ν_0 и с шириной в несколько раз шире пьедестала Раби. В этом диапазоне прилагаются все усилия для того, чтобы мощность нежелательных спектральных компонент была бы снижена до 10^{-7} раз от мощности полезного СВЧ сигнала.

Эффект неидеальности электронных систем. Если мы предположим, что атомный резонанс имеет ширину 500 Гц и если мы хотим, чтобы ошибка в определении резонансной частоты оставалась ниже 5×10^{-14}, то электроника, связанная с цезиевой лучевой трубкой должна измерять частоту резонанса с ошибкой, которая не должна превышать одну миллионную часть ширины линии,

Электронные системы, связанные с цезиевой лучевой трубкой поэтому должны проектироваться и изготавливаться с величайшей тщательностью. Требуемые характеристики наиболее легко достигаются в цифровых системах обработки сигналов.

Эффект неидеальностей наиболее эффективно уменьшается при уменьшении ширины наблюдаемого резонанса.

(3) Другие сдвиги частоты. Иногда возникают совершенно неожиданные сдвиги частоты. Именно поэтому затягивание Рэмси было идентифицировано и промоделировано совсем недавно, в конце восьмидесятых.

Исследование таких сдвигов является непрерывной заботой физиков, проектирующих и эксплуатирующих атомные стандарты частоты и, в частности, цезиевые пучковые стандарты частоты. С этой позиции очень важно создание независимых первичных цезиевых пучковых стандартов в нескольких лабораториях, с использованием по возможности совершенно разных конструкций и последующей проверкой того, что частоты, производимые инструментами, взаимно совместимы до соответствующих неопределенностей измерения.

(f) Долговременная нестабильность, невоспроизводимость и неточность

Сдвиги частоты атомных резонансов вызываются макроскопическим окружением атомов. Любое изменение в этом окружении может приводить к смещениям наблюдаемой частоты атомного перехода. Окружающая среда является первоисточником долговременной нестабильности и невоспроизводимости. Хотя существует множество частотных сдвигов, их величины очень малы по сравнению с частотой атомного резонанса, так что наибольшие сдвиги составляют около 10^{-10}. Благодаря тому, что относи-

тельные изменения этих сдвигов сами по себе очень малы, то тем более малы долговременная нестабильность и невоспроизводимость, как это будет показано позже. Точно также, если относительная неопределенность значения сдвига частоты очень мала, то и неточность в цезиевых стандартах времени и частоты тоже исключительно мала.

Долговременная нестабильность частоты. В цезиевых пучковых стандартах сдвиг частоты зависит главным образом от значения магнитного поля B_0, от распределения атомов по скоростям, от положения атомных траекторий по отношению к СВЧ магнитному полю в резонансной полости, от настройки частоты резонатора и от амплитудных и спектральных параметров СВЧ сигнала возбуждения. Значение того или иного из этих параметров может изменяться под влиянием окружающего магнитного поля, механических напряжений, температуры, атмосферного давления и даже влажности. Из этого следует, что в долгосрочном плане частота стандарта может изменяться в виде отклика на изменение окружающих условий или на эффекты старения материалов и электронных компонент.

Первичные пучковые цезиевые стандарты располагаются в тщательно контролируемой окружающей среде. В дополнение, значения важных параметров систематически измеряются. Долговременная нестабильность в таких стандартах не превышает значение 10^{-14} за время $\tau > 10$ дней.

В современных промышленных часах значение статического магнитного поля стабилизируется по свойствам самих атомов. Другая цепь стабилизации, использующая свойства цезиевого резонатора, удерживает амплитуду СВЧ сигнала возбуждения на постоянном уровне, подавляя таким способом эффекты затягивания резонатором. В результате система становится далеко не столь чувствительной к окружающим возмущениям. Долговременная нестабильность становится также, очень мала, и фактически она меньше 10^{-14} за время $\tau > 10$ дней.

Невоспроизводимость. Эта концепция особенно применима к промышленным моделям. Она описывает пригодность атомных стандартов частоты данной конструкции для воспроизведения без калибровки номинального значения частоты стандарта, если внешние окружающие условия находятся внутри установленных спецификацией диапазонов значений. Невоспроизводимость характеризует относительное отклонение выходной частоты от включения к включению либо из-за внутренних свойств инструмента, либо из-за какой-то части оборудования. Она обычно выражается в стандартном отклонении наблюдаемой разности частот в относительном выражении.

Невоспроизводимость уменьшается при улучшении конструкции и при лучшем осуществлении контроля на стадии изготовления. В зависимости от модели она лежит в диапазоне от 5×10^{-13} до 10^{-11}.

Неточность. Точность пучкового цезиевого стандарта частоты характеризует его способность без калибровки производить частоту, согласующуюся с определением секунды. Для достижения точности необходимо наблюдение изолированных атомов цезия, находящихся в покое. Однако в реальных приборах атомы движутся и подвергаются различным возмущениям. Неточность равна всеобъемлющей неопределенности, которая находится путем объединения всех неопределенностей, происходящих от разных сдвигов частоты, и деленных на невозмущенную частоту (раздел 5.5)

Неопределенности состоят из двух типов. Как пример рассмотрим, что происходит, когда измеряются эффекты разности фаз между двумя осциллирующими полями. Полуразность между измеренными частотами для каждого направления пучка равна, в принципе, сдвигу частоты, происходящему за счет разности фаз между двумя осциллирующими полями. Во время каждого измерения частота, производимая стандартом, флуктуирует случайным образом, например из-за дробового шума в пучке. Поэтому с измеренным значением сдвига частоты связано *случайное стандартное отклонение.* Мы также должны принять во внимание тот факт, что пространственная зависимость фазы осциллирующего поля и распределение траекторий в поперечном сечении атомного пучка точно не известны. Сдвиги частоты, связанные с пространственно зависимой разностью фаз, вряд ли имеют одно и то же значение для двух направлений пучка. Поэтому появляется систематическая ошибка в измеренном сдвиге частоты. Ее оценочная величина, характеризуемая стандартным отклонением, иногда называется *систематическим отклонением.* За неимением лучшего метода эта ошибка принимается как случайная [6.32]. Вариации в измеренных эффектах из-за разности фаз между двумя осциллирующими полями равны, таким образом, сумме квадратов двух стандартных отклонений.

Отметим, что для того чтобы измерить (экспериментально доступный) сдвиг частоты, влияющий на первичный стандарт, необходимо иметь вспомогательный стандарт. Этот стандарт необязательно должен быть точным. Он должен быть очень стабильным на интервалах, во время которых изменяются рабочие параметры испытуемого стандарта. Обычно в качестве вспомогательного стандарта используется водородный мазер.

В промышленных цезиевых часах значение выходной частоты корректируется на сдвиг из-за эффекта Зеемана второго порядка. Для пользователя все происходит так, как если бы сверхтонкий переход наблюдался в нулевом магнитном поле. В то же время выходная частота подвергается всем другим сдвигам частоты, таким как эффект Доплера второго порядка и эффект разности фаз между двумя осциллирующими полями. Оцениваемые значения этих сдвигов частоты рассматриваются как неопределенности в измеренном значении частоты атомного перехода. Они дают

вклад в неточности этих стандартов частоты. Установленные значения неточности лежат, в зависимости от конструкции, между 10^{-12} и 3×10^{-11}.

В лабораторных цезиевых пучковых стандартах ширина резонанса на порядок уже, чем в промышленных системах. Цель состоит в уменьшении сдвигов частоты, зависящих от ширины резонанса. Такой эффект может быть достигнут за счет изготовления длинного прибора с $L \sim 1$ м. В некоторых случаях используется магнитное отклонение чтобы селектировать атомы со скоростями, близкими к 90 м $с^{-1}$, в то время как наиболее вероятная скорость в пучке составляет около 250 м $с^{-1}$ [6.11, 6.12, 6.13, 6.14]. Различные сдвиги частоты измеряются или оцениваются с максимальной точностью. Неточности в лабораторных цезиевых пучковых стандартах с магнитной селекцией лежат в диапазоне между 7×10^{-15} и 10^{-13} в зависимости от конструкции [6.14, 6.33].

В таблице 6.3 раздела 6.8 приведены сравнения неточности цезиевых часов с неточностями других атомных стандартов.

Отметим, что неточность данного стандарта частоты (например, 3×10^{-12}) является необходимым верхним пределом его невоспроизводимости (например, 5×10^{-13}), а также его долговременной нестабильности частоты (например, 2×10^{-12}) для всего времени жизни инструмента.

6.4.2. Цезиевые пучковые часы с оптической накачкой

Среди полупроводниковых лазеров, разработанных для нужд оптической телекоммуникации, арсенид-галлиевые лазеры излучают длину волны, близкую к 0,85 мкм, которую легко настроить на совпадение с цезиевой линией D_2 (Рис.6.1). Переходы между $^2S_{1/2}$ и $^2P_{3/2}$ уровнями можно использовать для селекции и детектирования состояний атомов оптическим способом. Полупроводниковые лазеры, излучающие на длине волны 0,89 мкм (линия цезия D_1), разработаны не достаточно хорошо. Хотя эксперименты и продемонстрировали, что линия D_1 может быть исключительно полезной из-за меньшего числа уровней в $^2P_{1/2}$ состоянии, тем не менее использование этого преимущество не является обычной практикой [6.17].

(а) Описание

Цезиевые атомно-лучевые трубки с оптической накачкой включают в себя источник, который производит атомный пучок, и резонансную полость, похожую на ту, которая применяется в приборах с магнитным отклонением пучка. Однородное статическое магнитное поле B_0, в котором пролетает атомный пучок, проходит параллельно осциллирующим полям через два плеча резонатора и через область пролетного пространства между ними. Магниты заменены на зоны, в которых атомы взаимодействуют с

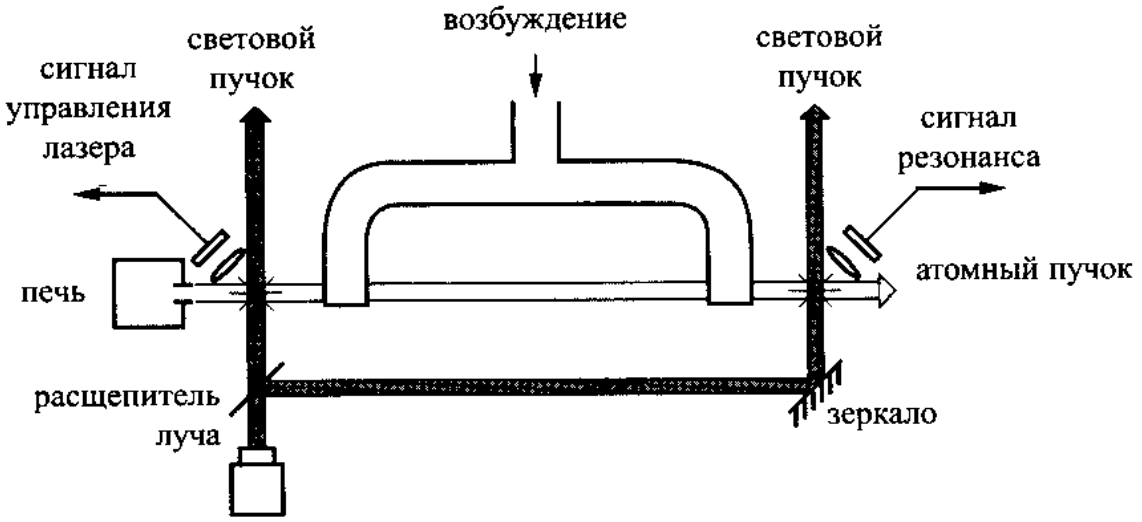


Рис. 6.15. Схематическая диаграмма цезиевой пучковой трубки с оптической подготовкой и оптическим детектированием атомных состояний. Предполагается использование одного лазера. Магнитные экраны не показаны

лазерным светом. Рис. 6.15 показывает простейшую схему, в которой единственный лазер обеспечивает нужды и подготовки атомов и детектирования их состояний [6.17].

Поляризованный должным образом лазерный пучок от лазера имеет мощность в несколько милливатт. Она разделяется на две части с помощью полупрозрачного расщепителя луча. Два пучка света, полученные таким образом, пересекают атомный пучок под прямым углом с каждой стороны резонансной полости. В результате создаются две зоны взаимодействия между атомами и светом. В каждой зоне создается магнитное поле. В той области, где происходят СВЧ переходы, оно параллельно СВЧ магнитному полю и его значение зависит от выбора оптических переходов. Некоторые переходы наиболее эффективны при магнитных полях порядка 5×10^{-5} Т, примерно в десять раз больших, чем обычные значения B_0, лежащие вблизи 6×10^{-6} Т [6.34].

Селекция состояний происходит в первой зоне оптического взаимодействия. Путем оптической накачки (раздел 6.2.2) атомы переводятся с одного уровня сверхтонкой структуры основного состояния, т.е. $F = 3$ или $F = 4$, на другой. Существует несколько схем оптической накачки. Рассмотрим, например, σ-свет, т.е. свет, поляризованный линейно в направлении, перпендикулярном статическому магнитному полю. Если он настроен на переходы между уровнями $F = 3$ основного состояния и $F' = 3$ состояния $^2P_{3/2}$, то говорят, что такой свет возбуждает переход (3-3)σ. Он переводит около 13% атомов в сверхтонкий подуровень основного состояния $F = 4$, $m_F = 0$. Подуровни $F = 4$, $m_F \neq 0$ заселяются несколько меньше, но их населенность одинакова для одинаковых по модулю уровней $| m_F |$, как это можно видеть из симметрии на диаграмме 6.16. Подуровни в состоянии $F = 3$ обедняются [6.34].

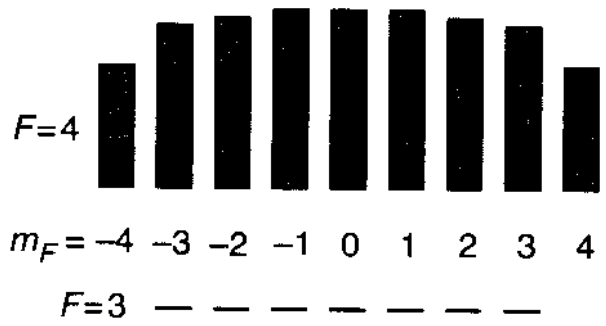


Рис. 6.16. Столбчатая диаграмма, показывающая распределение населенностей на уровнях $F = 3$ и $F = 4$ основного состояния для (3-3)σ накачки

Таким образом разность населенностей между уровнями $F = 4$, $m_F = 0$ и $F = 3$, $m_F = 0$ равна13% от числа всех атомов, произведенных источником. Эту величину надо бы сравнивать с величиной, полученной с помощью магнитной селекции по состояниям. В последнем случае она почти равна начальной населенности атомов, принадлежащих подуровню $F = 3$, $m_F = 0$, который отклоняется к оси цезиевой лучевой трубки, а именно 1/16 или около 6% начального числа атомов. В действительности доля полезных атомов значительно меньше этой величины, так как магнитное отклонение селективно по скоростям. Поэтому атомы, которые движутся слишком быстро или слишком медленно, исключаются из числа полезных атомов. А в случае оптической накачки угол отклонения атомного пучка, происходящий благодаря обмену моментами с лазерными фотонами, по крайней мере, в сотни раз меньше, чем при магнитной селекции состояний. Поэтому потери атомов ничтожны. Таким образом, мы видим, что подготовка состояний с помощью оптической накачки приводит к более сильному сигналу резонанса.

Поток флуоресцирующих фотонов, излучаемых спонтанно, при переходах атомов из состояния $^2P_{3/2}$ в состояние $^2S_{1/2}$, достигает максимума, когда частота излучения лазера совпадает с частотой оптического перехода. Фотоны флуоресценции собираются оптической системой и используются для управления частотой лазера методом, похожим в принципе на метод, описанный в разделе 6.4.1c. Для этого частота лазера модулируется, и сигнал флуоресценции детектируется синхронно с частотой модуляции. Таким образом, получается сигнал ошибки, который используется двумя различными способами. Он изменяет температуру лазерного диода с помощью прибора, основанного на эффекте Пельтье, и, таким образом, осуществляется грубая подстройка частоты лазера. Точная настройка осуществляется добавлением сигнала ошибки к току питания лазера, который также определяет лазерную частоту.

Часовой переход детектируется во второй оптической зоне взаимодействия. После того как атомный пучок проходит через СВЧ область

взаимодействия, населенность уровня $F = 3$, $m_F = 3$ становится пропорциональной вероятности совершения часового перехода. Число атомов, которые падающий лазерный свет переводит из уровня основного состояния $F = 3$ на уровень $F' = 3$ возбужденного состояния, пропорционально этой вероятности. То же самое происходит с потоком флуоресцирующих фотонов. Среднее число излученных фотонов в случае, когда детектируются атомы в состоянии $F = 3$, $m_F = 0$, равно примерно четырем. Они собираются с максимально возможной эффективностью с помощью спроектированных для этого оптических систем, чтобы обеспечить полезный сигнал максимально большим по сравнению со стохастическими флуктуациями. Флуктуации происходят как из-за теплового возбуждения фотодетектора, так и по другим причинам. Нежелательные источники засветки фотодиода по возможности исключаются из его окрестностей. Засветки производятся отражением или рассеянием пучка света, проходящего через цезиевую лучевую трубку.

В описанной установке селекция по состояниям и детектирование выполняются с помощью *накачивающего перехода*. На практике оптическая накачка завершается, когда атомы покидают оптическую зону взаимодействия длиной в несколько миллиметров. Отсюда следует что:

- атомы используются с максимально возможной эффективностью;
- распределение атомов по скоростям практически точно такое же, как максвелловское распределение атомов в пучке;
- детектируемый сигнал чувствителен к амплитудным и частотным шумам лазера.

На рис. 6.17 показан пример резонансного сигнала при сканировании СВЧ излучения по частоте часового перехода [6.17]. Сигнал получен на цезиевой лучевой трубке с такой оптической накачкой, как мы только что описали, Можно увидеть, что для этого резонанса боковые осцилляции Рэмси сильно ослаблены по сравнению с осцилляциями на рис. 6.7. Это объясняется широким распределением атомов по скоростям в цезиевой трубке с оптической накачкой.

Описанная выше цезиевая лучевая трубка с одним лазером хорошо подходит для компактных систем. Экспериментально изучались и другие конфигурации с различными оптическими переходами. Мы приведем два примера. Первый относится к лабораторному цезиевому пучковому стандарту с оптической накачкой, а второй используется с целью повышения эффективности подготовки атомов по состояниям.

В лабораторных приборах детектирование улучшается применением *циклических переходов* (раздел 6.2.2c). Например, используется следующая установка [6.35]. Атомный пучок подготавливается путем возбуждения перехода между уровнем $F = 4$ состояния $^2S_{1/2}$ и уровнем $F' = 4$ состо-

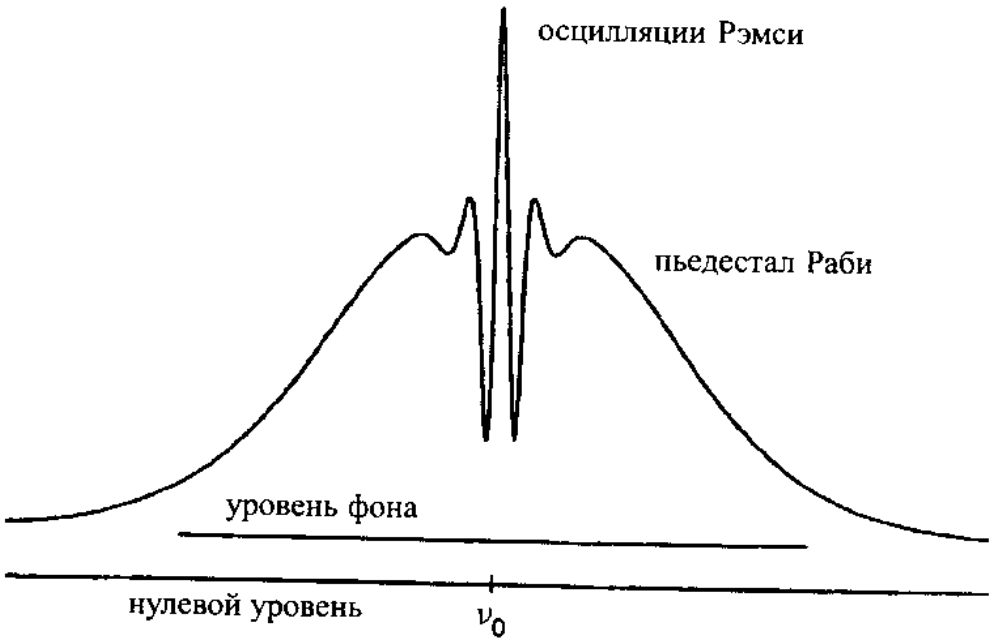


Рис. 6.17. Экспериментальная резонансная кривая в цезиевой лучевой трубке с оптической накачкой. Ширина центральной осцилляции 500 Гц. Взято из [6.17]

яния $^2P_{3/2}$. Этот переход заселяет подуровень $F = 3$, $m_F = 0$ и обедняет уровень $F = 4$, $m_F = 0$. СВЧ взаимодействие опять заселяет подуровень $F = 4$, $m_F = 0$. Затем этот подуровень детектируется вторым лазером, настроенным на циклический переход $F = 4 \leftrightarrow F' = 5$. В этом случае очень важна чистота спектра лазерного излучения. Таким образом, мы получаем несколько десятков фотонов на каждый атом, совершивший часовой переход. Этот возросший поток фотонов маскирует внутренний шум фотодетектора и таким образом улучает отношение сигнал/шум в процессе детектирования. Отношение сигнал/шум стремится к пределу, определяемому дробовым шумом детектируемых фотонов. Более того, число излученных каждым атомом флуоресцентных фотонов пропорционально времени пролета атома через световой пучок, так что детектирование с помощью циклических переходов более эффективно для медленных атомов. Для данного расстояния между двумя осциллирующими полями мы, таким образом, получим более узкий резонанс, чем он был бы при использовании только одного лазера.

Эффективность подготовки состояний может быть повышена путем накачки двумя лазерами с соответствующей поляризацией. Возможно несколько схем накачки. Например, один лазер настраивается на переход $F = 3 \leftrightarrow F' = 4$. Он перекачивает атомы на уровень $F = 4$ основного состояния. Второй лазер настраивается на переход $F = 4 \leftrightarrow F' = 4$ и обедняет уровень $F = 4$ за исключением подуровня $F = 4$, $m_F = 0$, потому что переходы с $F - F' = 0$, $m_F - m_{F'} = 0$ запрещены. Тогда все атомы стремятся попасть на подуровень (4, 0). Теория и эксперимент показывают, что по крайней мере 90% атомов, полученных из источника, таким образом, могут быть

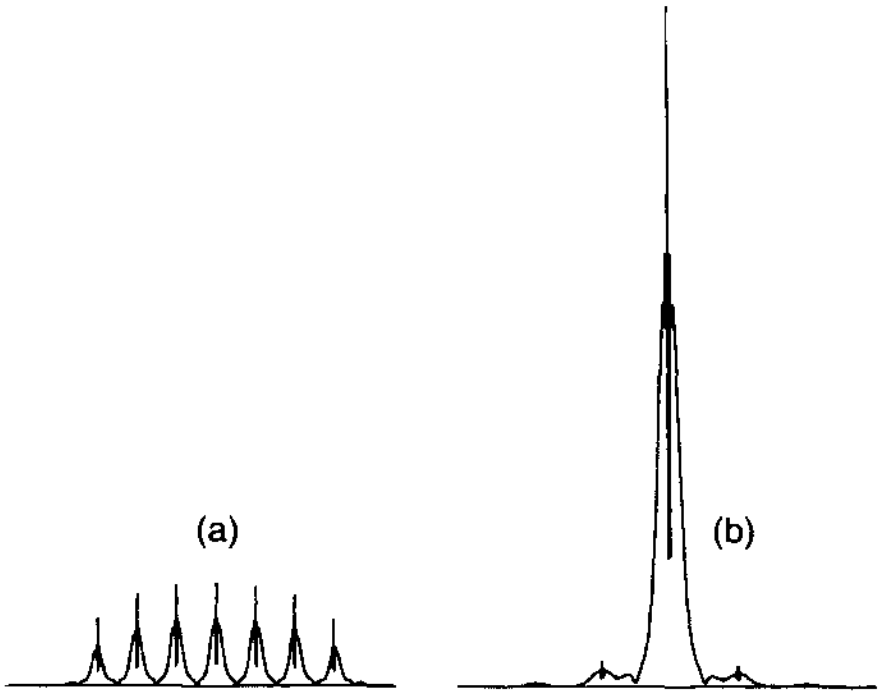


Рис. 6.18. СВЧ спектр атомов цезия. (a) Накачка с одним лазером, (b) накачка с двумя лазерами. Вертикальная шкала одна и та же, а горизонтальные шкалы разные. Взято из [6.36]

переведены на этот подуровень. На рис. 6.18 показан в сравнении экспериментальный СВЧ спектр семи переходов атома цезия, удовлетворяющих условию $\Delta F = 1$, $\Delta m_F = 0$ в зависимости от того, были ли подготовлены состояния с помощью одного лазера или с помощью двух [6.36]. В последнем случае интенсивность часового перехода очень сильно возрастает, в то время как интенсивность других переходов $\Delta F = 1$, $\Delta m_F = 0$ значительно ослабляется, так как подуровни $F = 4$, $m_F \neq 0$ обеднены. Однако было замечено, что с использованием обычных лазеров, доступных во времена проведения этих экспериментов, увеличенная эффективность подготовки состояний сопровождалась уменьшением отношения сигнал/шум при наблюдении сверхтонкого резонанса. Совсем недавно появилась возможность использовать все преимущества этой улучшенной подготовки состояний путем применения лазеров с очень узким спектром излучения [6.37].

(b) Нестабильность частоты за средние времена

Резонансный сигнал в цезиевых лучевых трубках с оптической накачкой сильнее, потому что отсутствуют потери атомов из-за селективного отклонения по скоростям и атомы подготавливаются более эффективно. Коэффициент улучшения, измеряемый через ток фотодетектора I_{pv}, составляет около сотни.

Однако это преимущество уравновешивается уширением резонанса. Действительно в стандартных трубках с магнитным отклонением распределение скоростей может быть смещено в сторону более низких скорос-

тей. Обычно наиболее вероятная скорость детектируемых атомов в промышленных трубках с магнитным отклонением близка к 140 мс$^{-1}$, а в трубках с оптической накачкой она лежит в диапазоне 200 –250 мс$^{-1}$.

В компактных моделях с одним лазером шумы фотодетектора, используемого во второй оптической области взаимодействия, преобладают над шумами наблюдения резонанса. Если для детектирования используется циклический переход, то достигается более благоприятная ситуация, в которой источником наблюдаемого шума становится дробовой шум собираемых фотонов.

Однако эти же преимущества сказываются и в более сильных резонансах на крыльях. За средние временна стабильность частоты цезиевых часов с оптической накачкой (рис.6.11) лучше, чем у приборов с магнитной селекцией состояний. Это было проверено и на компактных [6.38] и на громоздких приборах [6.19]. При использовании двух лазерных излучений для подготовки состояний путем оптической накачки, и третьего лазера для детектирования состояний через циклический переход была получена стабильность частоты $3{,}8\times10^{-13}\tau^{-1/2}$ [6.37].

(c) Остаточные сдвиги частоты

С помощью цезиевых трубок с оптической накачкой мы получаем лучший контроль и лучшее понимание величин сдвигов частоты, влияющих на переходы в сверхтонкой структуре. Появляется и другой сдвиг, называемый световым сдвигом. Он происходит из-за возмущения сверхтонких уровней энергии самим светом, однако он мал и не приводит к вредным последствиям.

Соседние переходы и эффект Зеемана второго порядка. Одним важным преимуществом оптической накачки является то, что она позволяет получать симметричные СВЧ спектры, как это можно увидеть на рис.6.12b. Приводящие к этому условия легко удовлетворяются. Свет накачки должен быть поляризован перпендикулярно магнитному полю B_0, и лазерная частота должна совпадать с центром частоты оптического перехода таким образом, чтобы сверхтонкие подуровни, характеризуемые одинаковыми значениями $|m_F|$, были одинаково населены. В таком случае затягивание Раби и Рэмси исчезают, потому что соседние переходы возмущают часовой переход симметрично.

То, что отсутствуют селектирующие магниты, создававшие сильное магнитное поле за пределами сердечников, позволяет проще решить задачу создания более однородного магнитного поля в СВЧ областях взаимодействия. Это, в свою очередь, уменьшает неопределенность, связанную со сдвигом частоты из-за эффекта Зеемана второго порядка. И, наконец, становится легче избавляться от переходов Майорана.

Эффект Доплера второго порядка и эффект разности фаз между двумя осциллирующими полями. Атомные скорости в резонаторе с оптической накачкой больше, чем в приборе с магнитным отклонением. Но зато распределение скоростей лучше известно. То же самое касается и атомных траекторий. Более того, одинаковые скоростные распределения применимы к каждой возможной траектории.

Эффект Доплера второго порядка и существование разности фаз между двумя осциллирующими полями по этой причине приводят к большему сдвигу частоты, но зато с меньшей неопределенностью. Т.к. распределение по скоростям шире чем для магнитного отклонения, то эти сдвиги частоты более сильно зависят от значения параметра b, пропорционального амплитуде осциллирующего поля согласно определению (6.18). Это не имеет больших последствий, если значение b контролируется таким образом, чтобы удовлетворить условию (6.76).

Эффект затягивания резонатором. Для данной расстройки резонатора и для данного расстояния между двумя осциллирующими полями сдвиг частоты больше в трубках с оптической накачкой, потому что атомы с большей скоростью уширяют атомный резонанс. Однако, как уже обьяснялось, этот сдвиг частоты исчезает, если значение b автоматически подстраивается.

Световой сдвиг. Энергии двух уровней сверхтонкой структуры основного состояния изменяются под действием лазерного возбуждения. Это приводит к сдвигу частоты в часах [6.4], известному как световой сдвиг.

В цезиевых лучевых трубках с оптической накачкой сдвиг частоты из-за света возмущает атомы в области СВЧ взаимодействия. Этот свет происходит от флюоресцирующих атомов в области общей для атомного и светового пучков и, возможно, так же от нежелательного отражения и рассеяния.

Теоретический анализ и экспериментальные исследования этого сдвига были выполнены в [6.18, 6.39, 6.40]. По оценкам сдвиг имеет относительное значение около 10^{-14} в длинных лабораторных трубках и около 10^{-13} в компактных трубках. Он может быть достаточно точно измерен, и не создает проблем на сегодняшней стадии разработки стандартов частоты на цезиевых лучевых трубках с оптической накачкой [6.18].

Другие сдвиги частоты. Сдвиги частоты, которые не были упомянуты выше (например, эффект излучения черного тела, столкновения, нежелательные спектральные компоненты, неидеальная электроника) сильно не отличаются в обоих типах атомных лучевых трубок.

(d) Долговременная нестабильность, невоспроизводимость и неточность частоты

В пучковых цезиевых стандартах частоты с оптической накачкой легче контролировать нежелательные сдвиги частоты, а их неопределенности в

принципе меньше чем в приборах с магнитным отклонением. Это благоприятствует уменьшению долговременной нестабильности, невоспроизводимости и неточности.

Относительно недавние оценки цезиевых лучевых стандартов частоты с оптической накачкой подтвердили их возможности. Неточность первичного лабораторного стандарта была оценена как $6{,}3\times10^{-15}$, а для компактного прибора 4×10^{-13} [6.18, 6.38].

Неопределенность двух лабораторных стандартов достигла величины 5×10^{-15} [6.18, 6.19]. Их частоты находятся в согласии с частотами лучших стандартов с магнитным отклонением. Точность компактных приборов оценивается величиной 4×10^{-13} [6.38].

Представляется вероятным, что стандарты частоты с оптической накачкой будут способны достичь уровня долговременной нестабильности и неточности, приближающегося к 10^{-15}.

(c) Заключения

Оптические методы, очевидно, привнесли с собой ряд преимуществ. Цезиевую лучевую трубку становится легче изготавливать, потому что атомные траектории почти не отклоняются. Более сложные компоненты с точки зрения производства, такие как детектор с горячей проволокой и особенно электронный умножитель (в промышленных приборах), больше не являются необходимыми. Тот факт, что стабильность улучшилась, являющийся сам по себе преимуществом, также означает, что остаточные сдвиги частоты легче измерить. Первичные стандарты частоты с оптической накачкой воспроизводят частоты, согласующиеся с частотами их двойников с магнитным отклонением. Это очень важный факт, показывающий что приборы, реализующие единицу времени и частоты различными способами, не дают систематического сдвига из-за каких-либо неизвестных причин.

6.4.3. Цезиевые часы с лазерным охлаждением

Уже было давно осознано, что характерные особенности цезиевых часов сильно зависят от скорости атомов, на переходах которых они работают. И действительно, как мы показали выше, нестабильность частоты уменьшается, если ширина резонансной линии сужается, т.е. если атомы движутся более медленно. К тому же многие частотные сдвиги являются возрастающими функциями скорости.

Разработка методов охлаждения атомов открыла многообещающие перспективы уменьшения нестабильности частоты, остаточных сдвигов частоты и неточности цезиевых атомных стандартов частоты. Мы представим два метода охлаждения. Затем мы проэкзаменуем первичный стандарт частоты на фонтане холодных атомов и опишем его характеристики. За

этим последует описание проектов разработки стандартов частоты на холодных атомах для космических применений. Наконец, мы обрисуем возможные преимущества использования рубидия вместо цезия.

(а) Охлаждение атомов лазерным излучением

Низкотемпературные печи не могут производить интенсивный пучок, т.к. давление паров является падающей функцией от температуры. Более того, распределение атомов по скоростям в парах находится в термодинамическом равновесии со стенками контейнера и быстро затухает для скоростей, которые значительно меньше, чем наиболее вероятная скорость. Например, в цезиевом пучке, произведенном печью при температуре 100 °C, вероятность наблюдения атомов со скоростью, равной 5 $мс^{-1}$ в 3×10^4 раз меньше, чем вероятность наблюдения атомов, скорость которых равна наиболее вероятной скорости в пучке, а именно, 260 $мс^{-1}$. К тому же медленно движущиеся атомы больше отклоняются столкновениями. Отсюда следует, что атомный пучок почти не содержит холодных атомов. Это и есть та причина, по которой Захариасу не удалось создать атомный фонтан в 1950-х годах [6.41]. Его идея состояла в том, чтобы направить атомный пучок вертикально вверх. Только наиболее медленные атомы не достигали бы потолка прибора и падали бы обратно. Эти атомы должны были подвергаться действию осциллирующего поля в зоне взаимодействия, расположенной вблизи атомного источника, как во время подъема, так и во время падения. Был построен прибор высотой в несколько метров для того, чтобы обеспечить пролетное время T между двумя зонами взаимодействия такое, чтобы оно было порядка одной секунды.

Возможности построить атомные стандарты частоты на очень холодных атомах не было до тех пор, пока не были открыты новые процессы охлаждения. Эти последние восходят к сформулированным между 1970–1975 годами различным предложениям по охлаждению атомов в газовой фазе путем действия на них непрерывным потоком фотонов. Затем последовали теоретические оценки, которые предсказали, что можно охладить и даже полностью остановить атомы через обмен энергией и моментами с фотонами [6.42, 6.43, 6.44]. Разработка полупроводниковых лазеров с очень высокой спектральной чистотой излучения на длине волны 0,85 мкм, привела к применению их к цезиевым атомам с середины 1980-х годов. С тех пор было изобретено, проанализировано и исследовано экспериментально большое количество лазерных охлаждающих процессов [6.45, 6.46, 6.47, 6.48, 6.49].

Мы опишем концептуально простейший способ замедления атомов, использовавшийся на первых стадиях охлаждающих процессов. Мы так же рассмотрим более тонкий метод, разработанный чтобы завершить операцию охлаждения, приводящую к температурам до 2 мкК.

Доплеровское охлаждение. Рассмотрим экспериментальную установку, показанную на рис. 6.19a. Атом подвергается действию двух монохроматических бегущих световых волн одинаковой интенсивности и с одинаковой частотой f_L, распространяющихся в противоположных направлениях. Рис. 6.19b показывает, что центр спектра поглощения атома расположен на резонансной частоте f_0. Этот спектр поглощения связан с переходом между уровнем основного состояния и уровнем возбужденного состояния $^2P_{3/2}$ атомов цезия. Частота световой волны настроена так, что

$$f_L < f_0 . \tag{6.77}$$

Если атом поглощает фотоны, распространяющиеся вправо, он получает дополнительные моменты в этом направлении. Через несколько десятков наносекунд после поглощения каждого фотона атом затухает путем излучения флуоресцентного фотона. Это излучение изотропно, и момент отдачи, связанный с этой флуоресценцией, в среднем равен нулю. Общий эффект таков, что атом толкается вправо. Аналогично, световой пучок, направленный влево, стремится толкать атомы в левом направлении.

Предположим, что атом находится в покое. Он поглощает в среднем одинаковое число фотонов с каждой волны. Поэтому средняя сила, действующая на атом, равна нулю. Если атом движется с некоторой скоростью v вдоль оси светового пучка, то включается эффект Доплера. В опорной системе атома все случается так, как если бы частота световой волны была сдвинута на величину $\Delta f_{Dop} = f_L v/c$ по абсолютной величине. Частота волны, навстречу которой атом движется, кажется более близкой к его

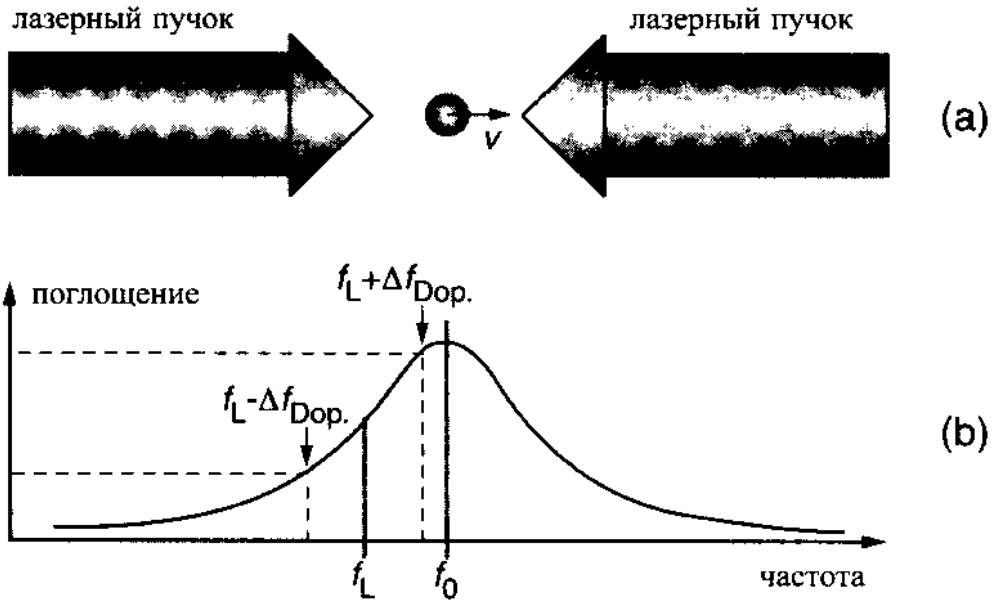


Рис. 6.19. Доплеровское охлаждение. (a) Атом, имеющий резонансную частоту f_0, подвергается действию двух световых волн с одной и той же амплитудой и одной и той же частотой $f_L<f_0$, распространяющихся в противоположных направлениях. (b) Атом поглощает больше из лазерного пучка, который направлен против его движения.

резонансной частоте, в то время как частота другой волны кажется более удаленной от резонансной частоты, как это показано на рис. 6.19b. Поэтому атом поглощает больше фотонов со стороны, которая направлена против его движения, чем со стороны, совпадающей с направлением его движения, и поэтому он испытывает эффект торможения. Этот эффект может быть описан как

$$\mathbf{F} = -\alpha \mathbf{v}\,, \tag{6.78}$$

где α – коэффициент вязкостного трения. Последний достигает максимума, когда частота двух световых волн меньше частоты атомного резонанса на величину, равную полуширине резонансной кривой, которая составляет для цезия около 2,5 МГц. Коэффициент трения пропорционален интенсивности световой волны.

Легко показать, что за каждый цикл поглощение-излучение скорость атома изменяется на

$$\Delta \upsilon = \frac{h}{M\lambda}, \tag{6.79}$$

где M – масса атома и λ – длина волны излучения. Для цезия Δv = 3,5× 10^{-3}мс$^{-1}$. Это значит, что атом должен испытать несколько десятков тысяч циклов излучение-поглощение, прежде чем он приобретет скорость, достаточно близкую к нулю. Поэтому лазерный свет должен быть довольно интенсивным, если охлаждение должно длиться только доли секунды.

Вязкостное затухание из-за эффекта Доплера не может привести к произвольно малой величине скорости. Действительно, атомы подвергаются остаточному возбуждению, происходящему по причине случайной природы самого процесса. Каждый флуоресцентный фотон излучается в произвольном направлении, и, соответственно, атом испытывает отдачу таким же случайным образом. Как число флуоресцентных циклов, происходящих за данное время, является случайным, так и изменение атомного момента за это время является случайным. Температура может быть связана с остаточным возбуждением. Можно показать, что доплеровское охлаждение не может уменьшить температуру ниже T_D, которая дается

$$kT_D = \frac{\hbar}{2\tau}\,, \tag{6.80}$$

где τ – время жизни возбужденного состояния. Значение T_D для цезия близко к 120 мкК.

Схема, показанная на рис. 6.19a осуществляет замедление по одной координате. При использовании трех пар лазерных пучков с взаимно перпендикулярными направлениями можно замедлять атомы по всем трем координатам. В каждой паре две световые волны имеют одну и ту же ча-

стоту и амплитуду и распространяются в противоположных направлениях. Это можно получить путем отражения лазерного пучка вдоль его траектории.

На пересечении трех пар пучков атомы подвергаются действию вязкостной силы, которая их замедляет и передает им очень маленькую скорость рассеяния. Вот почему область пересечения трех лазерных пучков называется *оптической трясиной.*

Уравнения (6.78) справедливы, если скорости атомов достаточно малы, так чтобы сдвиг частоты из-за эффекта Доплера сильно не превышал ширину атомного резонанса. Для того чтобы эффект торможения был значительным, начальная скорость атомов должна быть меньше скорости захвата v_c такой, что

$$v_c \sim \frac{\lambda}{2\pi\tau}, \tag{6.81}$$

которая не зависит от интенсивности лазерного излучения, и где λ – длина волны лазерного излучения, а τ – время жизни возбужденного состояния. Для цезия v_c порядка нескольких метров в секунду.

На практике метод охлаждения, рассматриваемый здесь, используется для атомов, содержащихся в ячейке с очень малым давлением. Их скорости распределены в очень широком диапазоне и, хотя вероятность найти атомы со скоростями ниже нескольких метров в секунду мала, но она не равна нулю. Когда такие атомы попадают в оптическую трясину, они захватываются и охлаждаются. Распределение атомов по скоростям в ячейке восстанавливается через столкновения с газовой фазой атомов или через столкновения со стенками, потому что эти столкновения перераспределяют скорости случайным образом. И, наконец, значительная часть атомов может удерживаться. Обычно для лазерных пучков диаметром 2 см и с интенсивностью порядка10 мВт $см^{-2}$ в зоне оптической трясины из атомов образуется облако диаметром около 1 см, содержащее около 10^7 атомов. Для охлаждения требуется около 100 мс.

Свойства оптической трясины могут быть улучшены путем наложения неоднородного магнитного поля. Таким способом мы можем создать возвращающую силу, пропорциональную расстоянию от центра оптической трясины. В таком случае атомы попадают в потенциальную яму, называемую *магнито-оптической ловушкой.* В ней могут быть получены более высокие плотности атомов. Однако присутствие магнитного поля и тот факт, что его необходимо выключать при извлечении атомов, создает практические ограничения на применение магнито-оптической ловушки. В результате обычно предпочитают охлаждать атомы в простой оптической трясине, если их хотят использовать дальше для работы в цезиевых часах.

Охлаждение с помощью эффекта Сизифа. Большим сюрпризом было обнаружение того факта, что атомы охлаждались до температуры ниже доплеровского предела (6.80).

Оказалось, что кроме доплеровского эффекта существуют более утонченные и даже более эффективные механизмы охлаждения. Один из них основан на пространственной вариации поляризации света в стоячей волне и на присутствии большого числа энергетических подуровней в основном состоянии атома. Охлаждение происходит благодаря совместному действию оптической накачки и изменений энергии (световые сдвиги) в подуровнях основного состояния [6.45, 6.46].

Рассмотрим стоячую волну, полученную суперпозицией двух световых пучков, распространяющихся в противоположных направлениях с равными амплитудами и с взаимно перпендикулярными линейными поляризациями. В стоячей волне поляризация света непрерывно меняется на расстояниях, равных примерно $\lambda/2$, от линейной, к правой круговой, далее линейной, к левой круговой и опять линейной.

Мы упростим объяснение механизма охлаждения, предположив, что в основном состоянии мы имеем только два подуровня. Их энергия возмущается световым излучением и зависит от его поляризации. Поэтому энергия изменяется пространственно, с периодичностью $\lambda/2$. Это показано на рис. 6.20 в одномерном представлении. Энергия подуровня 1 минимальна, когда энергия подуровня 2 максимальна и наоборот.

Давайте предположим, что в какое-то начальное время атом занимает подуровень 1, и он расположен на дне потенциальной энергии. Если он, например, движется вправо, то его потенциальная энергия возрастает, а

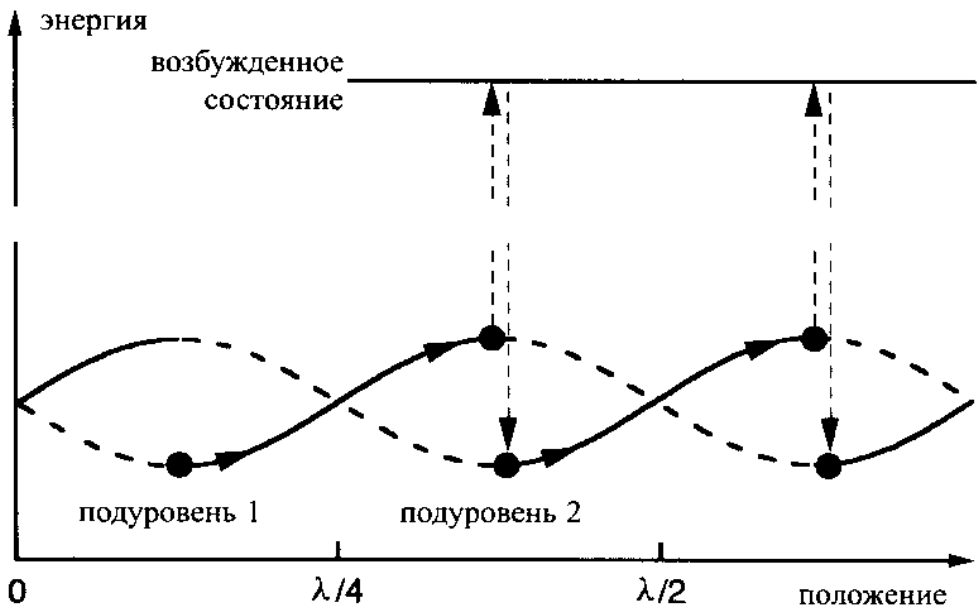


Рис.6.20. Охлаждение с помощью эффекта Сизифа. При своем движении атом вскарабкивается на потенциальный пик. Оптическая накачка сбрасывает атом в потенциальную яму.

кинетическая убывает, т.е. атом замедляется. Получается так, что когда атом достигает пика потенциальной энергии, поляризация света приводит к максимальной вероятности перекачки атома в возбужденное состояние. Почти немедленно после переизлучения фотона атом обнаруживает себя с большей вероятностью на подуровне 2. Но этот уровень находится как раз на дне потенциальной ямы. Продолжая свое движение, атом опять вскарабкивается на следующий пик и опять теряет кинетическую энергию. В конце концов, он захватывается на дне потенциальной ямы. Результат рассмотрения будет тот же самый, если атом движется налево, или как если бы он стартовал с подуровня 2.

Этот эффект охлаждения был назван эффектом Сизифа в честь героя греческой мифологии Сизифа, который был наказан тем, что он должен был закатывать камень на вершину горы. Как только он достигал вершины, камень тотчас срывался и скатывался на дно долины.

В этом случае оптическая трясина образуется механизмом, для которого не требуется эффект Доплера. Его особенности другим образом зависят от рабочих параметров. Коэффициент трения не зависит от интенсивности лазерного света I_L и он пропорционален расстройке δ между частотой света и атомной резонансной частотой. Скорость, при которой захватывается атом, дается уравнением, похожим на (6.81), в котором время жизни возбужденного состояния заменяется на постоянную времени, характеризующую оптическую накачку. Эта константа обратно пропорциональна I_L, так что скорость, при которой атом захватывается, пропорциональна I_L. На практике эта скорость меньше, чем в случае доплеровского охлаждения. Отсюда следует, что охлаждению с помощью эффекта Сизифа должно предшествовать доплеровское охлаждение.

Достигаемая температура пропорциональна интенсивности I_L и обратно пропорциональна отстройке частоты δ, если отстройка больше ширины резонанса. При охлаждении с помощью эффекта Сизифа предел нижней температуры определяется энергией отдачи атома при излучении им флуоресцентного фотона. Эта предельная температура определяется из

$$kT_R \sim \frac{h^2}{M\lambda^2} . \qquad (6.82)$$

Для атома цезия она близка к 0,1 мкК.

На практике сразу же после доплеровского охлаждения интенсивность I_L светового пучка уменьшается примерно в 10 раз, и отстройка частоты δ увеличивается примерно до десятикратной резонансной ширины, составляющей для цезия 5 МГц. При таких параметрах в течение несколько миллисекунд может достигаться температура около 2 мкК. Такая температура соответствует среднеквадратичной скорости 1,5 см $с^{-1}$.

Подбрасывание атомов. Облако атомов можно подбросить вверх с контролируемой начальной скоростью путем использования эффекта Доплера.

Рассмотрим еще раз схему, показанную на рис.6.19а, но предположим, что свет, распространяющийся вправо, имеет частоту $f_L + \Delta f_L$, а свет, распространяющийся влево, имеет частоту $f_L - \Delta f_L$. Если атом движется вправо со скоростью v_0 такой, что

$$v_0 = \frac{c\Delta f_L}{f_L} = \Delta f_L \lambda_L, \tag{6.83}$$

то в системе координат, находящейся в равномерном прямолинейном движении вместе с атомом, все происходит так, как если бы частота светового пучка, движущегося влево, возросла на $f_L v_0/c = \Delta f_L$ благодаря эффекту Доплера. В этой системе атом находится в покое, и он воспринимает частоту лазерного пучка как f_L. Аналогичные доводы показывают, что то же самое происходит и со световым пучком, движущимся вправо.

Поэтому атомы захватываются оптической трясиной движущейся со скоростью $v_0 = \Delta f_L \lambda_L$. Скорость подбрасывания определяется величиной Δf_L. Если $\Delta f_L = 5$ МГц, то для атома цезия мы получаем скорость $v_0 = 4{,}25$ м с$^{-1}$.

(b) Часы на фонтане атомов цезия с лазерным охлаждением

Мы видели, каким образом можно подбросить облако холодных атомов со скоростью в несколько метров в секунду. В принципе именно поэтому и возможно увеличение времени взаимодействия Δt в (6.8) примерно до одной секунды в приборе с разумными размерами. Однако в лаборатории атом будет падать под действием земного гравитационного поля, и это необходимо принять во внимание. Если бы атомы запускались горизонтально, то после всего лишь одной секунды полета они бы упали вниз на 5 м! Это привело бы к значительным практическим трудностям. Лучше воспользоваться преимуществом земного притяжения и запускать атомы вертикально вверх, так, чтобы они поднимались на несколько дециметров и падали обратно вниз. Таким образом, можно получить фонтан, идея которого впервые была выдвинута Захариасом.

Если запускать атомы с начальной скоростью v_0, то они достигнут кульминационной высоты H над точкой запуска

$$H = \frac{v_0^2}{2g}, \tag{6.84}$$

где g – ускорение силы тяжести. Мы найдем, что $H = 0{,}6$ м, если начальная скорость $v_0 = 3{,}4$ м с$^{-1}$. Время T_{path}, необходимое этим атомам чтобы добраться до вершины и вернуться в точку запуска, определяется из

$$T_{\text{path}} = \frac{2v_0}{g} = 2\left(\frac{2H}{g}\right)^{1/2}. \qquad (6.85)$$

Это время T_{path} равно 0,7 с при вышеуказанной скорости или высоте.

Описание атомного резонатора. Первый цезиевый фонтан на холодных атомах, созданный для исследования его свойств в качестве первичного стандарта частоты, был изготовлен в LPTF (лаборатория первичного времени и частоты) в Париже [6.21, 6.22, 6.23]. Схематически установка показана на рис. 6.21. Мы приведем упрощенное описание его работы.

Область, в которой атомы охлаждаются и запускаются, и где происходит подготовка атомных состояний, расположена на пересечении лазерных лучей, обозначенных М. Диаметр лучей равен 2 см. В этой области парциальное давление паров цезия, определяемое температурой холодной точки, составляет около 10^{-6} Па. Охлаждение и запуск выполняются через циклические переходы между сверхтонким уровнем $F = 4$ состояния $^2S_{1/2}$ и уровнем $F' = 5$ состояния $^2P_{3/2}$. Начнем с того, что мощность лазера и его частота настраиваются таким образом, чтобы захватить и охладить атомы с помощью доплеровского охлаждения. Для этого требуется несколько сотен миллисекунд. Затем атомы вовлекаются в движение вверх путем сдвига частот вертикальных лазерных лучей на несколько миллисекунд. Частота луча, направленного вверх, сдвигается на $\Delta f_L > 0$, а частота луча, направленного вниз, сдвигается на $-\Delta f_L$. Скорость запуска определяется (6.83). Мощность всех лазерных лучей затем уменьшается, а их частоты понижаются на одинаковую величину, примерно на 50 МГц. Двигающиеся вверх атомы охлаждаются из-за эффекта Сизифа. Полученное таким образом облако содержит примерно 10^7 атомов в объеме 1 см^3. Облако движется с начальной скоростью несколько метров в секунду. Температура атомов в этот момент равна нескольким мкК. Флуктуации скорости при такой температуре составляют примерно 1 см с$^{-1}$. Поэтому атомы в облаке движутся почти с одинаковой скоростью.

Атомные состояния подготавливаются во время, когда облако атомов все еще пересекает горизонтальные пучки. Эти пучки на несколько миллисекунд настраиваются на переход накачки $F = 4 \rightarrow F' = 4$, и атомы переходят на сверхтонкий уровень $F = 3$ основного состояния. Все лазерные пучки затем выключаются чтобы не вызвать световой сдвиг часового перехода.

Атомы попадают в область, где парциальное давление цезия исключительно мало, порядка 10^{-8} Па, чтобы уменьшить эффект от столкновений с остаточными парами цезия. Они пересекают в первый раз цилиндрическую резонансную полость. На резонатор постоянно подается сигнал возбуждения с частотой 9,192... ГГц. Атомы достигают вершины своей

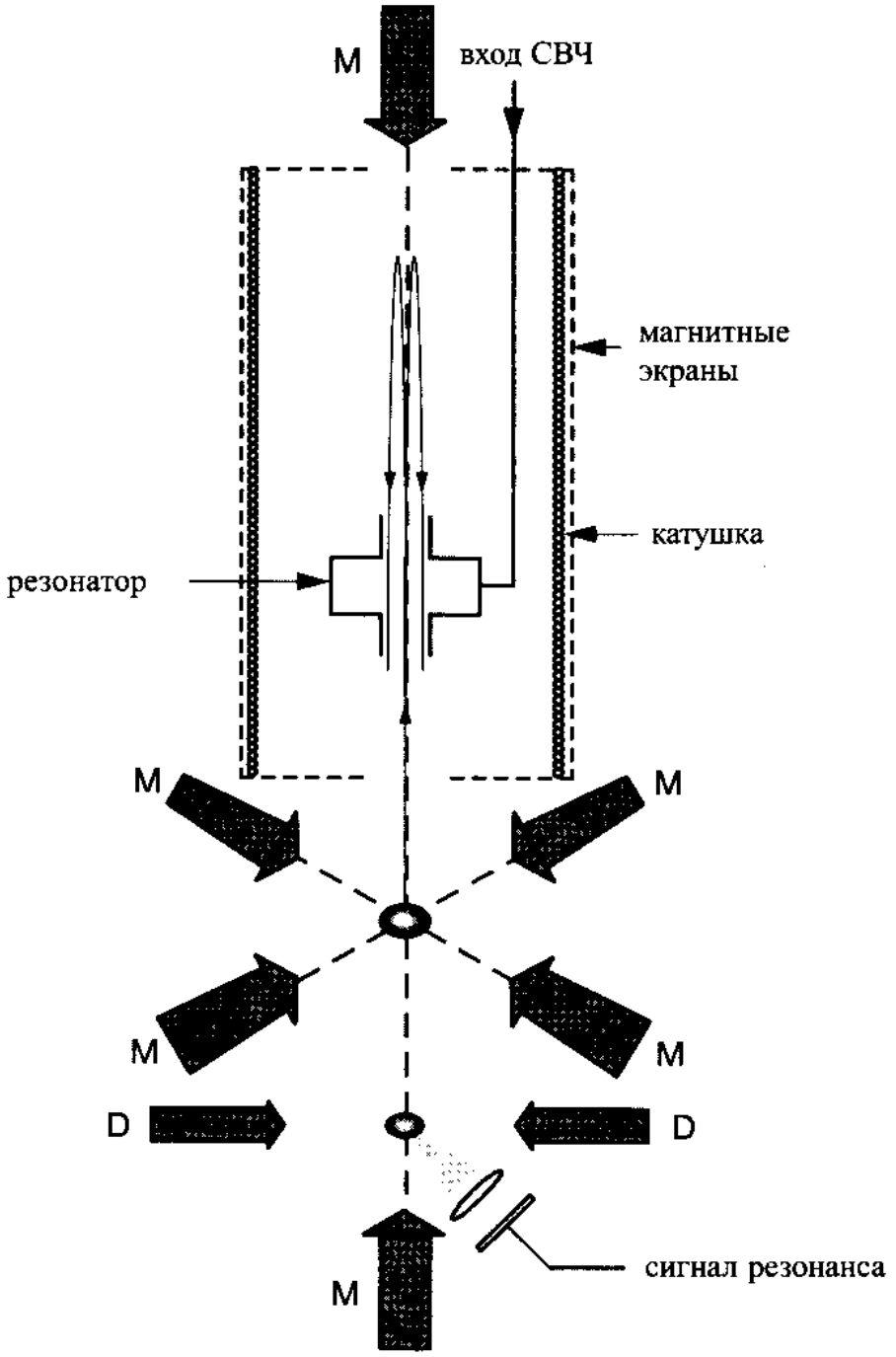


Рис. 6.21. Схематическая диаграмма цезиевого фонтана на холодных атомах. M – световые пучки, образующие оптическую трясину, а D – световой пучок для детектирования сверхтонкого резонанса.

параболической траектории, находящейся примерно на 30 см выше резонатора, и затем падают обратно, пересекая осциллирующее поле во второй раз. Время пролета между двумя пересечениями осциллирующего поля составляет около 0,5 с. В осевой области резонансной полости, которую пересекают атомы, магнитное СВЧ поле ориентировано вертикально. Все пространство, включающее в себя резонансную полость и зону, в которой атомы движутся обратно после достижений высшей точки, защищается магнитными экранами, выполненными из нескольких слоев

материала с очень высокой магнитной проницаемостью. Соленоид создает вертикальное статическое магнитное поле, которое и стабильно и однородно. Значение B_0 мало и близко к 10^{-7} Т.

Атомы падают до тех пор, пока они не достигают области детектирования, расположенной под точкой запуска. Так как они имеют ненулевую поперечную скорость, то не все они попадают в область детектирования. Типичное число атомов составляет порядка 10^5. Те атомы, которые совершили часовой переход, находятся теперь на сверхтонком подуровне $F = 4$, $m_F = 0$ основного состояния. Они детектируются в стоячей волне, образованной двумя световыми пучками, помеченными буквой D, настроенными на циклический переход $F = 4 \rightarrow F' = 5$.Число собранных фотонов пропорционально вероятности совершения часового перехода. Специальная процедура применяется для того, чтобы уменьшить изменения детектирующего сигнала из-за флуктуаций числа атомов, содержащихся в следующих друг за другом облаках.

Атомный резонанс детектируется следующими один за другим циклами. Информация, касающаяся отклика резонатора собирается примерно каждую секунду. Рис.6.22 показывает резонансную кривую, полученную путем изменения частоты возбуждения на 0,1 Гц от одного запуска к другому [6.21]. В этом эксперименте время пролета между двумя осциллирующими полями равно 0,5 с. Ширина центральной интерференционной полосы на полувысоте равна 1 Гц, в согласии с (6.19). Ширина пьедестала Раби, зависящего от того, как долго атомы пересекают осциллирую-

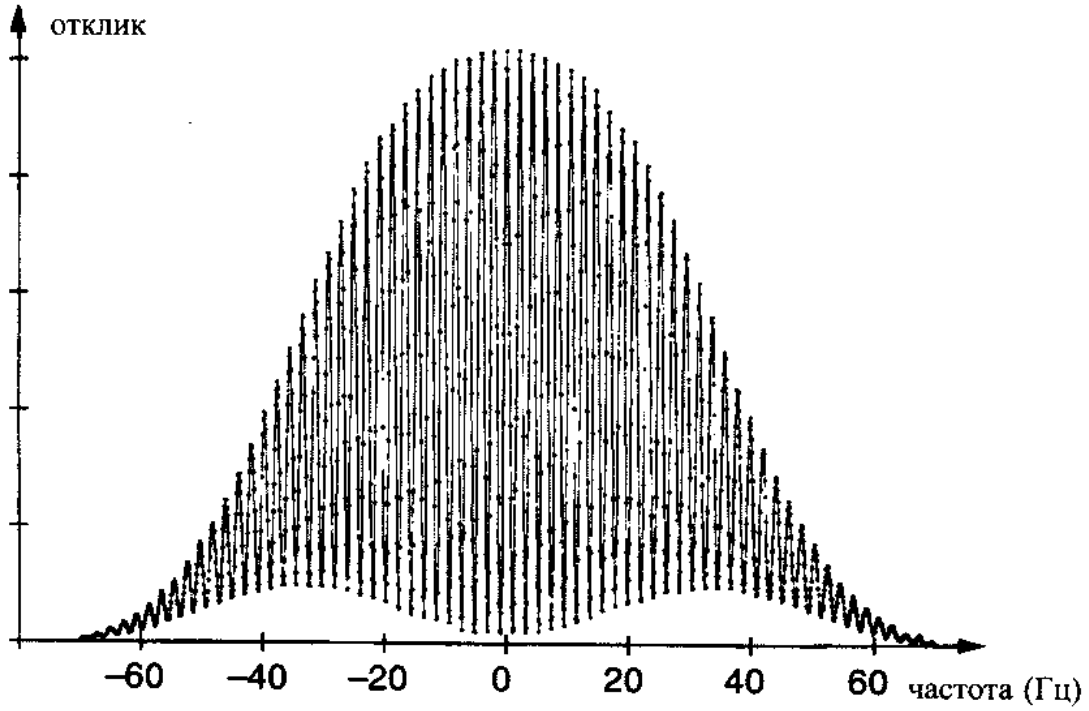


Рис. 6.22. Отклик цезиевого фонтана на холодных атомах. Начало шкалы частот расположено на частоте сверхтонкого перехода в основном состоянии. Взято из [6.21].

щее поле, составляет 60 Гц. Осцилляции Рэмси видны по всему исследуемому диапазону, потому что все атомы имеют почти одну и ту же скорость. Центральная осцилляция описывается формулой (6.17).

Автоматическая подстройка осциллятора по частоте атомного резонанса. Так как отклик атомного резонатора доступен только как последовательность, дискретная во времени, то метод автоматической подстройки осциллятора к атомному опорному резонансу должен быть приспособлен к этой последовательности. Частота возбуждения 9,192...ГГц синтезируется из сигнала кварцевого генератора, который должен подстраиваться по атомному резонансу. Его частота модулируется прямоугольной волной. Во время одного цикла измерений, когда приготовленное облако следует по своей баллистической траектории, частота возбуждения равна $\nu_i + \nu_m$, где ν_m – глубина модуляции, равная полуширине резонанса на полувысоте. В конце этого цикла число детектированных фотонов равно N^+. В следующем цикле частота возбуждения равна $\nu_i - \nu_m$ и детектируется отклик через число фотонов N^-. Разность $N^+ - N^-$ в среднем равна нулю, если ν_i равна частоте сверхтонкого перехода основного состояния атома цезия. Если разность не равна нулю, то она является сигналом ошибки, который затем приходит на цифровой интегратор. Его выход действует на частоту кварцевого генератора и таким образом осуществляется подстройка частоты кварцевого генератора по частоте атомного резонансного перехода.

(c) Нестабильность частоты

Нестабильность частоты за средние времена для этого типа стандартов определяется флуктуациями в числе детектированых фотонов. Можно показать, что вариация Аллана для относительных частотных флуктуаций, определенная в разделе 5.2.6, дается следующим выражением

$$\sigma_y(\tau) = \frac{\sqrt{2}}{\pi Q_{at}} \frac{\sigma_{\nabla N}}{N} T_c^{1/2} \tau^{-1/2}. \qquad (6.86)$$

Параметр Q_{at} является добротностью атомного резонанса, $\sigma_{\nabla N}$ – стандартное отклонение разности $\delta N^+ - \delta N^-$ между флуктуациями в N^+ и N^-, N – среднее значение величин N^+ и N^-, и T_c – длительность одного цикла.

Добротность имеет очень высокое значение, около 10^{10}, потому что используются медленные атомы, и отсюда ясно подразумевается возможность достижения исключительно малой нестабильности. В 1997 году измеренная величина нестабильности на прототипе прибора была равна $1{,}5\times10^{-13}\tau^{-1/2}$ [6.22]. Рис.6.11 показывает, что эта нестабильность меньше, чем у любых других типах цезиевых часов. Фактически в этих измерениях нестабильность частоты определялась спектральными свой-

ствами осциллятора, частота которого автоматически подстраивалась по атомному резонансу. Происхождение этого ограничивающего эффекта связано с нелинейными эффектами и импульсным режимом работы. Из-за импульсного режима в сигнале ошибки появляются высокочастотные фурье-компоненты, взаимодействующие со спектральными компонентами осциллятора. Однако объяснение этого эффекта выходит за рамки этой книги.

Эта нестабильность частоты была уменьшена после замены кварцевого генератора на генератор с улучшенной спектральной чистотой [6.23] и было достигнуто значение нестабильности, предсказываемое (6.86). Резонатор этого улучшенного генератора представляет собой сапфировое кольцо, охлажденное до температуры жидкого гелия [6.50]. Частотная нестабильность стандарта в этом случае составляет $4\times10^{-14}\tau^{-1/2}$. Это наименьшая величина, когда-либо полученная в первичных стандартах частоты. Из нее следует, что нестабильность частоты за $\tau = 1$ день может быть близкой к 10^{-16}.

(d) Сдвиги частоты и неточности

Сдвиги частоты, зависящие от скорости атомов. Так как атомы движутся очень медленно, то сдвиг частоты из-за эффекта Доплера второго порядка уменьшается примерно до 10^{-16}. Его неопределенность очень мала, так как скорости атомов известны с очень высокой точностью на всех их траекториях.

Разность фаз, которая может существовать в двух плечах резонансной полости, больше не представляет какой либо проблемы в фонтанных часах. Действительно, теперь мы имеем единственное осциллирующее поле, которое дважды пересекается атомами. Единственный остающийся здесь эффект – это пространственные изменения фазы осциллирующего поля внутри резонансной полости. Так как траектории атомов расходятся, то атомы подвергаются действию поля не с одной и той же фазой во время двух прохождений резонатора. При условии низких скоростей атомов остаточный эффект Доплера первого порядка оценивается величиной 2×10^{-16}.

Сдвиги частоты, зависящие от добротности атомного резонанса. В фонтане холодных атомов цезия добротность атомного резонанса Q_{at} в сотни раз выше, чем в лабораторных стандартах, работающих на тепловых пучках. Сдвиг частоты, зависящий от Q_{at}, по этой причине сильно подавлен. Он, в частности, проявляется в затягивании частоты резонатором и меняется пропорционально Q_{at}^{-2}. Этот эффект может быть уменьшен до уровня 10^{-16}. Неопределенность в этом сдвиге частоты становится пренебрежимо малой.

Сдвиги частоты, зависящие от соседних переходов и от статического магнитного поля. Так как ширина пьедестала Раби значительно уменьшается, а оптическая накачка гарантирует высокий уровень симметрии СВЧ спектра цезия, то соседние переходы могут быть расположены ближе к часовому переходу, если мы уменьшим значение магнитного поля B_0 примерно до 10^{-7} Т. К тому же однородность магнитного поля может быть проверена путем подбрасывания облака атомов на разные высоты. Неопределенность в сдвиге частоты из-за эффекта Зеемана второго порядка может быть уменьшена до несколько единиц 10^{-16}.

Сдвиги частоты из-за столкновений. В облаке холодных атомов плотность атомов незначительно отличается от плотности в типовых цезиевых пучках. Однако сечение спин-обменных взаимодействий, приводящих к сдвигу частоты, возрастает при уменьшении температуры [6.51]. Этот сдвиг частоты можно измерить путем изменения плотности атомов [6.21, 6.22, 6.52]. При типичных условиях сдвиг имеет значение 10^{-14}. Экстраполяция к нулевой плотности исключает эффект сдвига частоты из-за столкновений. Результирующая неопределенность измерений этого сдвига в настоящее время составляет меньше 10%. Сдвиг частоты из-за столкновений может быть уменьшен. Путем разработки более тонких методов охлаждения, чем представленные в разделе 6.4.3а, поперечные скорости атомов можно уменьшить так, что в любой части траектории плотность атомов ненамного превосходит остаточную плотность в зоне детектирования. Эта остаточная плотность является величиной, определяющей долговременную нестабильность. Другой способ уменьшения сдвига – это лучшая подготовка атомов путем удаления атомов с подуровнями $m_F \neq 0$, которые дают вклад в столкновения, но не дают вклада в сигнал.

Другие сдвиги частоты. Сдвиги частоты, не зависящие от структуры установки, такие как сдвиг из-за излучения черного тела, могут быть оценены с неопределенностью ниже 10^{-15}. Сдвиги частоты инструментального происхождения, такие как, например, сдвиги из-за переходов Майорана или из-за наличия нежелательных спектральных компонент в СВЧ сигнале возбуждения могут быть достаточно хорошо определены, чтобы их остаточное значение было учтено на уровне ниже 10^{-15}.

Неточности. Путем создания фонтанов на холодных атомах цезия был осуществлен впечатляющий рывок в реализации идеи изолированного атома в покое. Был достигнут уровень неточности 10^{-15} [6.23]. Дальнейшие совершенствования, вероятно, подведут нас, в пределах нашего понимания, близко к значению 10^{-16}. Многие метрологические лаборатории разрабатывают конструкции фонтанов на холодных атомах цезия с целью достижения наивысших точностей.

(е) Космические цезиевые часы с лазерным охлаждением

Для того чтобы увеличить добротность Q_{at} атомного резонанса и, следовательно, уменьшить нестабильность частоты и в то же время понизить вариации в сдвигах частоты, зависящих от добротности, необходимо увеличивать время пролета T между двумя осциллирующими полями. В условиях гравитации высота фонтана, как это можно увидеть из (6.85), должна быть увеличена в 4 раза, чтобы удвоить значение Q_{at},. При этом могут появиться ограничения, связанные, например, с неоднородностью магнитного поля. С другой стороны в условиях отсутствия гравитации при данном расстоянии L между осциллирующими полями пролетное время T зависит только от начальной скорости пучка атомов v_0. Поэтому мы можем получить большие значения T при разумных значениях L (например, $L = 0{,}5$ м) путем уменьшения начальной скорости (например, $v_0 = 10$ см $с^{-1}$). Такая ситуация может быть реализована на борту спутника. И атомы, и оболочка их содержащая, свободно падают одинаковым путем так, что после запуска атомы движутся равномерно и прямолинейно в системе координат, связанной с часами. Существует проект по размещению на орбите часов на холодных атомах. Он называется Фарао по начальным буквам названия проекта на французском языке. Цель этого проекта – изучение поведения цезиевых часов на холодных атомах в условиях микрогравитации. В этот проект вовлечено несколько французских исследовательских центров под общим патронажем национального агентства по изучению космоса (CNES) Франции.

Прототип таких часов был построен и успешно испытан в условиях уменьшенной гравитации на реактивном самолете во время полета по параболической траектории [6.53]. Ожидается, что в космосе неточность и нестабильность частоты за $\tau = 1$ день будут составлять около 10^{-16}.

Такие часы и водородный мазер [6.54] возможно будут установлены на международной космической станции в 2006 году для выполнения проекта европейского космического агентства ACES (ансамбль атомных часов в космосе) [6.55]. СВЧ и лазерные линии связи обеспечат средства сличений времени и частоты с наземными станциями.

Цели проекта ACES включают в себя улучшенные измерения гравитационного сдвига, лучшую проверку изотропности скорости света и поиск возможных вариаций постоянной тонкой структуры α.

Аналогичные проекты разрабатываются в США [6.56, 6.57].

6.4.4. Рубидиевые часы с лазерным охлаждением

С атомами рубидия 87 можно осуществлять оптическую накачку точно также как и с атомами цезия 133, потому что также существуют полупроводниковые лазеры на подходящей длине волны (см. таблицу

6.1). Интерес к созданию фонтана на холодных атомах рубидия проявился с того момента, когда было предсказано, что столкновительный частотный сдвиг в сверхтонком переходе рубидия 87 будет значительно меньше, чем в цезии 133 при равных плотностях [6.58]. Это очень интересное свойство экспериментально было подтверждено в работах [6.59, 6.60]. Оно обеспечивает возможность уменьшить наиболее беспокоящий частотный сдвиг и одновременно уменьшить нестабильность частоты путем некоторого увеличения плотности атомов. Вполне возможно, что может быть реализована неточность порядка 10^{-16}. В настоящее время в процессе подготовки находится эксперимент по сличению метрологических свойств фонтанов на холодных атомах цезия и на холодных атомах рубидия.

Результатом первых сличений стало измерение частоты сверхтонкого перехода в рубидии 87 с неопределенностью $1{,}3\times10^{-14}$ (см. таблицу 6.1) [6.61].

6.5. Водородные мазеры

Водородный мазер был изобретен на основе исследований, выполненных Рэмси и его группой в Гарварде, США, в конце 1950-х годов [6.24]. Первоначальная цель исследований состояла в сужении сверхтонкого резонанса в цезиевом пучке путем пропускания его через два разделенных осциллирующих поля. Прежде чем пройти через них атомы задерживались в ячейке на пути их следования таким образом, чтобы они не могли выйти из ячейки, пока они несколько раз не столкнутся со стенками. Стенки были покрыты исключительно инертными материалами, такими как фторопласт или парафин. Эксперимент оказался удачным. Временной интервал T между взаимодействиями с двумя осциллирующими полями был увеличен. Однако этот успех был компенсирован уширением сверхтонкой резонансной линии и ее уменьшением за счет самих столкновений. Кроме того, частота резонанса оказалась сдвинутой.

Потом было показано, что при таких столкновениях атомы водорода возмущаются меньше чем атомы цезия, потому что водород имеют меньшую массу и менее выраженную электрическую поляризуемость. Однако энергия ионизации атома водорода очень велика и в отличие от атома цезия его невозможно детектировать с помощью горячей проволоки. Оптические методы также очень сложны для применения, так как длины волн, относящиеся к переходам межу основным и возбужденным состояниями, лежат в области вакуумного ультрафиолета (см. таблицу 6.1). Поэтому было решено детектировать атомы водорода путем использования стимулированного излучения, как до этого было сделано в мазере на аммиаке [6.62, 6.63]. Водородный мазер был успешно построен в 1959 –1960 годах [6.64].

Водородный мазер вскоре стал широко использоваться в США как спектрометр и как очень высокостабильный стандарт частоты [6.64, 6.65]. Многие исследовательские центры в Австралии, Канаде, Китае, Франции, Японии, Западной Германии, Румынии, Великобритании, Швейцарии и СССР стали изучать и развивать водородные мазеры.

В настоящее время в мире используется несколько сотен водородных мазеров. Большинство из них было изготовлено в СССР [6.66]. Как можно видеть на рис.6.11, в своей активной автоколебательной модификации водородные мазеры являются стандартами времени и частоты с наименьшей нестабильностью частоты за время усреднения между 1 секундой и несколькими часами. (В показанном на этом рисунке мазере частота резонатора не контролируется системой автоматической настройки. При наличие системы автоматической настройки резонатора уменьшается долговременная нестабильность водородного мазера). В своей много более компактной пассивной модификации (таблица 6.2, раздел 6.9) на средних временах усреднения водородный мазер имеет типичную нестабильность, близкую к $10^{-12}\tau^{-1/2}$, т.е. меньшую, чем достигается в промышленных цезиевых пучковых стандартах. Неточность (активного или пассивного) водородного мазера не так мала, как в лабораторных пучковых стандартах (таблица 6.3). Это объясняется столкновениями со стенками колбы, содержащей атомы. Сегодня водородные мазеры используются главным образом в метрологии времени и частоты, радиоастрономии, геофизике и в экспериментальных проверках релятивистской теории.

6.5.1. Основные принципы работы водородного мазера

Частота сверхтонкого перехода основного состояния атома водорода близка к 1420 МГц (таблица 6.1) Соответствующее значение длины волны в вакууме близко к 21 см. Переход, используемый в водородном мазере, тот же самый, что и переход наблюдаемый астрономами, изучающими межзвездные облака атомов водорода. Такие наблюдения ведут среди прочего к измерению радиальной скорости компонент таких облаков, так как частота перехода сдвигается из-за эффекта Доплера.

Первоначальная конструкция водородного мазера схематически показана на рис. 6.23.

В радиочастотной разрядной лампе молекулярный водород диссоциирует на атомы. Этот *источник атомов водорода* присоединяется одним или несколькими капиллярными трубками к высоковакуумной области. Таким образом, из атомов водорода формируется коллимированный поток круглого сечения. На выходе из источника все четыре сверхтонких уровня (рис. 6.3) заселены равномерно.

Разделение по состояниям осуществляется магнитным отклонением (раздел 6.2.2). Атомный пучок пересекает *многополюсный магнит*. Он состоит из четырех или шести полюсов, расположенных через равные угловые интервалы вокруг оси пучка и намагниченных попеременно – северный, южный. Полюса удалены примерно на 1 мм от оси. В поперечном сечении создается магнитное поле. В первом приближении его модуль не зависит от полярного угла и постоянно растет от нуля на оси до величины, близкой к 1 Т около полюсов. Поле отклоняет атомы в состояниях $F=1$, $m_F=0$ и $F=1$, $m_F=1$ в направлении оси, в то время как атомы в состояниях $F=1$, $m_F=-1$ и $F=0$, $m_F=0$ выталкиваются от оси. Трубка в *накопительной колбе* служит как диафрагма и можно считать, что выбираются только атомы в состояниях $F=1$, $m_F=0$ и $F=1$, $m_F=1$. Среди них атомы в состоянии (1, 0) играют основную роль. Поток атомов, входящих в колбу, составляет от 10^{12} до 10^{13} атомов в секунду. Соответствующее парциальное давление составляет порядка 10^{-5} Па. Накопительная колба выполняется из кварца, чтобы обеспечить малые диэлектрические потери. Она обычно имеет объем от 1 до 2 л. Внутренняя поверхность стенок покрывается фторосодержащими полимерами типа фторопласт. В этом случае атом водорода может испытать от 10^4 до 10^5 столкновений со стенками, прежде чем его волновая функция подвергнется значительным возмущениям. Практически это означает, что достигается среднее время удержания близкое к одной секунде.

Накопительная колба располагается в центре цилиндрической *резонансной полости,* ось которой совпадает с осью атомного пучка. Резонатор настроен на частоту атомного перехода. Для конструкции, показанной на рис. 6.23, диаметр и длина резонатора составляют около 27 см.

Общая форма силовых линий СВЧ магнитного поля показана тонкими линиями. Атомы удерживаются около центра резонансной моды в области, где фаза магнитного поля постоянна. Это значит, что частота перехода не подвержена действию эффекта Доплера первого порядка (раздел 6.2.3) и эффективная ширина атомного резонанса по порядку величины составляет 1 Гц.

Соленоид, ось которого совпадает с осью резонатора, создает постоянное однородное поле около 10^{-7} Т. Если СВЧ магнитное поле примерно параллельно статическому магнитному полю в той области где находятся атомы, то согласно правилам отбора разрешен часовой переход

$$F=1,\ m_F=0 \leftrightarrow F=0,\ m_F=0.$$

Водородный мазер имеет также и магнитные экраны, сделанные из четырех или пяти слоев материала с очень высокой магнитной проницаемостью. Они заключают в себе резонансную полость и уменьшают вариации окружающего магнитного поля в 10^4 – 10^5 раз.

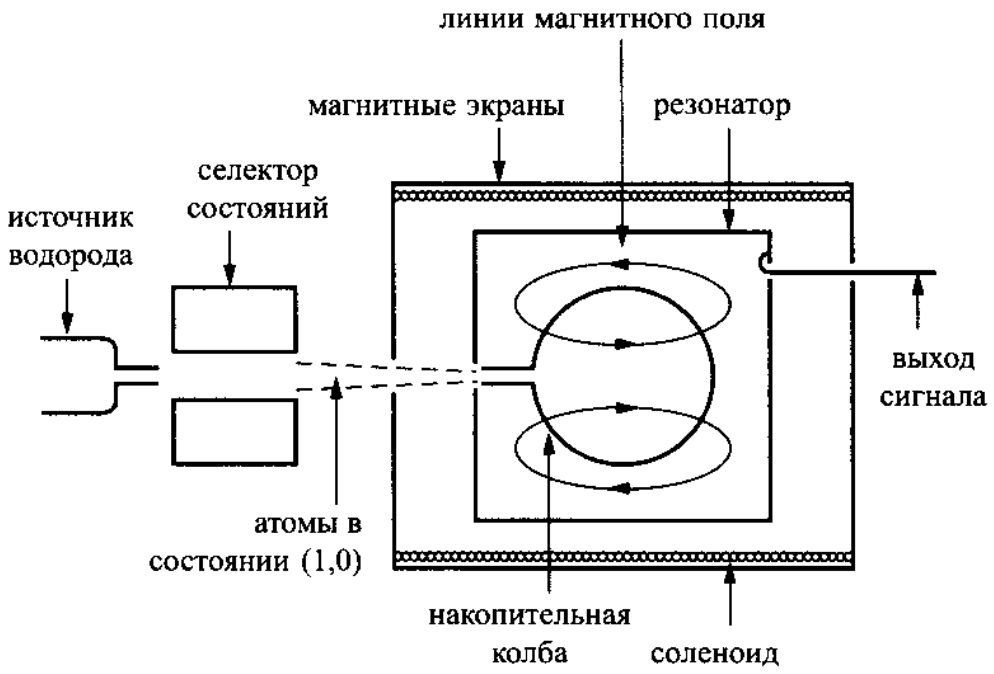


Рис. 6.23. Схематическая диаграмма водородного мазера. Атомный пучок обычно расположено вертикально

В большинстве современных приборов ось, параллельная атомному пучку, располагается вертикально.

6.5.2. Условия генерации

При входе в область накопительной колбы поток атомов в состоянии F =1, m_F = 0 значительно превышает поток атомов в состоянии F = 0, m_F = 0. Так как энергетический уровень (1, 0) выше, чем уровень (0, 0) (рис. 6.2), то атомный пучок приносит энергию в резонансную полость. Атомы таким образом способны усиливать СВЧ поле, уже присутствующее в резонаторе, при условии, что его частота очень близка к частоте сверхтонкого перехода. Они также могут поддерживать автоколебания, если энергия, вносимая атомным пучком, достаточна для того, чтобы компенсировать потери в стенках резонансной полости.

Давайте сравним поток энергии, доставляемый в резонатор атомным пучком, с ухудшающей условия генерации мощностью, рассеиваемой в резонаторе.

(а) Мощность, рассеиваемая в резонаторе

Согласно общим результатам теории электромагнетизма мощность P_d, рассеиваемая в любой резонирующей системе на круговой частоте ω_c связана с накопленной энергией W через

$$P_d = \omega_c \frac{W}{Q_c}, \tag{6.87}$$

где Q_c – добротность резонатора. Запасенная энергия равна удвоенной магнитной энергии в резонаторе. Так как магнитное поле синусоидально, то усредненная по времени запасенная энергия равна

$$W = \frac{1}{2\mu_0} \int_V B^2 \mathrm{d}\upsilon, \tag{6.88}$$

где интегрирование производится по всему объему резонатора. Величина μ_0 – это магнитная проницаемость вакуума, а B – амплитуда магнитного поля. Амплитуда магнитного поля является пространственно переменной функцией, но мы можем написать результат интегрирования в сжатом виде

$$W = V_c \frac{\langle B^2 \rangle_c}{2\mu_0}, \tag{6.89}$$

где V_c –объем резонатора и $\langle B^2 \rangle_c$ –среднеквадратичное значение амплитуды магнитного поля по объему резонатора. Это подразумевает, что

$$P_d = \frac{\omega_c V_c \langle B^2 \rangle_c}{2\mu_0 Q_c}. \tag{6.90}$$

(b) Мощность, генерируемая атомным пучком

Предположим, что в резонаторе установилось электромагнитное поле. Из-за своего беспорядочного движения в накопительной колбе атомы усредняют пространственную зависимость компонент СВЧ магнитного поля. Все происходит так, как будто они подвергаются действию магнитного поля с постоянной амплитудой и напряженностью $\langle B_z \rangle_b$, где B_z компонента поля B, параллельная статическому магнитному полю. Эта компонента может индуцировать часовой переход ΔF =1, Δm_F = 0. Усреднение выполняется по всему объему накопительной колбы.

Далее, частота Раби для атомов, как уже было рассмотрено в разделе 6.4.1b, определяется из

$$b = \mu_B \frac{\langle B_z \rangle_b}{\hbar}. \tag{6.91}$$

В дальнейшем мы должны будем рассмотреть также *коэффициент заполнения*, определяемый из следующей формулы

$$\eta = \frac{\langle B_z \rangle_b^2}{\langle B^2 \rangle_c}. \tag{6.92}$$

Этот коэффициент характеризует в резонаторе степень связи между облаком атомов и электромагнитным полем.

Приблизительное выражение вероятности перехода. Если мы на момент предположим, что в водородном мазере квантовые состояния атомов могут быть представлены вектором состояний $|\Psi\rangle$, введенном в разделе 6.4.1b, то можно использовать результаты этого раздела.

В соответствии с взаимодействием длительностью θ вероятность совершения перехода из энергетического состояния E_2 в энергетическое состояние E_1, $E_1 < E_2$, определяется из

$$P(\theta) = |\gamma_2(\theta)|^2. \tag{6.93}$$

Мы имеем

$$\begin{pmatrix} \gamma_1(\theta) \\ \gamma_2(\theta) \end{pmatrix} = \mathcal{M}^{(1)} \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{6.94}$$

где при резонансе $\mathcal{M}^{(1)}$ дается формулой (6.39). Мы получаем

$$P(\theta) = \sin^2 \frac{b\theta}{2}. \tag{6.95}$$

Вероятность того, что атом прекратит излучение или из-за ухода из накопительной колбы, или из-за столкновения со стенкой или с другими атомом, представлена феноменологически путем введения экспоненциального распределения для времени взаимодействия [6.64]. Полагая

$$f(\theta) = \frac{1}{T_0} \exp - \frac{\theta}{T_0}, \tag{6.96}$$

где T_0 – постоянная времени, вероятность, усредненная по времени взаимодействия дается выражением:

$$P = \int_0^\infty f(\theta) P(\theta) \mathrm{d}\theta. \tag{6.97}$$

Затем мы получим

$$P = \frac{1}{2} \frac{T_0^2 b^2}{1 + T_0^2 b^2}, \tag{6.98}$$

этот результат является только приблизительным. Коэффициент T_0^2 должен быть, как мы увидим ниже, заменен на произведение двух констант T_1 и T_2. Появление этих двух констант является результатом стандартной интерпретации экспериментов по магнитному резонансу в ограниченных газах, жидкостях или твердых телах.

Точное выражение для вероятности переходов. Точные свойства находящихся в накопительной колбе атомов водорода вывести не так просто как для атомов цезия в цезиевом пучковом резонаторе. Действительно, столкновения атомов водорода со стенками и друг с другом возмущают их вектор состояний до такой степени, что описание с помощью этого вектора состояний становится не адекватным. Мы должны обратиться к понятию матрицы плотности [6.25], которая остается до некоторой степени за пределами представления в этой книге. Поэтому мы только процитируем основные результаты [6.4, 6.65].

Точное выражение вероятности перехода

$$F = 1, m_F = 0 \rightarrow F = 0, m_F = 0$$

для данного резонанса дается

$$P = \frac{1}{2}\frac{T_1 T_2 b^2}{1 + T_1 T_2 b^2}. \tag{6.99}$$

Постоянные времени T_1 и T_2 характеризуют разность населенности и когерентность состояний (1, 0) и (0, 0) соответственно. Они зависят от эффектов столкновений со стенками и с газовой фазой и от потока атомов через канал накопительной колбы. Типичные значения этих так называемых *времен релаксации* составляют $T_1 = 0{,}3$ с и $T_2 = 0{,}5$ с.

Когда амплитуда электромагнитного поля возрастает и увеличивается значение b, вероятность P стремится к 1/2. Поэтому населенности энергетических уровней E_1 и E_2 стремятся уравняться. Это уравнивание определяется коэффициентом $T_1 T_2 b^2$ в правой стороне (6.99), который называется *коэффициентом насыщения.* Обычно для водородных генераторов коэффициент насыщения близок к единице.

Мощность, доставляемая атомным пучком. Каждый атом, производящий переход с уровня энергии E_2 на уровень E_1, высвобождает энергию $\hbar\omega_0$. Мощность P_{at} освобождаемая атомным пучком, таким образом, равна

$$P_{\text{at}} = \frac{1}{2} I \hbar \omega_0 \frac{T_1 T_2 b^2}{1 + T_1 T_2 b^2}. \tag{6.100}$$

(c) Условия генерации

Если мощность, доставляемая атомным пучком, достаточна для поддержания электромагнитного поля в резонаторе, то мощность, рассеиваемая в нем, равна мощности в устойчивом состоянии.

Если частота настройки резонатора равна частоте атомного резонанса, то мы получим из (6.90) и (6.100), используя (6.92) для коэффициента заполнения

$$P_{at} = P_d = \frac{1}{2}\hbar\omega_0(I - I_t), \tag{6.101}$$

где

$$I_t = \frac{\hbar V_c}{\mu_0 \mu_B^2 \eta Q_c T_1 T_2}. \tag{6.102}$$

Уравнение (6.101) показывает, что величина P_d может быть только положительной, и что, таким образом, генерация может поддерживаться, если

$$I > I_t. \tag{6.103}$$

Это неравенство выражает условия генерации. Величина I_t является пороговой величиной атомного потока.

Для типичных практических потоков при V_c = 15 дм^3, η=2,5, Q_c =3,5×10^4, T_1 = 0,3 с и T_2= 0,5с мы находим, что I_t =1,1×10^{12} атомов в секунду.

6.5.3. Активный водородный генератор

(а) Уровень генерации

Несмотря на выражение (6.101) рассеиваемая в резонаторе мощность не пропорциональна атомному потоку I. Фактически I_t также зависит от этого потока. Из-за столкновений между атомами водорода T_1 и T_2 являются затухающими функциями от атомной плотности и, следовательно, от I. К тому же рассеиваемая мощность зависит от добротности резонатора, так как I_t обратно пропорциональна Q_c. В работе [6.65] было установлено каким образом P_d зависит от I и Q_c. Рис. 6.24 показывает зависимость I от различных значений Q_c.

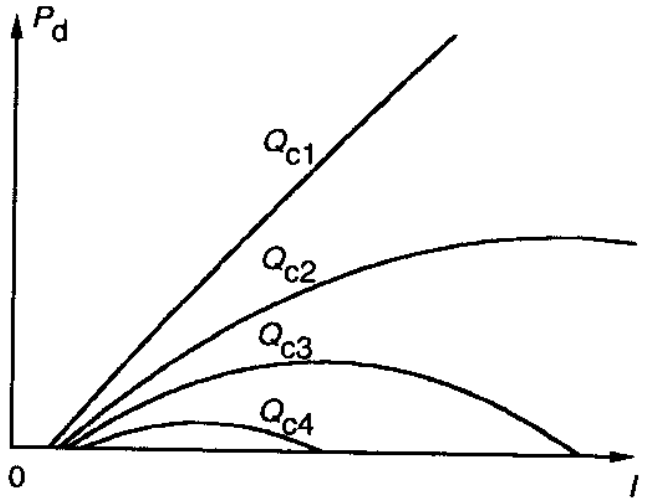


Рис. 6.24. Мощность, рассеиваемая в резонаторе, как функция входящего потока атомов для $Q_{c1} > Q_{c2} > Q_{c3} > Q_{c4}$

Из рисунка видно, что так как P_d может быть только положительной, то условия генерации будут выполняться только при достаточно высокой добротности резонатора. Действительно, малые значения Q_c приводят к большим значениям I_t, которые также возрастают для меньших значений T_1 и T_2. Следовательно, в таком случае не существует атомного потока I, который бы удовлетворял условиям генерации (6.103). Минимальное значение добротности составляет около 2×10^4.

В активном водородном генераторе типичная добротность резонатора составляет около $3{,}5\times10^4$, а атомный поток примерно в три раза превышает пороговый поток. Мощность, рассеиваемая в резонаторе, близка к 10^{-12} Вт. Мощность, связанная с выходными цепями составляет примерно одну десятую от этого значения.

(b) Частота генерации

Активный водородный генератор является прибором, который генерирует сигнал на сверхвысокой частоте, которая определяется главным образом атомным переходом. Эта частота слабо зависит от настройки частоты резонатора.

Здесь ситуация совершенно отличается от ситуации с цезиевыми часами. В цезиевых часах благодаря очень малому взаимодействию с каждым осциллирующим полем электромагнитное поле в резонаторе остается практически неизменным при прохождении атомного пучка. В водородном генераторе атомная среда и электромагнитное поле сильно связаны, так как именно присутствие атомов поддерживает это поле. Поэтому следует ожидать более сильной зависимости частоты генерации от настройки частоты резонатора.

Частота генерации не может быть выведена из вышеприведенных соотношений. Для этого необходимо рассмотреть индуцированные полем резонатора магнитные моменты в атомной среде.

Используем комплексные выражения обозначив через $B(\omega)\exp(-i\omega t)$ магнитное поле, действующее на атомы. Модуль $B(\omega)$ равен $\langle B_z\rangle_b$, введенному ранее. Квантовое теоретическое рассмотрение атомов в накопительной колбе приводит к комплексной амплитуде магнитного момента этого газа, а именно:

$$M(\omega)=\mu_B T_1 T_2 I\,\frac{-i[1+iT_2(\omega-\omega_0)]}{1+T_1T_2b^2+T_2^2(\omega-\omega_0)^2}\,\frac{\mu_B B(\omega)}{\hbar}. \qquad (6.104)$$

Этот магнитный момент очень селективен по частоте. Если коэффициент насыщения мал, то модуль $M(\omega)$ делится на $\sqrt{2}$ для $\omega-\omega_0=\pm\,1/T_2$. Ширина атомного резонанса W равна таким образом

$$W = \frac{2}{T_2} \tag{6.105}$$

и добротность равна $Q_{at} = \omega_0/W$, или

$$Q_{at} = \frac{\omega_0 T_2}{2}. \tag{6.106}$$

Эта добротность близка к $(1–2)\times10^9$.

Индуцированный магнитный момент, осциллирующий с угловой частотой ω, воздействует на электромагнитное поле в резонаторе. Используя уравнения Максвелла для резонатора, содержащего атомную среду, мы найдем соотношение

$$\left[\omega_c^2 - i\omega\frac{\omega_c}{Q_c} - \omega^2\right]B(\omega) = \frac{\mu_0\eta}{V_c}\cdot\omega^2 M(\omega). \tag{6.107}$$

С левой стороны мы узнаем характерный коэффициент для цепей, резонирующих на частоте ω_c с добротностью Q_c. С правой стороны исходный член пропорционален осциллирующему магнитному моменту атомов.

Введем теперь выражение (6.104) для $M(\omega)$ в (6.107) и уравняем мнимые и действительные части обеих сторон уравнения. Используя то, что угловые частоты ω, ω_c и ω_0 практически очень близки, мы получим

$$1 + T_1T_2b^2 + T_2^2(\omega-\omega_0)^2 = \frac{\mu_0\mu_B^2\eta Q_c T_1 T_2}{\hbar V_c} I, \tag{6.108}$$

и

$$\omega - \omega_0 = \frac{Q_c}{Q_{at}}(\omega_c - \omega_0). \tag{6.109}$$

При резонансе (6.108) возвращает нас назад к выражению (6.102), т.е. к тому пороговому потоку атомов, при котором возникают автоколебания.

Уравнение (6.109) выражает частоту генерации $\omega/2\pi$ в терминах атомной резонансной частоты и расстройки частоты резонатора. Оно показывает, что при данной расстройке сдвиг частоты генерации водородного мазера пропорционален отношению добротностей Q_c/Q_{at}. Отметим, что в цезиевом пучковом резонаторе этот сдвиг частоты перехода пропорционален квадрату этого отношения.

Уравнение (6.109) дает также точный метод настройки частоты резонатора. Добротность атомного резонанса может изменяться путем изменения интенсивности пучка. Под влиянием столкновений вместе с изменением плотности атомов в накопительной колбе изменяется время

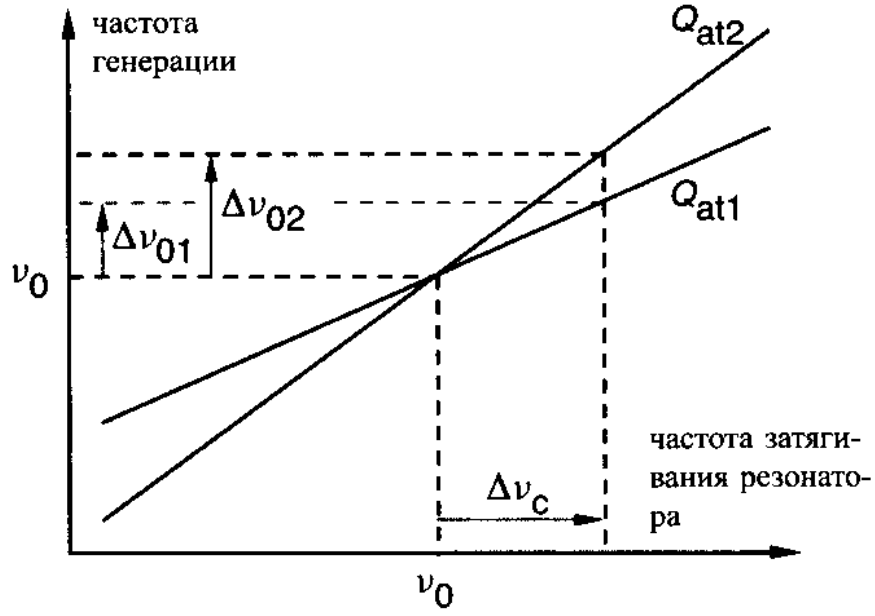


Рис. 6.25. Настройка частоты резонатора путем изменения добротности атомного резонанса. $Q_{at1} > Q_{at2}$.

релаксации T_2. Поэтому для данной $\omega_c - \omega_0$ частота генерации зависит от значения Q_{at} если $\omega_c - \omega_0 \neq 0$.

Рис. 6.25 иллюстрирует применение этого метода для настройки частоты резонатора. Если добротность атомного резонанса изменяется с Q_{at1} на Q_{at2}, то при расстройке резонатора $\Delta\nu_c$ частота генерации изменяется как $\Delta\nu_{02} - \Delta\nu_{01}$ и пропорциональна $\Delta\nu_c$. Если мы получим $\Delta\nu_{02} - \Delta\nu_{01} = 0$ для любого значения Q_{at}, то резонатор настроен правильно. Настройка может осуществляться вручную или автоматически (раздел 6.5.3e).

Настройка резонатора должна быть очень точной. Относительный сдвиг частоты генерации из-за расстройки резонатора равен

$$\frac{\Delta\nu_{osc}}{\nu_0} = \frac{Q_c}{Q_{at}} \frac{\nu_c - \nu_0}{\nu_0}. \qquad (6.110)$$

Для того чтобы получить $\Delta\nu_{osc}/\nu_0 = 3{,}5\times10^{-15}$ при $Q_c = 3{,}5\times10^4$ и $Q_{at} = 10^9$ относительная расстройка резонатора не должна превышать $(\nu_c - \nu_0)/\nu_0 = 10^{-10}$. Поэтому резонатор должен конструироваться и изготавливаться с наибольшей тщательностью. Он должен поддерживаться при неизменной температуре в пределах милликельвин. Настройка частоты резонатора часто осуществляется системой электронной автоподстройки (раздел 6.5.3e).

(c) Практическая конструкция водородного генератора

Некоторые основные указания как построить водородный генератор, были уже даны, когда мы рассматривали различные его части. Мы упоминали

источник атомов, многополюсные магниты, накопительную колбу, статическое магнитное поле и магнитную экранировку. Цель этой части дать краткое практическое описание молекулярного источника водорода, системы откачки газов, резонатора и различных электронных цепей, необходимых для обеспечения правильной работы водородного генератора.

Источник молекулярного водорода. Источник молекулярного водорода представляет собой либо объем со сжатым газом, либо металлический гидрид, выделяющий водород при нагреве.

Система откачки газов. Так как непрерывно используется поток водорода, то требуется система откачки. Она включает в себя один или несколько ионных насосов с производительностью 100 л $с^{-1}$, либо несколько насосов, содержащих материал, такой как титановый порошок, который эффективно поглощает водород. Вместе с ними устанавливаются маленькие ионные насосы, которые откачивают другие газы, не поглощаемые титаном.

Резонатор. Применяется несколько методов изготовления резонатора примерно с одинаковыми результатами в каждом случае.

Один из них близок к схематическому изображению на рис. 6.23 [6.67]. Диаметр и высота резонатора примерно равны (27 см), что сводит к минимуму потери в стенках. Практическая добротность в вакууме достигает 50 000. Накопительная колба сферическая с диаметром около 16 см. Резонатор изготавливается из материала с очень низким коэффициентом температурного расширения, таким как плавленый кварц или стеклокерамика. Слой серебра, напыляемый на внутренние стенки, ограничивает объем, доступный для электромагнитных волн. Система термостабилизации действует не непосредственно на стенки резонатора, а на стенки вакуумной камеры и на другие элементы конструкции. Время установления теплового равновесия велико, от 30 до 40 часов.

В другой конструкции высота резонатора примерно в полтора раза больше диаметра. Форма накопительной колбы тоже до некоторой степени вытянута [6.68]. Это позволяет улучшить наполнение резонатора атомной средой. Термостабилизация в этом случае должна быть очень эффективна. Она действует непосредственно на стенки резонатора в вакууме. Другие стенки вне вакуума также термостабилизируются. Остаточные флуктуации температуры резонатора находятся на уровне 1 мкК. Тепловая постоянная времени меньше чем в предыдущем случае и составляет около 10 часов.

С 1980 годов одним из главных вопросов в конструкции водородных генераторов было уменьшение объема резонатора и, следовательно, объема всего прибора. В настоящее время накопительная колба помещается в диэлектрическом цилиндре. Если введенные таким образом дополнитель-

ные потери малы, то поддерживаются условия автогенерации. Две различные конструкции объединяются этой особенностью. В одной цилиндр выполняется из кварца толщиной в несколько миллиметров. Это недорогое решение снижает объем резонатора на 50% [6.69]. В другой цилиндр толщиной 10 – 15 миллиметров изготавливается из синтетического сапфира. Этот материал имеет диэлектрическую постоянную около 10 и очень малые диэлектрические потери. Поэтому объем резонатора уменьшается примерно в 4 раза [6.70]. Во всех этих конструкциях резонатор снабжен двумя петлями связи. Одна используется для выхода генерируемой мощности, а другая содержит варактор для тонкой подстройки резонатора.

Сопутствующие электронные приборы. В источнике атомов водорода разряд поддерживается радиогенератором, имеющим выходную мощность около 5 Вт на частоте примерно 100 МГц.

Поток атомов водорода поддерживается на постоянном уровне путем автоматической поддержки давления в разрядной трубке и в связанном с ней объеме. Давление устанавливается путем изменения потока молекулярного водорода, поступающего в атомный источник. В настоящее время используется свойство никеля или палладий-серебряного сплава становиться более проницаемыми для водорода при повышении их температуры.

Объем резонатора поддерживается при температуре 45 – 50 °С с отдельным контролем различных частей водородного генератора. Используется по крайней мере десяток различных цепей. Термодатчики и термисторы выбираются исходя из стабильности их характеристик и включаются в мост Уитсона. Любой разбаланс моста приводит к изменению мощности нагревательных элементов, расположенных по поверхности, которая должна стабилизироваться по температуре.

Другие электронные цепи обеспечивают постоянный ток соленоида и напряжение подстройки на варикапе, который подстраивает частоту резонатора.

(d) Управление частотой кварцевого генератора по частоте генерации мазера

В автоколебательном режиме водородный мазер является активным стандартом частоты. Однако сигнал, который он генерирует, не может быть использован непосредственно главным образом из-за того, что его мощность слишком мала, порядка 10^{-13} Вт. К тому же его частота должна быть преобразована к частоте, которая может быть использована непосредственно потребителем.

Частота кварцевого генератора может контролироваться по частоте мазера. Решение этой задачи опирается на очень хорошо разработанную технику гетеродинного приема и технику фазовой автоподстройки.

На рис.6.26 показана упрощенная схема работы электроники. Высококачественный кварцевый генератор производит сигнал на частоте 5, 10 или 100 МГц. Пусть ν_{sc} будет его частота, привязанная к частоте мазера. Этот кварцевый генератор через усилитель развязки обеспечивает сигнал, доступный потребителю. Он также питает синтезатор частоты, специально сконструированный для данной работы. Схематически он обеспечивает сигнал местного генератора на частоте $K_1\nu_{sc}$, близкой к частоте генерации мазера ν_r. Константа K_1 является рациональной дробью. Сигнал местного генератора смешивается с усиленным выходным сигналом мазера. Получается сигнал на промежуточной частоте $\nu_{int} = |\nu_r - K_1\nu_{sc}|$. Этот сигнал усиливается и сравнивается по фазе с другим сигналом, также генерируемым синтезатором на частоте $\nu'_{int} = K_2\nu_{sc}$, где K_2 – другая рациональная дробь. Если фазы сравниваемых сигналов отличаются, то на выходе фазового компаратора появляется сигнал ошибки,. Этот сигнал поступает в цепь управления частотой подстраиваемого генератора через низкочастотный операционный фильтр, который определяет передаточную функцию цепи управления. В стационарном режиме сигнал ошибки на выходе фазового компаратора равен нулю. В этом случае фазы, а следовательно и частоты двух сигналов равны. Тогда $\nu_{int} = \nu'_{int}$ и, следовательно, $\nu_{sc} = \nu_r/K_s$, где $K_s = K_1 + K_2$. Частота управляемого кварцевого генератора связана таким образом с частотой атомного генератора.

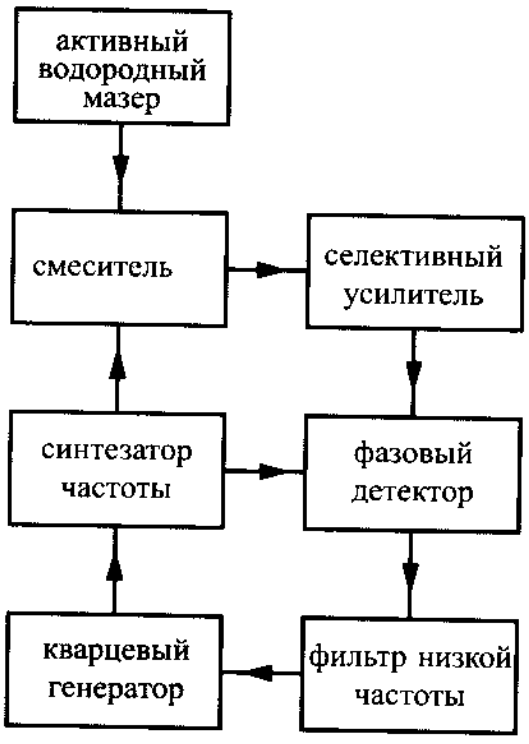


Рис. 6.26. Блок-схема электроники для управления частотой кварцевого генератора по частоте генерации водородного мазера

На рис.6.27 представлена эквивалентная функциональная схема фазового управления частотой. Обозначения такие же, как в разделе 6.4.1с. На данной схеме фазы двух сигналов, подходящих к фазовому компаратору, обозначены как $\Delta\Omega_r(p)/p$ и $K_s\Delta\Omega_{sc}(p)/p$. Поэтому сигнал ошибки равен

$$\Delta E(p) = K_d \frac{\Delta\Omega_r(p) - K_s\Delta\Omega_{sc}(p)}{p}. \qquad (6.111)$$

Здесь справедливы уравнения (6.47) и (6.48), описывающие изменения выходного напряжения операционного фильтра и угловой частоты генерации соответственно. Простейший операционный фильтр, который можно использовать, имеет постоянный коэффициент передачи A на низкой частоте. Следовательно

$$G(p) = A, \qquad (6.112)$$

и отсюда следует

$$\Delta\Omega_{sc}(p) = \frac{\mathcal{T}p}{1+\mathcal{T}p}\Delta\Omega_f(p) + \frac{1}{1+\mathcal{T}p}\frac{\Delta\Omega_r(p)}{K_S}, \qquad (6.113)$$

где

$$\mathcal{T} = \frac{1}{AK_dK_SK_q}. \qquad (6.114)$$

На практике постоянная времени $\mathcal{T}$ порядка 0,1 с.

Уравнение (6.113) имеет такой же вид, как (6.50), хотя операционный фильтр и не содержит интегратор. Действительно, когда управляют фазой, интегрирование обеспечивается тем, что сигнал ошибки пропорционален фазовому сдвигу сигнала, в то время как управляющий сигнал действует не на фазу, а на частоту кварцевого генератора. Отметим, что на

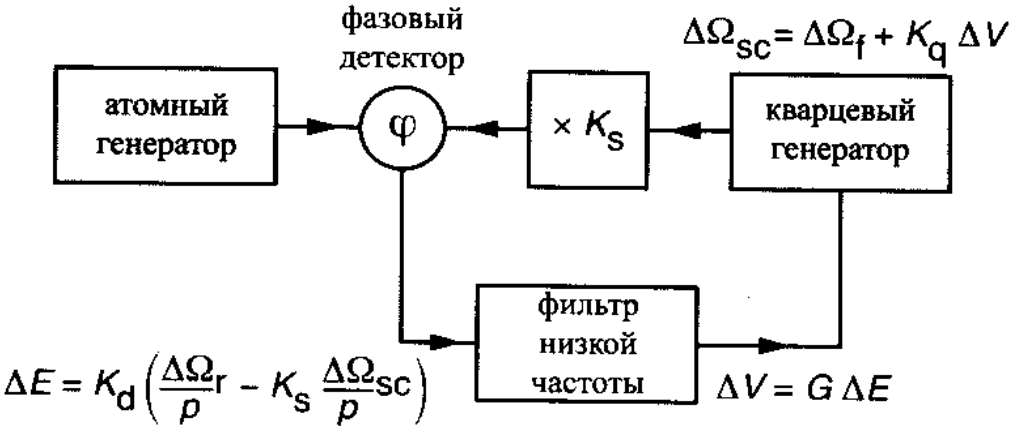


Рис. 6.27. Автоматическое управление частотой кварцевого генератора с помощью атомного генератора: функциональная схема

практике передаточная функция операционного фильтра не такая простая, как мы здесь предположили. Однако это не влияет на общие особенности рассматриваемой схемы автоматической подстройки.

Здесь применимо также заключение, приведенное в конце раздела 6.4.1c. Генератор с автоматической подстройкой воспроизводит быстрые флуктуации свободного кварцевого генератора и медленные изменения частоты атомной опоры.

(e) Автоматическая подстройка резонатора

Как упоминалось в разделе 6.5.3b, на настройку частоты резонатора налагаются строгие требования, если частота генерации должна оставаться постоянной. В этом состоит причина, почему большинство современных водородных генераторов оборудовано системами автоматической настройки резонатора. Здесь мы дадим краткое описание трех используемых способов достижения этой цели. В каждом случае получается сигнал ошибки. Он пропорционален степени расстройки резонатора по отношению к должной частоте генерации, равной частоте атомного перехода. Сигнал ошибки воздействует на варикап, связанный с резонатором.

Первый метод основан на измерениях генерируемой частоты мазера с автоматической настройкой резонатора. Для этого метода необходим второй опорный мазер. Обычно два мазера меняются местами через регулярные интервалы времени. В двух других методах информация о расстройке резонатора выводится непосредственно из поведения самого мазера.

Первый метод: модуляция атомного потока. Этот метод известен также как спин-обменная настройка. В его основе лежит уширение атомного резонанса из-за столкновений между атомами водорода в накопительной колбе с обменом электронами и, следовательно, спинового обмена. Это наиболее прямой метод, использующий выражение (6.109), представляющее изменение частоты генерации как функцию настройки частоты резонатора $\nu_c = \omega_c/2\pi$ [6.71, 6.72].

Атомный поток медленно модулируется с периодом в несколько сотен секунд между двумя значениями, связанными с добротностью Q_{at1} и с добротностью Q_{at2}. На рис.6.25 показано, что изменение частоты генерации пропорционально $\Delta\nu_c$, что оно меняет знак вместе с $\Delta\nu_c$ и что оно исчезает, если удовлетворяется условие $\nu_c = \nu_0$. Сигнал ошибки получается как выходной сигнал реверсивного счетчика. Этот прибор считает частоту биений между управляемым мазером и опорным мазером так, что он суммирует импульсы для одного полупериода модуляции и вычитает их для другого полупериода.

Обычно изменение частоты $\Delta\nu_{02} - \Delta\nu_{01}$ меньше, чем сдвиг частоты мазера из-за настройки частоты резонатора. Сдвиг равен $\Delta\nu_{01}$ или $\Delta\nu_{02}$ в

зависимости соответствующей добротности резонанса, Q_{at1} или Q_{at2} (рис.6.25). Более того, на измерения $\Delta\nu_{02} - \Delta\nu_{01}$ действуют случайные флуктуации частот обоих мазеров, и управляемого и опорного. Следовательно, сами водородные мазеры должны обеспечивать удовлетворительное отношение сигнал/шум для измерения этой величины.

На практике постоянная времени для коррекции настройки частоты резонатора управляемого мазера составляет около одного часа. Для периодов усреднения τ меньше чем один час стабильность частоты мазера ухудшается. Достоинство этой системы настройки проявляется только на больших интервалах времени, для τ больше чем один день.

Второй метод: модуляция частоты резонатора. Этот метод использует инерционность амплитуды магнитного момента атомов, обсуждавшихся в разделе 6.5.3b [6.69]. Это следствие высокого значения релаксационных времен T_1 и T_2, близких к одной секунде.

Частота резонатора модулируется прямоугольной волной с периодом в несколько сотых долей секунд, много меньшим чем T_1 и T_2. Электромагнитное поле в резонаторе представляет собой отклик на возбуждение осциллирующим магнитным моментом. Схематически это выглядит так, как будто это возбуждение имеет постоянную амплитуду и частоту. Отклик зависит от настройки частоты резонатора. В таком случае уровень осцилляций также модулируется, если только частота настройки резонатора не совпадает с частотой атомного перехода. Если это условие не выполняется, то появляется сигнал ошибки. Он может быть использован для автоматической подстройки резонатора таким образом, чтобы частота генерации совпадала с частотой атомного перехода.

Этот метод гарантирует обнаружение расстройки частоты с удовлетворительным отношением сигнал/шум. Постоянная времени управления резонатором составляет около одного часа. Стабильность частоты за средние времена не изменяется, а медленные уходы в настройке частоты резонатора и, следовательно, частоты мазера, таким образом подавляются.

Третий метод: непосредственный контроль частоты резонатора. Предположим на момент, что резонатор не содержит атомов водорода. Мы могли бы тогда использовать один из известных методов автоматической настройки резонатора на заданную частоту. Один из возможных методов описан в разделе 6.4.1c и показан схематически на рис.6.9, но с заменой кривой атомного резонанса на кривую отклика резонатора. Практически в резонатор вводится частотно-модулированный сигнал с девиацией $\Delta\nu'_0$ около средней частоты.

В применении к водородному мазеру должна соблюдаться осторожность в том, чтобы вводимый в резонатор сигнал не возмущал частоту генерации мазера. С этой целью очень существенно было бы подавить

несущую частоту вводимого сигнала, когда она равна частоте атомного перехода. Это достигается очень просто, если форма модулирующего сигнала опроса резонатора является прямоугольной. В этом случае несущая частота исключается, если частота модуляции является четной кратной частью от глубины модуляции $\Delta\nu'_0$ [6.73]. Это условие может быть легко выполнено путем использования техники синтеза частот. Фактически сигнал частоты опроса синтезируется из сигнала кварцевого генератора, частота которого автоматически подстраивается по частоте мазера.

Сигнал опроса резонатора может иметь бульшую интенсивность, чем интенсивность генерация мазера. Отсюда следует, что сигнал ошибки при расстройке частоты резонатора имеет хорошее отношение сигнал/шум. На практике постоянная времени управления резонатором может быть малой, около 10 секунд. При этом улучшается долговременная стабильность водородного генератора без заметного изменения стабильности за средние времена [6.74].

6.5.4. Пассивный водородный мазер

Когда условия автоколебаний не выполняются, т.е. когда $I < I_t$, то мазер ведет себя как усилитель. На резонатор подается сигнал на частоте 1,42... ГГц от внешнего генератора. Откликом стандарта является просто усиленный сигнал. Мазер поэтому становится пассивным стандартом частоты с очень селективным по частоте откликом.

(а) Усиление мазера

Для того чтобы смоделировать внешнее возбуждение резонатора на угловой частоте ω, в правую сторону уравнения (6.107) должен быть добавлен член, отражающий сигнал источника частоты [6.4]. Комплексная амплитуда магнитного поля в резонаторе является тогда решением уравнения

$$\left[\omega_c^2 - i\omega\frac{\omega_c}{Q_c} - \omega^2\right]B(\omega) = \frac{\mu_0\eta}{V_c}\omega^2 M(\omega) - \frac{\omega_c^2}{Q_c}B_e(\omega), \qquad (6.115)$$

где $M(\omega)$ все еще дается формулой (6.104). Дополнительный член таков, что когда нет атомов и когда мы имеем $\omega = \omega_C$, то магнитное поле с комплексной амплитудой $B'(\omega)$ имеет такой же модуль, как $B_e(\omega)$. Положим

$$iB'(\omega) = B_e(\omega). \qquad (6.116)$$

Комплексный коэффициент усиления усилителя является отношением величины $B(\omega)$, когда атомы есть, к ее величине $B'(\omega)$, когда там атомов нет. Для упрощения мы предположим, что резонатор настроен на частоту атомного резонанса и ограничимся в обсуждении случаем, когда коэффициент насыщения много меньше единицы. Тогда мы получим

$$|G(\omega)| = \left[1 + \frac{\alpha(\alpha-2)}{1+T_2^2(\omega-\omega_0)^2}\right]^{-1/2}, \tag{6.117}$$

для модуля ненасыщенного коэффициента усиления, где мы определили

$$\alpha = \frac{I}{I_t}. \tag{6.118}$$

На рис. 6.28 показано, как изменяется модуль ненасыщенного коэффициента усиления от величины $\omega - \omega_0$ для нескольких значений α. На резонансе коэффициент усиления меняется монотонно от 1, когда $\alpha = 0$, до бесконечности, когда $\alpha = 1$. В первом случае отклик мазера является просто откликом резонатора, во втором случае достигается порог генерации. Резонансная кривая сужается при приближении к порогу генерации. Ее ширина на полувысоте равна $2/T_2$ для очень малых значений α. Она стремится к 0, если α стремится к 1.

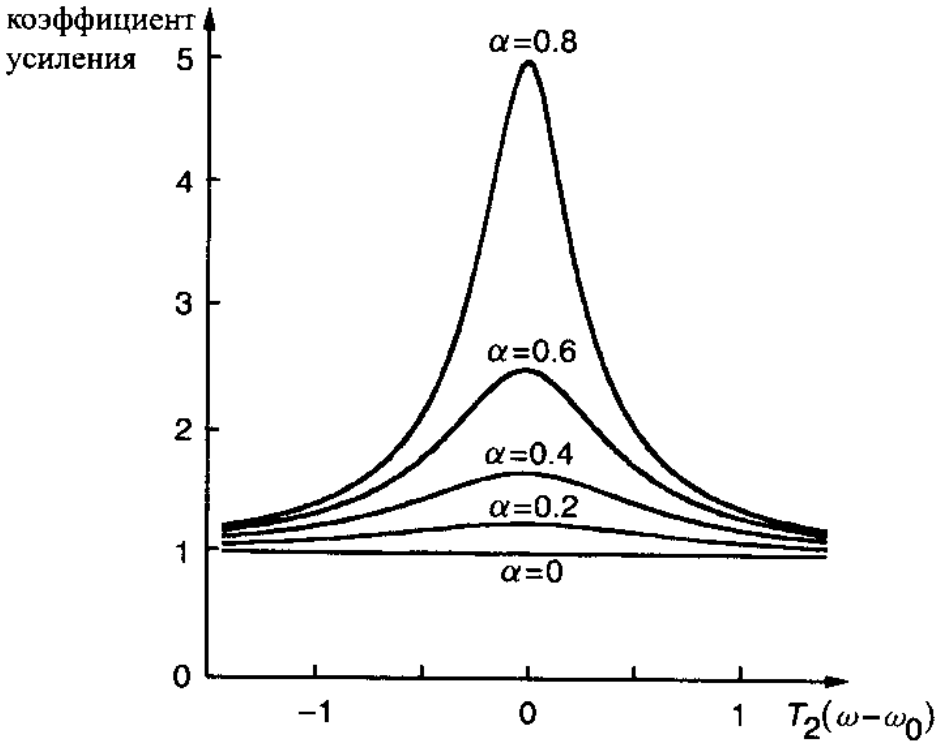


Рис. 6.28. Ненасыщенный коэффициент усиления пассивного водородного мазера

(b) Резонансная частота

В пассивном водородном мазере связь между электромагнитным полем и атомной средой также очень сильная, т.к. среда усиливает электромагнитное поле в резонаторе. На практике зависимость частоты атомного ре зонанса от частоты настройки резонатора также дается выражением

(6.109) [6.75]. Здесь ω должна рассматриваться как угловая частота атомного перехода, в то время как ω_C является угловой частотой резонатора.

(c) Практическая конструкция атомного резонатора

Активный и пассивный мазеры существенно отличаются объемами своих резонаторов. Для того чтобы уменьшить объем в 10 раз в резонатор должны вноситься материалы, потери в которых делают невозможным достижение добротности больше, чем 15×10^3.

Условия генерации больше не выполняются, и поэтому мазер будет работать пассивно. (Однако, возможно увеличение добротности электронным путем, т.е. путем использования цепей, усиливающих добротность так, что могут выполняться условия генерации. Этот метод применялся ранее, но сейчас он не используется).

Естественно, что предпринимаются все усилия для уменьшения объема других частей мазера без значительного изменения их структуры по отношению к активному мазеру. Поэтому мы рассмотрим только различные решения, предпринятые для наибольшего уменьшения объёма резонатора. Они схематически показаны на рис. 6.29.

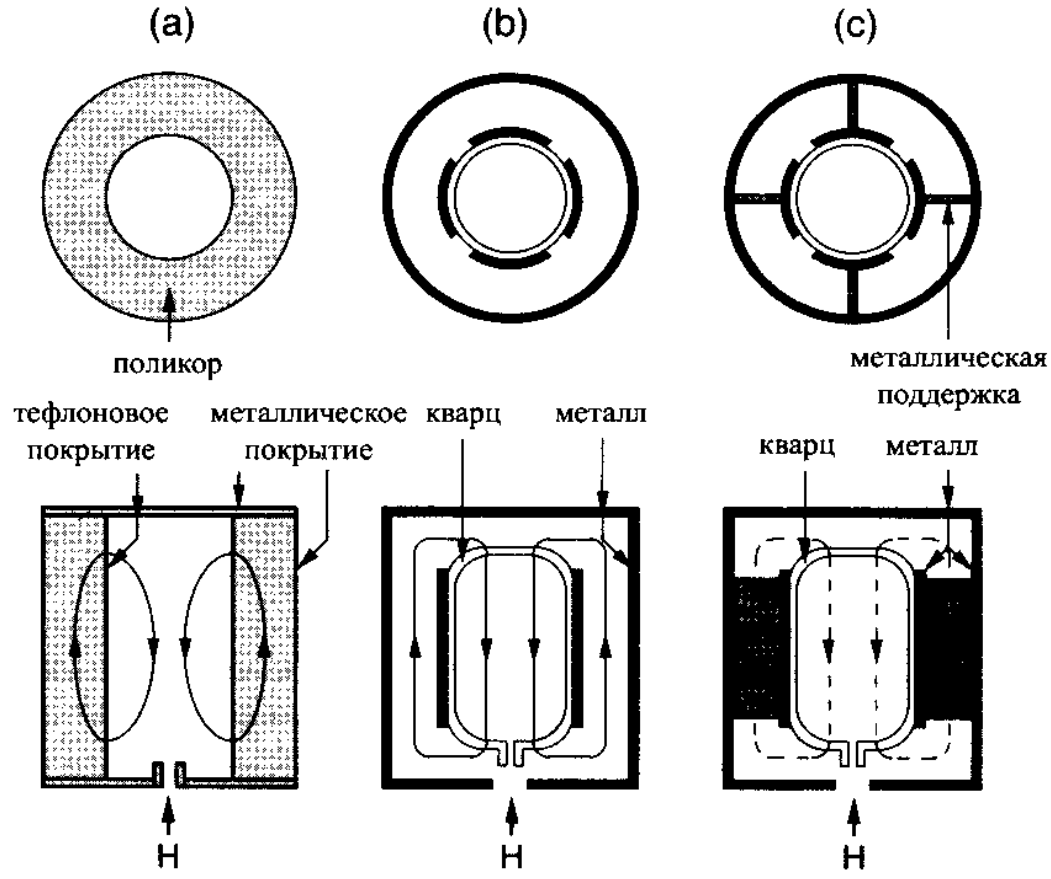


Рис. 6.29. Резонаторы для пассивного водородного мазера. В *(c)* силовые линии поля лежат вне плоскости поперечного сечения

Резонатор может быть частично заполнен кольцами из поликора или сапфира (рис. 6.29a) [6.76, 6.77]. В этом случае диэлектрический материал служит как поддержка для внешнего металлического покрытия, а также для покрытия фторопластом изнутри. Это определяет границы резонатора и накопительной ячейки соответственно.

Другая альтернатива – введение металлических электродов в резонатор. Они окружают накопительную колбу и имеют форму цилиндрических секторов. В одном варианте они жестко скрепляются с накопительной колбой, не касаясь внешней оболочки резонатора (рис. 6.29b) [6.78]. В другом варианте они соединяются с внешними стенками резонатора через радиальные металлические поддержки (рис. 6.29c) [6.79]. Структура резонатора напоминает тогда структуру магнетрона. Их легче воспроизвести и они более жесткие, чем те, которые показаны на рис. 6.29b.

(d) Автоматическая подстройка частоты кварцевого генератора и автоматическая подстройка частоты резонатора

Кварцевый генератор может автоматически подстраиваться таким же методом, который применяеется в цезиевых часах. В этом случае сигнал опроса на частоте 1,42 ГГц должен быть модулирован по частоте с очень медленной модуляцией на частоте около 0,1 Гц, т.е. меньше, чем ширина атомного резонанса.

Этот метод не лучший, потому что на модулированный сигнал опроса будет действовать фликкерный частотный шум. Действительно, чем ближе спектральные компоненты модулированного сигнала лежат к несущей частоте, тем больше они подвергаются действию этого типа шума. К тому же постоянная времени автоматической подстройки кварцевого генератора должна быть велика по сравнению с периодом модуляции, и поэтому она должна будет иметь слишком большую величину.

В пассивном водородном мазере резонанс наблюдается путем изменения амплитуды электромагнитного поля в резонаторе. Частота опроса может быть стабилизирована путем применения варианта метода Паунда [6.80]. Этот метод обычно применяется на сверхвысоких частотах в экспериментах по магнитному резонансу. В противоположность описанному выше методу частота модуляции должна быть выше, чем ширина опорного резонанса. Таким образом, удаляются препятствия, описанные выше. При использовании данного варианта и частота кварцевого генератора, и частота резонатора автоматически подстраиваются по частоте атомного резонанса. Резонатор должна быть термостабилизирован, так как его внутренняя структура означает, что он имеет высокий коэффициент теплового расширения.

На рис. 6.30 в сильно упрощенном виде показана блок-схема электроники. Сигнал опроса, генерируемый кварцевым генератором, модулиру-



Рис. 6.30. Пассивный водородный мазер. Схематическая диаграмма, показывающая, как могут автоматически подстраиваться кварцевый генератор и резонатор мазера

ется с частотой модуляции $\nu_м$, близкой к 15 кГц. Это много больше, чем ширина атомного резонанса, и это примерно тот же порядок величины, что и половина полосы пропускания резонатора. Амплитудный и фазовый сдвиг проходящей несущей волны зависит от разности между ее частотой и частотой атомного резонанса. Это выглядит как функциональная зависимость амплитуды и фазы от частоты резонатора. На выходе резонатора сигнал усиливается. Его различные спектральные компоненты интерферируют в квадратичном детекторе. Детектированный сигнал модулирован и из него извлекаются компоненты частоты модуляции $\nu_м$. Две цепи, вносящая каждая свой соответствующий фазовый сдвиг, разделяют затем эти компоненты, выражающие отклонение несущей частоты и частоты атомного резонанса с одной стороны и несущей частоты и частоты резонатора с другой стороны [6.8]. Эти два сигнала затем демодулируются синхронно с модуляцией сигнала опроса. Таким образом, генерируются два сигнала ошибок. После интегрирования один сигнал идет на управление частотой кварцевого генератора, а другой на управление частотой резонатора.

6.5.5. Метрологические свойства водородного мазера

(а) Нестабильность за средние времена

Для времени выборки меньше, чем примерно 0,1 с, нестабильность частоты и активного и пассивного водородного мазера примерно такая же как и у кварцевого генератора, управляемого по атомному резонансу. Если $\tau > 0{,}1$ с, то нам необходимо различать случаи активного и пассивного мазеров.

Активный мазер. В активном водородном мазере электромагнитное поле в резонаторе индуцирует атомный магнитный момент, который затем поддерживает электромагнитное поле. Однако в резонансной моде присутствует тепловой шум, и связанное с ним случайное электромагнитное поле возмущает генерацию когерентных осцилляций. Шумовые компоненты, находящиеся в фазе с осцилляциями, создают частотные флуктуации. Можно показать, что спектральная плотность мощности в относительных флуктуациях частоты (раздел 5.2.2) дается из [6.4, 6.82].

$$S_y(f) = \frac{kT}{P_{\text{at}} Q_{\text{at}}^2}, \tag{6.119}$$

где k постоянная Больцмана, T – термодинамическая температура резонатора, P_{at} – мощность, рассеиваемая в резонаторе, и Q_{at} – добротность атомного резонанса. Это белый частотный шум. Таблица 5.2 показывает, что вариация Аллана для флуктуаций частоты тогда равна

$$\sigma_y(\tau) = \frac{1}{Q_{\text{at}}} \left(\frac{kT}{2P_{\text{at}}} \right)^{1/2} \tau^{-1/2}. \tag{6.120}$$

Это выражение описывает нестабильность частоты за средние времена, для τ больше чем 10 с. На малых временах, для τ между 0,1 и 10 с, нестабильность частоты обусловлена низким уровнем мощности, генерируемой мазером. В этом случае мы должны рассматривать эффекты случайных сигналов, добавляющихся к сигналу генерации. Они происходят от тепловых шумов самого резонатора и от шумов цепей, усиливающих сигнал мазера. Компоненты этих флуктуаций, находящиеся в квадратуре с осцилляциями, приводят к флуктуациям в наблюдаемом сигнале. Спектральная плотность мощности относительных флуктуаций частоты в этом случае равна

$$S_y(f) = \frac{4kT}{P_{at}} \frac{f^2}{\nu_0^2} \left(1 + F \frac{P_{at}}{P_r} \right), \tag{6.121}$$

где ν_0 – частота генерации. Параметр F – коэффициент шума усилителя и P_r – эффективная мощность, принимаемая усилителем. Стандартная вариация Аллана, связанная с этим белым шумом, равна

$$\sigma_y(\tau) = \frac{3kTf_h}{2\pi\nu_0^2 P_{at}} \left(1 + F \frac{P_{at}}{P_r} \right) \tau^{-1}, \tag{6.122}$$

где f_h – частоты среза прибора, используемого для наблюдения флуктуаций частоты. Стандартные отклонения суммируются квадратично.

Обычно, если частота среза приборов, измеряющих нестабильность частоты, равна несколько герц, то как показано на рис. 6.11, мы имеем $\sigma_y(\tau) \sim 10^{-13}\,\tau^{-1}$ для $\tau < 10$ с и $\sigma_y(\tau) \sim 3 \times 10^{-14}\,\tau^{-1/2}$ для $\tau > 10$ с.

Пассивный мазер. Хотя метод, используемый для подстройки кварцевого генератора по атомному резонансу, не такой же как для цезиевых часов, уравнение (6.63) остается справедливым до следующего по порядку величины члена. Для пассивного мазера

$$\sigma_y(\tau) = \frac{K}{Q_{at}} \left(\frac{kT}{P_{at}} \right)^{1/2} \tau^{-1/2}, \tag{6.123}$$

где K – константа порядка 10, зависящая от экспериментальных условий. На практике мы имеем $\sigma_y(\tau) \sim (1-2) \times 10^{-12}\,\tau^{-1/2}$ для $1 < \tau < 10^5$ с [6.83].

(b) Остаточные сдвиги частоты

Первые три сдвига частоты, которые мы рассмотрим, являются общими для всех типов атомных стандартов частоты. Последние три сдвига относятся только к водородным мазерам.

Эффект Зеемана второго порядка. Статическое магнитное поле B_0, направленное параллельно СВЧ магнитному полю, сдвигает энергетические уровни (рис. 6.2). Этот эффект Зеемана является эффектом второго порядка для уровней с $m_F = 0$. Сдвиг частоты равен

$$\Delta\nu_Z = 2{,}7730 \times 10^{11} B_0^2, \tag{6.124}$$

где B_0 выражается в Тесла. B_0 обычно близко к 10^{-7} Т. Это означает, что $\Delta\nu_Z/\nu_0 = 1{,}95 \times 10^{-12}$. Значение этого частотного сдвига измеряется путем индуцирования низкочастотных переходов $\Delta F = 0$, $\Delta m_F = \pm 1$ между подуровнями триплета $F = 1$. На практике относительная неопределенность в значении этого частотного сдвига меньше чем 10^{-13}.

Эффект Доплера второго порядка. Этот сдвиг релятивистского происхождения связан с движением атомов относительно опорной лабораторной системы. Атомы водорода приходят в тепловое равновесие через столкновения со стенками накопительной колбы. Их кинетическая энергия равна 3 $kT/2$. Из (6.16) сдвиг частоты тогда дается выражением

$$\frac{\Delta\nu_D}{\nu_0} = -\frac{3kT}{2Mc^2}, \tag{6.125}$$

где M – масса атома водорода. Предполагая, что температура накопительной колбы равна 313 К, мы получаем $\Delta\nu_D/\nu_0 = -4{,}31 \times 10^{-11}$. Хотя этот сдвиг и является наибольшим сдвигом, действующим на водородный мазер, но если температура стенки ячейки измерена с точностью около 0,1 К, то это приводит к исключительно малой неопределенности значений частоты сдвига, равной $1{,}4 \times 10^{-14}$.

Затягивание резонатора. Оно описывается (6.109) и уже обсуждалось в некоторых деталях. Если резонатор тщательно настроен вручную или автоматически, то неопределенность, вносимая этим эффектом, меньше чем 10^{-14}.

Столкновения между атомами водорода. Столкновения между атомами приводят к сдвигу частоты сверхтонкого перехода. Сдвиг частоты может быть почти полностью компенсирован подходящей ручной или автоматической настройкой резонатора, так что остаточный эффект на частоту, генерируемую мазером, будет составлять несколько единиц тринадцатого знака.

Магнитная релаксация. Сдвиг частоты происходит при движении атома в СВЧ поле при наличии градиента статического поля. Сдвиг зависит от величины этого градиента, от величины статического поля и состава пучка. Если приняты меры предосторожности, чтобы создать по возможности более однородное поле, то сдвиг частоты меньше 10^{-13} при $B_0 = 10^{-7}$ Т.

Столкновения со стенками накопительной колбы. Столкновения со стенками изменяют фазу индуцированного дипольного момента атома. Соответствующий сдвиг частоты имеет вид

$$\Delta\nu_W = \frac{K}{D}, \tag{6.126}$$

где K – константа, зависящая от свойств покрытия и от температуры, а D – диаметр накопительной колбы. Например, $K \sim -0{,}4$ Гц на сантиметр при температуре 40° Цельсия для покрытия FEP 120 Тефлон, производимого фирмой Дюпон де Немурс. Следовательно, $\Delta\nu_W/\nu_0 \sim -1{,}9 \times 10^{-11}$ для $D = 15$ см. Этот сдвиг может быть измерен в принципе путем изменения

D. На практике трудности из-за воспроизведения покрытия приводят к неопределенности порядка 10% от значения сдвига частоты из-за столкновения со стенками.

(c) Долговременная нестабильность и неточность

Долговременная нестабильность частоты. На больших интервалах времени нестабильность частоты является результатом систематических эффектов, связанных с размерной нестабильностью резонатора и вариациями стеночных эффектов [6.84]. Первая упомянутая причина нестабильности ведет к дрейфу частоты порядка 10^{-16} до 10^{-15} в день. Вторая не имеет такой важности для некоторых приложений, таких как, например радиоинтерферометры со сверхдлинной базой. Для хранения времени должна использоваться система автоматической настройки резонатора. Другая причина нестабильности связана с медленно изменяющимися свойствами фторосодержащих полимеров, используемых для покрытия накопительных колб. Соответствующие изменения в частоте тоже 10^{-16} до 10^{-15} в день. Эти изменения частоты могут происходить в противоположных направлениях.

Долговременная нестабильность водородного мазера зависит от того, является ли он активным или пассивным, а также от производителя. Наиболее стабильными в долговременной области являются активные типы с системами автоматической настройки резонатора. Водородные мазеры имеют такую же долговременную стабильность, как первичные цезиевые стандарты частоты.

Неточность. Точность воспроизведения мазеров характеризует их способность генерировать без калибровки частоту, согласующуюся с частотой сверхтонкого перехода ν_H в невозмущенном основном состоянии атома водорода.

На практике неточность определяется неопределенностью значения сдвига частоты из-за столкновений атомов водорода со стенками накопительной колбы и близка к $\sigma_1 = (2\text{–}5) \times 10^{-12}$.

Значение частоты ν_H, полученной из измерений, а не из определений как в случае цезия, ухудшается с нестабильностью. Его стандартное отклонение равно 3 мГц в абсолютной величине (табл. 6.1), или $\sigma_2 = 2 \times 10^{-12}$. Стандартное отклонение σ_H неопределенности частоты воспроизводимой водородным мазером дается формулой

$$\sigma_H = \left(\sigma_1^2 + \sigma_2^2\right)^{1/2}. \tag{6.127}$$

6.5.6. *Криогенный водородный мазер*

Во многих областях науки и техники было бы очень полезно применять еще более стабильные генераторы, чем существующие сейчас лучшие во-

дородные генераторы. Например, генератор частоты с улучшенной стабильностью за средние и большие времена был бы очень полезен при изучении метрологических свойств часов на фонтане холодных атомов (раздел 6.4.3 и 6.4.4) и часов на удержанных ионах ртути (раздел 6.7.3). Такие генераторы могли бы также удовлетворить нужды навигации в дальнем космосе, и они могли бы использоваться для более точного измерения гравитационного сдвига частоты с целью получения отличий в различных выводах общей теории относительности.

Так как (6.120) и (6.122) показывают, что кратковременная нестабильность и нестабильность за средние времена усреднения для водородного мазера являются возрастающими функциями температуры резонатора, то ясно, что можно получить преимущества при охлаждении резонатора. Что касается долговременной нестабильности, то известно, что размерная стабильность материалов улучшается при низких температурах. Аналогично, мы можем надеяться на улучшение стабильности свойств материалов, используемых для внутреннего покрытия накопительной колбы, если есть способ их непрерывного обновления. Эти причины стимулировали ряд проектов по изучению свойств водородного мазера при низких температурах вплоть до 4,2 К.

Однако наиболее решающие результаты по отношению к возможным высокостабильным водородным мазерам были получены из фундаментального физического изучения атома водорода. Было обнаружено, что рекомбинация и релаксация этого атома сильно уменьшаются гелиевыми пленками при температурах ниже 1 К. Кроме того, сдвиг частоты за счет столкновений с пленкой составляет величину того же порядка, что и в мазерах, работающих при комнатной температуре, а уширение сверхтонкого резонанса из-за столкновений между атомами водорода на три раза порядка меньше по величине, чем при тех же плотностях при комнатной температуре. По этой причине атомная добротность Q_{at} и мощность P_{at} рассеиваемая в резонаторе могут быть значительно увеличены. Уравнения (6.120) и (6.122) показывают, что такое увеличение дало бы вклад в уменьшение нестабильности частоты мазера точно также, как и уменьшение температуры. Эти выводы использовались для оценки нестабильности криогенного водородного мазера, работающего при температуре 0,5 К [6.85]. Согласно этим оценкам она должна составлять около 10^{-18} для $\tau = 10^3$ с. Однако этот вывод был сделан в предположении, что в этом случае ранее представленные теоретические результаты будут все еще применимы. В реальности все не так. При температуре 0,5 К давление гелия таково, что средняя длина свободного пробега атома водорода составляла бы только около 1 см. Атомы водорода не могли бы больше усреднять пространственные вариации статического и СВЧ магнитных полей. Более того, появляются другие сдвиги частоты. Они происходят из-за

увеличения времени столкновения между атомами, которое само по себе происходит из-за того, что средние скорости атомов становятся малыми при малых температурах. Поэтому требуется дальнейшее изучение вопроса.

Для того чтобы построить криогенный водородный мазер необходимо использовать гелиевый ожижитель. Это одна из причин сложности и дороговизны такого прибора. Один из таких мазеров использовался в фундаментальных физических измерениях [6.86]. Во время написания этой книги еще не было возможности создания двух криогенных водородных мазеров с характеристиками, необходимыми для их метрологической оценки через взаимные сличения.

6.6. Часы на рубидиевой ячейке

Оптическое детектирование перехода в сверхтонкой структуре щелочных атомов, содержащихся в отпаянной резонансной ячейке, было впервые изучено в США в конце 1950 годов [6.87, 6.88, 6.89]. В то же самое время был разработан первый стандарт частоты на парах щелочного металла [6.90, 6.91]. Сразу же было осознано, что можно создавать компактные атомные стандарты частоты, и первые промышленные образцы таких стандартов появились в начале 1960 годов [6.92]. В этих промышленных моделях по практическим причинам из щелочных металлов предпочтение отдавалось атомам рубидия.

Рынок атомных часов с рубидиевой ячейкой оказался значительным, и многие фирмы производили и все еще производят эти типы атомных стандартов. Ежегодное производство достигает несколько тысяч приборов. Выбор был сделан в пользу высококачественных приборов, которые являются и надежными и компактными. Их объем может быть уменьшен до 0,25 литра (таблица 6.2) [6.93, 6.94].

По своей структуре часы на рубидиевой ячейке являются вторичными атомными стандартами частоты (т.е. их частота должна калиброваться по первичному источнику). Их долговременная нестабильность лежит между нестабильностью кварцевых генераторов и цезиевых пучковых часов. Они используются главным образом в системах определения положения, навигации и в телекоммуникационных системах.

6.6.1. Описание

Рубидиевые часы являются стандартом частоты пассивного типа. В них используется *метод двойного резонанса*, т.е. оптический и СВЧ резонансы и это показано схематически на рис. 6.31. Небольшая *лампа* объемом меньше 1 см^3 содержит около 1 мг рубидия 87 или смеси изотопов 85 и 87 вместе с инертными газами, такими как криптон. Радиочастотный генератор поддерживает разряд в лампе. Рабочая температура выбирается от 125 до 140 °C.

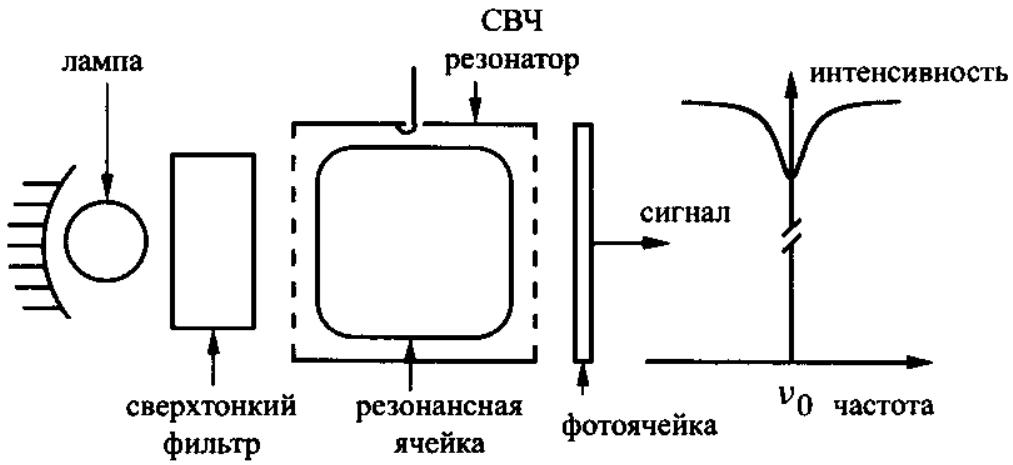


Рис. 6.31. Схематическая диаграмма часов на рубидиевой ячейке

В модификации, показанной на рис.6.31, свет проходит через *сверхтонкий фильтр*. Фильтр содержит рубидий 85 и инертный газ, например аргон, при давлении около 10^4 Па.

Резонансная ячейка, имеющая объем несколько $см^3$, содержит рубидий 87. В ней также присутствует *буферный газ* при низком давлении (около10^3 Па). Буферный газ представляет собой смесь азота и инертных газов.

В представленной здесь конструкции сверхтонкий фильтр выполнен отдельно. В более компактных моделях сверхтонкий фильтр объединен с самой резонансной ячейкой. Она содержит естественную смесь изотопов рубидия, состоящую на 72% из рубидия 85 и 28% рубидия 87, такой сверхтонкий фильтр называют интегрированным. Количество рубидия, содержащегося в резонансной ячейке, очень мало, порядка одного миллиграмма.

Эта ячейка помещается в *резонансной полости* (далее резонатор), настроенной на частоту сверхтонкого перехода в рубидии 87, т.е. около 6834,68 МГц. Сюда приходит сигнал возбуждения. Используются различные методы для достижения подходящей конфигурации электромагнитного поля в ячейке и в то же время стремятся получить минимально достижимый объем резонатора. Например, может быть использована конструкция резонатора, показанная на рис. 6.29с. Две параллельные стенки резонатора имеют отверстия для прохода света, генерируемого лампой. Температура резонансной ячейки поддерживается около 80 °С.

Интенсивность пройденного через ячейку света измеряется с помощью кремниевого *фотоэлектрического детектора*. Когда частота сигнала возбуждения изменяется вблизи частоты сверхтонкого перехода в рубидии 87, то вариации в токе фотоэлектрической ячейки имеют вид перевернутой резонансной кривой.

Магнитные экраны защищают ячейку от окружающего магнитного поля и его флуктуаций. Расположенные должным образом катушки внутри этих экранов создают магнитное поле величиной около 10^{-5} Т.

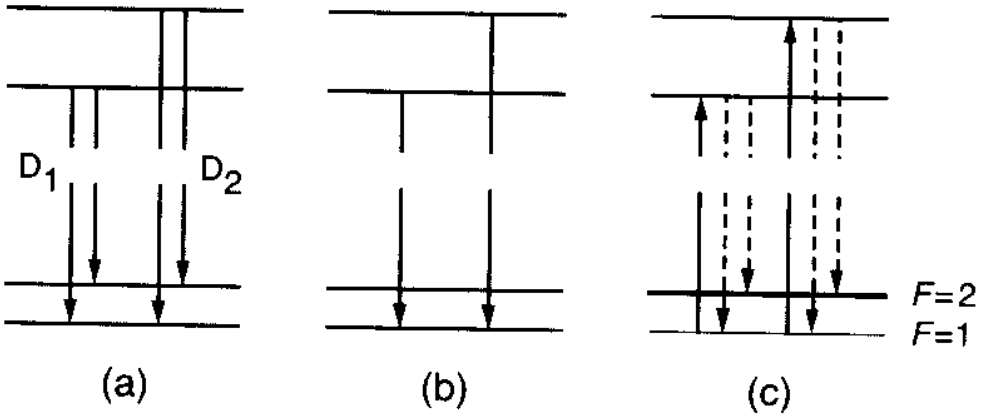


Рис.6.32. Рубидиевые часы. (a) Спектр, излучаемый рубидиевой лампой. (b) Отфильтрованный спектр. (c) Оптическая накачка в ячейке.

6.6.2. Принцип действия

Внутри резонансной ячейки происходят сложные физические явления. Они проявляются в результате действия двойного резонанса в присутствия буферного газа. Здесь мы представим только очень упрощенное объяснение атомного резонатора, проиллюстрированное на рис. 6.32.

В лампе атомы рубидия 87 излучают резонансные линии D_1 и D_2 на длинах волн 795 и 780 нм соответственно. Эти линии имеют сверхтонкую структуру, обусловленную существованием двух сверхтонких уровней $F = 1$ и $F = 2$ в основном состоянии. На рис. 6.32a показан очень упрощенный спектр излучаемого лампой света.

Благодаря благоприятному изотопическому совпадению атомы рубидия 85 в сверхтонком фильтре поглощают по одной из сверхтонких компонент в линиях D_1 и D_2 так, что фильтр является непрозрачным для этих компонент света, излучаемого лампой, в то время как остальные компоненты света лампы проходят через фильтр. Буферные газы, содержащиеся в фильтре, сдвигают и уширяют линии резонанса в рубидии 85. Их давление настраивается на оптимальную эффективность фильтрации. На рис. 6.32b показан спектр света, прошедшего через фильтр.

В резонансной ячейке атомы рубидия 87, находящиеся на сверхтонком подуровне $F = 1$ основного состояния (см. раздел 6.2.2b и c), селективно поглощают падающий свет. При этом атомы переходят в возбужденное состояние, откуда они возвращаются в одно из двух состояний P. Они возвращаются на одно из двух основных состояний либо через флюоресценцию, либо через столкновения с атомами азота, причем последний механизм является доминирующим.* Поглощательно-релаксационный цикл

* Релаксация через столкновения не является излучательной. Давление азота выбирается таким, чтобы интенсивность флюоресцентного света атома в ячейке была очень малой. Спектр этого света был бы похож на спектр, показанный на рисунке 6.32a. Такая флюоресценция уменьшила бы эффективность оптической накачки, т.к. она возбуждала бы атомы и с уровня F=2, и с уровня F=1.

уменьшает населенность уровня $F = 1$ и увеличивает населенность $F = 2$. Такой цикл оптической накачки показан на рис. 6.32. Ячейка становится прозрачной для падающего света, т.к. число атомов, способное поглощать этот свет, становится малым.

До сих пор мы не упоминали СВЧ поле в резонаторе. Вышеупомянутое описание справедливо, если это поле не вызывает перехода между двумя сверхтонкими подуровнями основного состояния. Давайте предположим, что частота сигнала возбуждения равна частоте сверхтонкого перехода. Тогда стимулируется переход $F = 2 \rightarrow F = 1$, и уровень $F = 1$ опять населяется. Часть падающего света поглощается и, следовательно, интенсивность света, попадающего на фотоячейку, немного уменьшается. Рис. 6.31 показывает резонансный профиль при сканировании СВЧ частоты относительно частоты атомного резонанса. Этот лоренцовский профиль может быть описан уравнением

$$I = I_n\left[1 - \frac{k}{1 + 4(\nu - \nu_0)^2 / \Delta\nu^2}\right], \qquad (6.128)$$

где I – поток собираемых фотонов, равный вне резонанса I_n. Ширина $\Delta\nu$ на полувысоте резонанса составляет около 500 Гц, параметр k близок к 0,5 %.

Буферный газ в резонансной ячейке выполняет несколько важных задач. В дополнение к безизлучательному затуханию рубидия 87 буферный газ подавляет доплеровское уширение сверхтонкого резонанса (раздел 6.2.3b). Кроме того он уменьшает скорость, с которой атомы сталкиваются со стенками ячейки, и уменьшает таким способом релаксацию и сдвиг частоты, связанный с этими столкновениями. Давление буферного газа устанавливается так, чтобы скорость диффузии атомов рубидия была очень мала, порядка 1 см $с^{-1}$. На самом деле атомы рубидия стационарны только во время их релаксации, составляющее около 1 мс. А поэтому они не могут усреднять значения оптического и СВЧ полей в ячейке. Вследствие этого свойства ячейки становятся неоднородными, и это сильно усложняет количественный анализ.

Если осуществляется сверхтонкая фильтрация в самой резонансной ячейке, как это делается в малогабаритных конструкциях, то нежелательные сверхтонкие компоненты падающего света последовательно уменьшаются вместе с прохождением света по траектории внутри ячейки, так что детальное объяснение поведения часов еще более усложняется.

6.6.3. Электроника

Так как рубидиевые часы являются пассивными атомными стандартами частоты, то атомный резонанс должен быть использован для того, чтобы автоматически подстраивать частоту кварцевого генератора таким же спо-

собом, как и в цезиевых пучковых часах (раздел 6.4.1c). Другие электрические цепи генерируют радиочастотный сигнал, необходимый для возбуждения разряда в лампе. Существенно наличие отдельных систем термостатирования лампы, резонансной ячейки и сверхтонкого фильтра (если он существует отдельно). Это дает возможность подстройки плотности рубидия и давления газа, содержащегося в различных частях прибора, что гарантирует удовлетворительную работу лампы и достаточное отношение сигнал/шум при детектировании сверхтонкого резонанса и при контроле нежелательных сдвигов частоты.

6.6.4. Метрологические свойства часов с рубидиевой ячейкой

(a) Нестабильность частоты за средние времена

Как и для любого пассивного стандарта приближение для нестабильности за средние времена получается из (6.63). Здесь шум является дробовым шумом электрических цепей фотоячейки. По порядку величины мы имеем $Q_{at} \sim 10^7$ и $(P_S/P_N)^{1/2} \sim 2\times10^3$ в полосе пропускания 1 Гц, откуда следует что $\sigma_y(\tau) \sim 10^{-11}\tau^{-1/2}$.

На практике в зависимости от изготовителя нестабильность за средние времена имеет значения между $5\times10^{-12}\tau^{-1/2}$ и $10^{-10}\tau^{-1/2}$. Для периодов выборки τ больше чем 10^3 или 10^4 секунд $\sigma_y(\tau)$ больше не изменяется как $\tau^{-1/2}$, потому что начинают оказывать влияние систематические эффекты. На рис. 6.11 показана типичная нестабильность частоты рубидиевых часов в сравнении с нестабильностью других часов.

(b) Сдвиги частоты атомного перехода

Кроме сдвигов частоты, присущих всем атомным часам, существуют два особенно губительных сдвига частоты, специфические для часов с рубидиевыми ячейками. Один из них типичен для экспериментов по двойному резонансу, когда два перехода имеют общий уровень и возбуждаются одновременно. Этот сдвиг называется *световым сдвигом*, потому что он связан с оптической накачкой. Другой сдвиг происходит из-за столкновений атомов рубидия с молекулами буферного газа, при которых силы взаимодействия приводят к изменениям в сверхтонком расщеплении.

Световой сдвиг. Эффект исчезает, если спектр света накачки расположен по центру линии поглощения рубидия. Это условие зависит от состава и содержания лампы накачки и сверхтонкого фильтра или от резонансной ячейки, если она объединена с фильтром. На практике световой сдвиг никогда не может быть полностью подавлен. Его величина составляет несколько герц.

Сдвиг из-за столкновений. Этот сдвиг частоты пропорционален давлению буферного газа. В типичных условиях его величина составляет около одного килогерца. Его значение изменяется от одних часов к другим, потому что при заполнении и отпайке рубидиевых ячеек невозможно контролировать с необходимой точностью состав и давление буферного газа.

Часы с рубидиевой ячейкой генерируют сигнал с заранее предопределенной номинальной частотой, например с частотой 10 МГц. Это осуществляется путем настройки частоты синтезатора, генерирующего СВЧ сигнал возбуждения. Эта настройка достигается путем присвоения разных значений коэффициенту K_s, введенному в разделе 6.4.1c для разных часов одного и того же изготовления. Эта операция вызывает необходимость калибровки по отношению к какому-нибудь опорному источнику. Понятие точности в данном случае не имеет смысла, если мы рассматриваем это понятие в отношении рубидиевых часов.

(c) Долговременная нестабильность и эффекты окружающей среды

На больших временах усреднения τ больше, чем $10^3 - 10^4$ с, частота часов с рубидиевой ячейкой дрейфует на $(1–5)\times10^{-11}$ в месяц. Причины дрейфа до сих пор полностью не выяснены. Возможно дрейф связан с химической реакцией между рубидием и стеклом и изменением состава буферного газа и его давления из-за диффузии через фильтр и стенки ячейки.

По сравнению с другими атомными стандартами времени и частоты рубидиевый стандарт наиболее чувствителен к окружающим условиям [6.95]. Наиболее существенны эффекты изменения температуры, достигающие 10^{-10} на градус Цельсия. Изменения температуры воздействуют на спектр света, излучаемого лампой и на эффективность сверхтонкой фильтрации, а следовательно, и на световой сдвиг. Они также изменяют вероятность столкновений между атомами рубидия и молекулами буферного газа в резонансной ячейке и оказывают таким образом воздействие на сдвиг частоты сверхтонкого перехода.

Технический выбор состава и давления используемого газа, а также выбор температуры лампы, сверхтонкого фильтра (если он разделен), и резонансной ячейки, в значительной степени определяются необходимостью уменьшения эффектов окружающей среды. Несмотря на существующие при этом трудности в настоящее время производятся рубидиевые часы, которые показывают очень низкую чувствительность к окружающим условиям. Они используются одни или совместно с цезиевыми часами на борту спутников глобальной системы определения положения и навигации (GPS) (см. раздел 9.2.2).

6.6.5. Современные исследования и перспективы

Теоретические и экспериментальные работы показывают, что разрядную лампу можно эффективно заменить полупроводниковым лазером, работающим на длине волны 780 нм [6.96, 6.97, 6.98]. Этот тип лазера обычно используется в дисковых проигрывателях и, следовательно, он оказывается дешевым в производстве. Квази-монохроматический свет лазера стабилизируется по линии рубидия D_2 и это делает возможным осуществление эффективной оптической накачки при существенно меньшем потоке света по сравнению с используемой разрядной лампой. Как следствие возрастает относительная высота резонанса при общем снижении тока на выходе фотоячейки. Благодаря этому можно получить стабильность частоты за средние времена на уровне $10^{-13}\tau^{-1/2}$. С технической точки зрения применение лазера также проще, чем использование лампы, и это дает возможность улучшить долговременную стабильность частоты.

В лабораторных исследованиях созданы мазеры на рубидии 87, работающие в активном режиме [6.99, 6.100]. С этой целью насколько это возможно оптимизируется осуществление инверсной населенности между уровнями $F = 2$, $m_F = 0$ и $F = 1$, $m_F = 0$, а резонансная ячейка располагается в резонаторе с максимально возможной добротностью. Основная притягательность рубидиевого мазера заключается в его отличной стабильности частоты, достигающей 10^{-13} за время усреднения τ, близкое к одной секунде. Однако эта стабильность уменьшается при временах $\tau > 1$ с из-за световых сдвигов и сдвигов из-за столкновений с буферным газом.

6.7. Часы на удержанных ионах

Уже несколько десятилетий было известно, что ионы могут быть локализованы (удержаны) в вакууме с помощью электромагнитного поля. Эта локализация означает, что ионы удерживаются в ограниченном пространстве на интервалах времени до одного дня и больше *без привлечения физических стенок.*

В ловушке Пеннинга используются статическое электрическое и магнитное поле, в то время как в ловушке Пауля используется радиочастотное электромагнитное поле. На практике наличие магнитного поля в ловушке Пеннинга не является достоинством для применений в стандартах частоты. Поэтому мы ограничимся только рассмотрением удержания в ловушке Пауля [6.101].

Многие ионы имеют сверхтонкую структуру. Действительно, для ионов $^{3}He^{+}$, $^{9}Be^{+}$, $^{25}Mg^{+}$, $^{88}Sr^{+}$, $^{113}Cd^{+}$, $^{135}Ba^{+}$, $^{137}Ba^{+}$, $^{171}Yb^{+}$, $^{173}Yb^{+}$ и $^{199}Hg^{+}$ сверхтонкое расщепление в основном состоянии известно с точностью лучше чем единица на 10^{-8}. В 2002 году для создания приборов, которые можно было бы рассматривать в качестве стандартов частоты и времени, исполь-

зовались только ионы $^{171}Yb^{+}$ и $^{199}Hg^{+}$. Так как наилучшие результаты были достигнуты на ионах ртути, то мы сконцентрируем наше внимание на ртутных ионных стандартах с удержанием в ловушке Пауля. Первоначально идея была представлена Мэйжором в конце 1960 годов в лабораториях NASA (National AeroSpace Agency – национальное аэрокосмическое агентство). Впервые сверхтонкий резонанс удержанных ионов ртути $^{199}Hg^{+}$ экспериментально наблюдался в NASA [6.102], а затем в университете Майнца (Германия) [6.103].

Основные этапы в разработке соответствующего стандарта частоты были выполнены в лаборатории атомных часов (Франция) [6.104], в Хьюлетт-Паккард (США) [6.105, 6.106] и в лаборатории реактивного движения (США) [6.107, 6.108, 6.109]. Эти приборы можно сделать компактными для использования в полевых условиях. Экспериментальная нестабильность частоты для этих приборов составляла $10^{-13}\tau^{-1/2}$. Совсем недавно в национальном институте стандартов и технологии (NIST) (США) была успешно выполнена работа по созданию стандарта частоты на ионах ртути с лазерным охлаждением. Была достигнута относительная нестабильность частоты $3{,}3\times10^{-13}\tau^{-1/2}$ и неточность 3×10^{-15} [6.110]. Это уже более сложный и громоздкий прибор.

Мы начнем с представления принципов удержания ионов в ловушке Пауля. Затем мы перейдем к описанию стандарта частоты на ионах ртути, использующего облако ионов, которое приводится в тепловое равновесие путем столкновения с буферным газом при комнатной температуре. Здесь используется оптическая накачка с помощью изотопической лампы. Мы кратко обсудим стандарт частоты на ионах иттербия и закончим представлением лазерного охлаждения ионов и особенностями стандарта частоты на ионах ртути с лазерным охлаждением.

6.7.1. Удержание ионов в радиочастотной ловушке

(а) Распределение потенциалов

Начнем с рассмотрения двух распределений потенциалов, которые используются для того, чтобы ограничить перемещение заряженных частиц. Это соответственно линейная ловушка (используемая также в некоторых газоанализаторах) и трехмерная ловушка.

Двухмерное квадрупольное поле. Распределение потенциала в декартовых координатах (x,y,z) в виде

$$V(x,z)=C(x^2-z^2) \tag{6.129}$$

может быть реализовано физически, так как лапласиан этого потенциала равен нулю. Величина C является константой. Потенциал равен нулю

на оси. В плоскости (x,z) эквипотенциальные кривые являются прямоугольными гиперболами.

Если мы введем проводящие электроды, повторяющие форму этих эквипотенциальных поверхностей, и подведем к ним соответствующее напряжение, то в каждой точке между электродами будет реализовано распределение потенциала, описываемое (6.129). В частности, если возвратные точки электродов расположены эквидистантно от оси y, то будет реализована конфигурация, показанная на рис. 6.33a, в которой

$$C = \frac{V_0}{x_0^2} = \frac{V_0}{z_0^2}, \tag{6.130}$$

где V_0 – абсолютное значение потенциала, до которого заряжается электрод, а $x_0 = z_0$ является расстоянием между осью y и возвратными точками электродов.

Компоненты электрического поля между электродами равны

$$E_x = -\frac{2V_0}{z_0^2}x, \quad E_y = 0, \quad E_z = \frac{2V_0}{z_0^2}z. \tag{6.131}$$

На практике электроды не могут быть бесконечными вдоль асимптот гипербол. Поэтому они отсекаются так, как это показано на рис. 6.33b. Для упрощения конструкции они обычно уменьшаются до простых стержней с круглым сечением. Затем показывается, что около нейтральной оси, там где фактически удерживаются ионы, распределение потенциалов приближается к распределению, данному в (6.129).

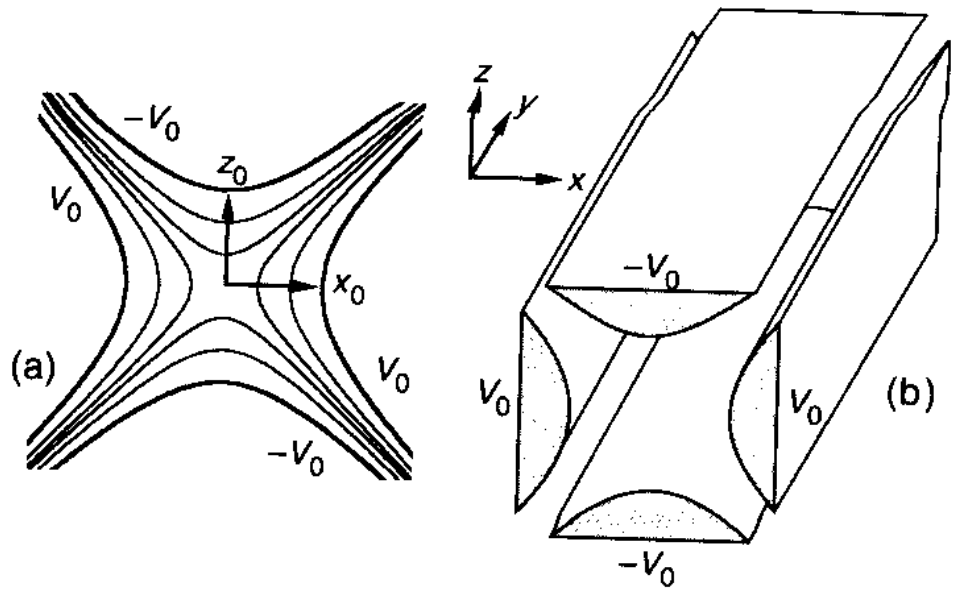


Рис. 6.33. Линейная ловушка (в двух размерах). (a) Эквипотенциальные линии в поперечной плоскости. (b) Структура электродов.

Трехмерное квадрупольное поле. Следующее распределение потенциалов, выраженное в цилиндрических координатах

$$V(r,z) = C(r^2 - 2z^2), \tag{6.132}$$

может также быть реализовано физически. Потенциал равен нулю в начале координат и эквипотенциальные поверхности являются гиперболоидами вращения вокруг оси z. Гиперболоиды сходятся на круговой поверхности, центр которой лежит на оси z, а две другие поверхности отрезаются осью z, и их можно рассматривать как крышки. Круговой электрод с минимальным расстоянием r_0 относительно оси z и крышки с возвратными точками на расстоянии $z_0 = r_0/\sqrt{2}$ от центра имеют одинаковое абсолютное значение потенциала. Если эти эквипотенциальные поверхности заменить физическими электродами и придать им потенциал V с $|V| = V_0$, то мы получим конфигурацию, в которой

$$C = \frac{V_0}{r_0^2} = \frac{V_0}{2z_0^2}, \tag{6.133}$$

как показано на рис. 6.34a.

Компоненты электрического поля между электродами выражаются следующим образом:

$$E_r = -\frac{V_0}{z_0^2} r, \quad E_z = 2\frac{V_0}{z_0^2} z. \tag{6.134}$$

Как и в предыдущем случае электроды фактически имеют конечные размеры. Практически они выполняются так, как показано схематически на рис.6.34b. Для удобства они могут не совсем повторять теоретическую гиперболическую форму. Однако вблизи центра потенциал адекватно представляется выражением (6.132).

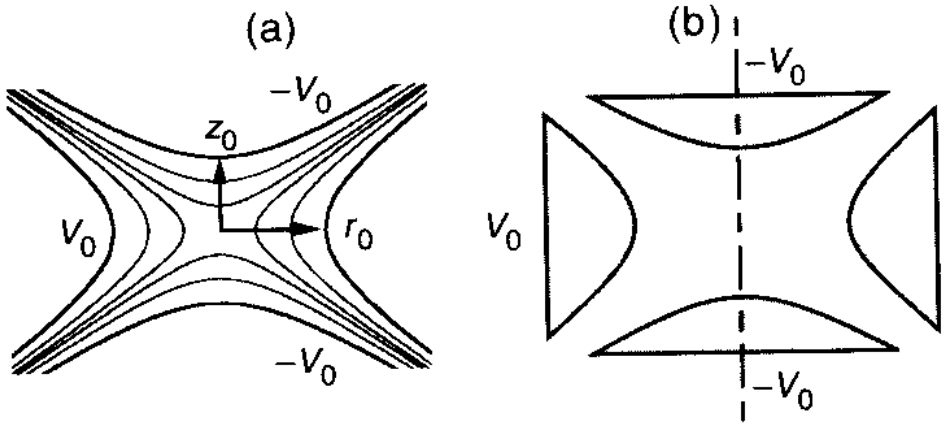


Рис. 6.34. Трехмерная ловушка. (a) Эквипотенциальные поверхности в меридиональной плоскости. (b) Структура электродов.

(b) Удержание заряженных частиц

Уравнения (6.131) и (6.134) показывают, что ортогональные компоненты электрического поля пропорциональны расстоянию от оси линейной ловушки или от центра трехмерной ловушки, но с противоположными знаками. Если приложить постоянный потенциал, то на заряженные частицы буду действовать силы одной компоненты, стремящейся вернуть частицы на ось или к центру, и другой компоненты, стремящейся выбросить их из ловушки.

Однако если использовать переменную разность потенциалов, то можно получить удержание частиц. Конечно, движение частиц может быть описано аналитически через фундаментальные уравнения динамики. В этом случае мы получим два дифференциальные уравнения Матье. Они имеют устойчивое решение для некоторых значений параметров, определяемых через амплитуду и частоту приложенной к электродам разности потенциалов, через параметры ловушки r_0 и z_0 и через массу и заряд частиц. В этом случае частицы осциллируют около оси или около центра, не достигая стенок. Таким образом физически реализуется удержание. Полное изучение движения ионов показывает, что условия устойчивости можно оптимизировать, если к разности потенциалов добавить постоянную составляющую [6.101].

В последующем мы представим простые качественные соображения по объяснению происхождения *возвратной силы*, которая удерживает ионы в обоих направлениях.

Рассмотрим линейную ловушку и без потери общности предположим, что в ней находится положительный ион. Силы электрического поля заставляют ион участвовать в так называемом *микродвижении* - синусоидальном движении с малой амплитудой. Это движение происходит с той же частотой, что и частота электрического поля, но совершенно с противоположной фазой. На рис.6.35 показано изменение электрического поля во времени в двух точках на плоскости (x, z), расположенных на x и на оси z с координатами $(x_1>0, 0)$ и $(0, z_1>0)$ соответственно. Электрическое поле находится в противофазе, так как из (6.131) E_x и E_z имеют противоположные знаки. Здесь также показано локальное микродвижение ионов. Это движение позволяет им обнаруживать электрическое поле. Однако, так как это поле неоднородно и пропорционально расстоянию до оси, то и электрическая сила, действующая на ион, искажается так, как это показано на рисунке. Следовательно, среднее значение силы не равно нулю, и мы видим, что сила направлена к оси. Этот возвратный эффект действует на ионы, заставляя их участвовать в двумерном поперечном гармоническом движении около оси ловушки. Это движение называется *макродвижением* или *вековым движением*. Это движение происходит с низкой

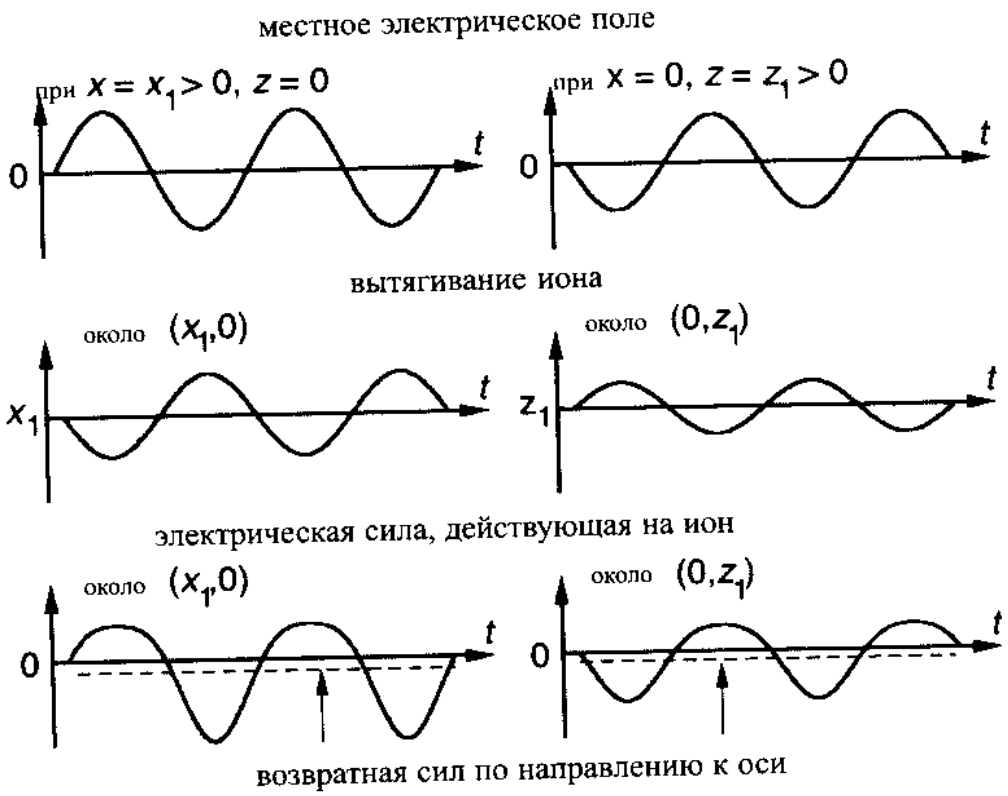


Рис. 6.35. Происхождение возвращающей силы, заставляющей ионы двигаться к оси в линейной ловушке

частотой по сравнению с частотой электрического поля. Продольное удержание обеспечивается двумя концевыми электродами, расположенными на оси и находящимися под постоянным отталкивающим потенциалом.

С помощью аналогичного объяснения можно показать, что в трехмерной ловушке удержание происходит в направлении оси и в любом радиальном направлении. В этом случае удержание действительно является трехмерным.

Путем наложения разности потенциалов с переменными составляющими на электроды линейной или трехмерной ловушки ионы могут удерживаться вблизи оси или центра соответственно. Ионы удерживаются в вакууме под действием сил электрического происхождения, ограничивающих движение ионов по отношению к электродам. Они участвуют в пространственном вековом движении, на которое накладывается высокочастотное микродвижение с малой амплитудой. Для ловушки, в которой расстояние между электродами составляет около одного сантиметра и приложенная переменная разность потенциалов около 100 В на частоте 100 к Гц, частота векового движения для ионов ртути составляет примерно 10 кГц, а амплитуда меньше, чем внутренние размеры ловушки.

При их почти гармоническом движении ионы ведут себя так, как будто они находятся в параболической потенциальной яме с глубиной около 10 В.

6.7.2. Удержание облака ионов

(a) Плотность ионного облака

Плотность ионного облака, удерживаемого таким способом, очень мала из-за электростатических сил отталкивания между заряженными частицами. Эта плотность составляет примерно 10^6 ионов в $см^{-3}$, что соответствует парциальному давлению в области облака 10^{-8} Па, сопоставимому с давлением остаточных газов в сверхвысоком вакууме. Таким образом, ионная среда очень разрежена и взаимодействия через столкновения являются редкими.

(b) Спектры поглощения ионов

Если ионы облучаются электромагнитной волной с постоянной частотой ν, то они чувствуют ее как частотно-модулированную волну. Это связано с их собственным движением и связанным с ним эффектом Доплера. Спектр сигнала, принимаемого ионами включает, таким образом, в себя несущую частоту ν и боковые компоненты на частотах $\nu \pm nf_\mu$ и $\nu \pm mf_{\mu'}$, где n и m – целые числа, а f_μ и $f_{\mu'}$ частоты векового движения и микродвижении соответственно. Так как время удержания ионов может быть очень большим, то они имеют очень резкий сверхтонкий резонанс. Его ширина много меньше, чем f_μ, а следовательно, и чем $f_{\mu'}$. Если частота ν изменяется, то ионы стремятся поглотить падающее излучение каждый раз, когда одна из спектральных компонент совпадает в пределах ширины резонанса со сверхтонкой резонансной частотой ν_0.

Если перейти с точки зрения иона на точку зрения наблюдателя, то все происходит так, как если бы спектр поглощения (или излучения) ионов состоял бы из очень узких линий на частотах $\nu_0 \pm nf_\mu$ и $\nu_0 \pm mf_{\mu'}$. Схематически спектр показан на рис. 6.36.

Если частота облучения ν очень близка к частоте сверхтонкого перехода ν_0, то боковые компоненты не возбуждаются и все происходит так, как если бы эффект Доплера не действовал. Этот результат уже упоминался в

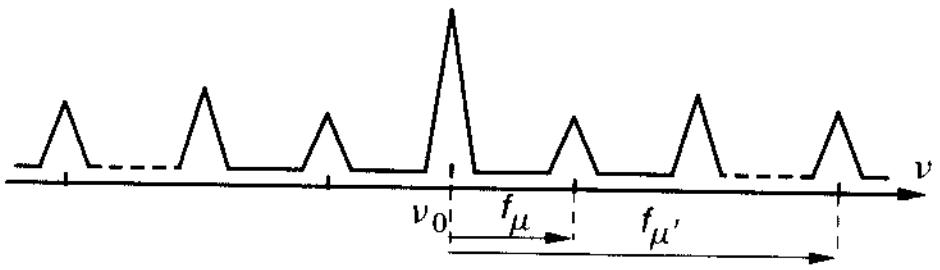


Рис. 6.36. Схематическая иллюстрация спектра поглощения ионов, удерживаемых в радиочастотной ловушке

разделе 6.2.3. Можно показать, что условия Дике (6.15), ограничивающее размеры области, занимаемой ионами, предполагает, что центральная компонента спектра поглощения почти не будет ослабляться, а боковые компоненты будут иметь малую амплитуду.

(c) Энергия ионов

Резонансный сигнал, который можно получить от группы атомов или ионов, пропорционален числу наблюдаемых частиц. Поэтому предпринимаются все усилия для увеличения числа частиц настолько, насколько это возможно. Однако для ионов существует корреляция между их числом и их энергией. Энергия является возрастающей функцией от числа ионов. То же самое справедливо и для сдвига частоты из-за эффекта Доплера второго порядка, в соответствии с (6.16).

Легко найти качественное объяснение этой связи. Из-за кулоновского отталкивания ионное облако занимает больший объем при увеличении числа ионов. Так как электрические силы, действующие на ион, пропорциональны расстоянию до оси или центра ловушки (из(6.131) и (6.134)), то кинетическая энергия ионов возрастает. Для трехмерной ловушки средняя кинетическая энергия близка к 2eV. Несмотря на большую массу иона ртути, эффект Доплера второго порядка приводит к сдвигу частоты до уровня – 10^{-11}.

Кинетическую энергию ионов можно уменьшить путем добавления гелия при очень низком давлении, порядка 10^{-3} Па, чтобы не вызвать дальнейший сдвиг частоты, который может еще ухудшить ситуацию. Средняя кинетическая энергия гелия при комнатной температуре равна 4×10^{-2} eV. Этот *столкновительный метод охлаждения* на практике не очень эффективен. Сдвиг частоты из-за эффекта Доплера второго порядка может быть уменьшен до -5×10^{-12}.

Для линейной ловушки ситуация лучше, так как ионы могут расширяться в цилиндрической области вдоль оси ловушки. Для данного числа ионов поперечный размер облака может быть уменьшен за счет увеличения его длины, большей чем его диаметр. Таким образом уменьшается средняя кинетическая энергия [6.107, 6.108, 6.109]. Использование гелия при давлении 10^{-3} Па уменьшает сдвиг из-за эффекта Доплера второго порядка до -10^{-12}. Можно также показать, что в линейной ловушке средняя кинетическая энергия менее чувствительна к флуктуациям переменной разности потенциалов на электродах.

Эти преимущества объясняют предпочтение линейной ловушки для удержания в стандартах частоты на ионах ртути.

6.7.3. Описание стандарта частоты на облаке ионов ртути

(a) Спектроскопические свойства $^{199}Hg^+$

Ядерный спин 199 изотопа ртути равен 1/2 и если создать ион ртути Hg^+ путем удаления электрона с внешней оболочки, то он будет иметь структуру, очень похожую на структуру щелочных металлов с одним непарным электроном на внешней оболочке.

Ион ртути $^{199}Hg^+$ имеет такую же простую сверхтонкую структуру основного состояния, как и атом водорода, имеющий лишь четыре сверхтонких подуровня, а именно: $F = 0$, $m_F = 0$, и $F = 1$, $m_F = -1, 0, 1$. Поэтому соотношение уровней, полезных для работы стандарта частоты, т.е. уровней (0, 0) и (1, 0), к общему числу уровней является максимально возможным. Это делает более доступным наблюдение сигнала сверхтонкого резонанса на ограниченном числе ионов. Отметим также большое сверхтонкое расщепление, соответствующее частоте около 40.5 ГГц. Для заданной ширины сверхтонкого резонанса добротность резонанса будет, таким образом, выше, чем для атомов, рассматриваемых ранее. Эти же самые свойства благоприятны для достижения более высокой стабильности частоты.

(b) Оптическая накачка $^{199}Hg^+$ и наблюдение сверхтонкого резонанса

Сверхтонкие состояния иона не могут быть разделены с помощью магнитной селекции, так как отклоняющие силы, связанные с электрическим зарядом много больше, чем силы, возникающие из-за магнитного момента непарного электрона. По этой причине селекция состояний осуществляется с помощью оптической накачки.

В данном случае существует удачное изотопическое совпадение между спектром поглощения $^{199}Hg^+$ и спектром излучения $^{202}Hg^+$. Спектр последнего излучается лампой, содержащей 202 изотоп ртути и возбуждаемой радиочастотным разрядом. Этот изотоп не имеет сверхтонкой структуры, и так получилось, что длина волны 194 нм линии D_1 изотопа $^{202}Hg^+$, очень близка к длине волны перехода между сверхтонким уровнем $F = 1$ основного состояния $^2S_{1/2}$ иона $^{199}Hg^+$ и его первым возбужденным состоянием $^2P_{1/2}$. Это положение очень похоже на положение показанное, на рис.6.4а, за исключением того, что падающий свет возбуждает ионы на уровень β (здесь уровень $F = 1$) и оптическая накачка заселяет уровень α (здесь уровень $F = 0$).

Сигнал флуоресценции, излучаемый ионом, спонтанно затухающим из первого возбужденного состояния в основное состояние, не очень интенсивен, потому что уменьшается населенность уровня $F = 1$. Если теперь ионы облучаются электромагнитной волной с частотой, близкой к частоте сверхтонкого перехода, то уровень $F = 1$ заселяется и свет флуоресценции на длине волны 194 нм увеличивается. Его интенсивность измеряется и при сканировании частоты сигнала относительно частоты сверхтонкого перехода ν_0 получается резонансная кривая с максимумом на частоте перехода.

Если оптическая накачка совершалась бы с помощью монохроматического света, то она была бы более эффективна и резонанс был бы более интенсивный. Однако в настоящее время (2002 год) еще нет непрерывных ультрафиолетовых лазеров. Когерентное излучение было получено на длине волны 194 нм с применением техники синтеза оптических частот, т.е. путем умножения и смешения частот в нелинейных кристаллах [6.111]. Экспериментальная установка очень сложная и дорогостоящая, поэтому все еще используется лампа на ионах $^{202}Hg^+$, которая является относительно компактным прибором.

(c) Экспериментальная установка и резонансная кривая

Спроектированная в лаборатории реактивного движения в США экспериментальная установка схематически показана на рис.6.37.

Нейтральные пары ртути 199 вводятся контролируемым способом при очень низком давлении путем нагрева окиси ртути. Пары ртути встречаются с потоком электронов из электронной пушки. Произведенные таким образом ионы удерживаются с помощью должным образом выбранной разности потенциалов на электродах линейной ловушки. Ловушка состоит из четырех симметрично расположенных цилиндрических стержней на расстоянии 1 см друг от друга. Кинетическая энергия ионов уменьшается гелием при давлении около 10^{-3} Па.

Свет от лампы на ионах $^{202}Hg^+$ направляется на ионное облако. Для того, чтобы уменьшить интенсивность рассеянного света, который мог бы мешать наблюдению флуоресцентного света от ионов этот рассеянный свет после прохождения ионной ловушки собирается в ловушке света. Свет флуоресценции фокусируется на фотоумножитель в направлении, перпендикулярном плоскости рисунка.

СВЧ сигнал возбуждения на частоте 40,5 ГГц подается через рупор. Электромагнитная волна распространяется под прямым углом к оси ловушки.

Магнитные экраны ослабляют влияние земного магнитного поля и его флуктуаций. Установка содержит катушки, создающие однородное и стабильное магнитное поле с напряженностью примерно такой же, как и у

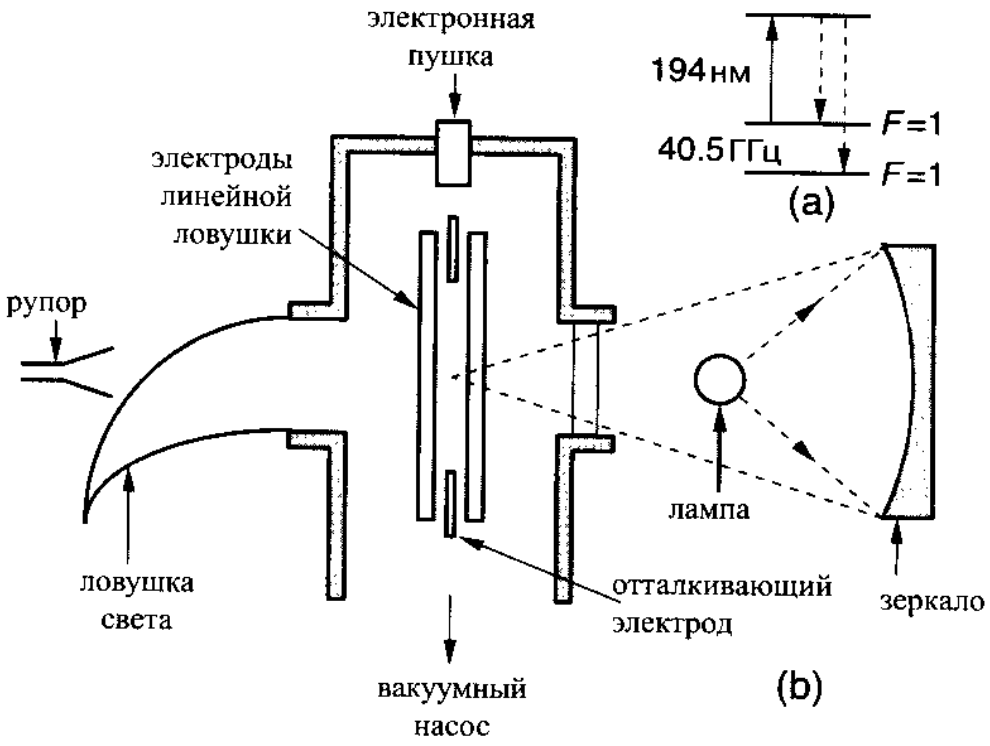


Рис. 6.37. Резонатор с удержанными ионами. (a) Полезная часть спектра $^{199}Hg^{+}$. (b) Упрощенное представление удержания иона ртути в резонаторе. Свет флуоресценции наблюдается под прямым углом к плоскости рисунка.

цезиевых часов. Используется высоковакуумный насос и ряд других приборов.

Резонанс наблюдается с помощью метода Рэмси. Однако в этом случае два последовательных взаимодействия с электромагнитным полем происходят не благодаря движению атомов (или ионов) как в цезиевых часах. В стандарте частоты на ионах $^{199}Hg^{+}$ два последовательных взаимодействия получаются путем создания импульсного осциллирующего поля. Во время последовательности наблюдений это поле облучает область, занятую ионом, двумя короткими импульсами, разделенными большим интервалом времени T. Таким образом, реализуется метод Рэмси во времени.

Последовательность действия при наблюдениях обычно начинается с перезагрузки ловушки, для того чтобы поддерживать число ионов (и, следовательно, их энергию) на постоянном уровне от одного цикла к другому. На короткое время включается электронная пушка. Затем примерно на 1,5 с зажигается лампа и за это время осуществляется подготовка атомов в состоянии $F = 0$. Включается СВЧ излучение на время около 0,3 с. Затем СВЧ поле выключается на промежуток времени T продолжительностью от 2 до 30 секунд и опять включается на 0,3 с. Опять включается лампа и измеряется интенсивность света флуоресценции для того чтобы обнаружить эффект СВЧ излучения. На рис. 6.38 показана полученная резонансная

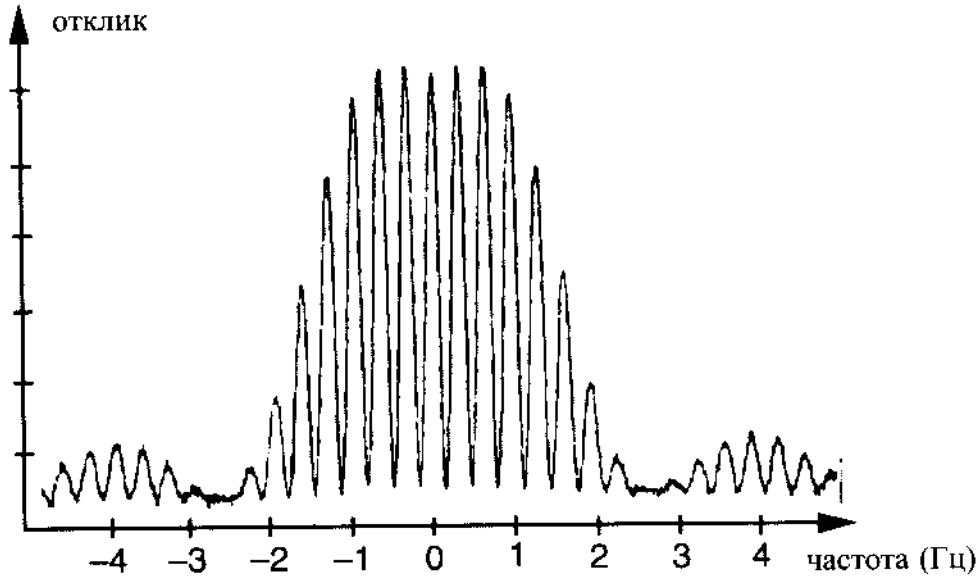


Рис. 6.38. Отклик линейной ловушки ионов ртути. Начало частот совмещено с частотой перехода в сверхтонком уровне основного состояния. Из [6.107, 6.108, 6.109].

кривая при $T = 2{,}5$ с [6.107, 6.108, 6.109]. Она представляет вероятность перехода, как представлено в (6.43).

Добротность Q_{at} центральной полосы резонансной кривой на рис.6.38 составляет 2×10^{11}. Она может быть увеличена пропорционально T. Временной метод Рэмси, использованный в удержанных ионах, в настоящее время повсеместно применяется в спектроскопии сверхвысокого разрешения для получения наибольшей добротности атомных резонансов.

(d) Управление частотой генератора

Отклик атомной системы может наблюдаться только в дискретные времена в конце каждой последовательности взаимодействия. Частота кварцевого генератора управляется по ионному сверхтонкому резонансу таким же способом, как это осуществляется на фонтане холодных атомов.

6.7.4. Метрологические свойства стандартов частоты на облаке холодных ионов ртути

(a) Нестабильность частоты за средние времена

Если времена выборки велики по сравнению с постоянной времени цепи подстройки кварцевого генератора и, следовательно, дольше T, то нестабильность частоты стандарта на удерживаемых ионах ртути определяется из (6.86).

Шум детектирования определяется дробовым шумом в детектируемом потоке света. В нем присутствуют две компоненты, а именно: – флуоресцентный свет от ионов и свет от нежелательных отражений падающего

света. На практике падающий свет много интенсивнее света флуоресценции. Это приводит к необычайно большому дробовому шуму. Однако высокое значение добротности Q_{at} благоприятствует достижению высокой стабильности частоты. Была получена вариация Аллана около $7\times10^{-14}\tau^{-1/2}$, что, как можно увидеть на рис.6.11, ставит стандарт на ионах ртути в ранг одного из самых стабильных, [6.107, 6.108, 6.109].

(b) Остаточные сдвиги частоты

Основной сдвиг частоты, влияющий на стандарт частоты на ионах ртути, происходит из-за квадратичного эффекта Зеемана, эффекта Доплера второго порядка и из-за столкновений с буферным газом – гелием.

Статическое магнитное поле имеет типичное значение $3{,}5\times10^{-6}$ Т. Сдвиг частоты сверхтонкого перехода в ионе ртути $^{199}Hg^{+}$ в основном состоянии равен в этом случае 3×10^{-12}. Это много меньше, чем в цезиевых пучковых часах, в которых этот сдвиг имел бы значение $1{,}7\times10^{-10}$.

Сдвиг частоты из-за эффекта Доплера второго порядка и из-за столкновений с буферным газом связаны между собой, так как плотность буферного газа определяет кинетическую энергию ионов. При парциальном давлении гелия 10^{-3} Па сдвиг частоты из-за обоих связанных между собой эффектов составляет около 10^{-12}.

Отметим, что несмотря на применение удерживающих электрических полей квадратичный эффект Штарка очень мал и вызываемый им сдвиг частоты составляет несколько единиц на 10^{-14}.

(c) Долговременная нестабильность частоты и неточность

Долговременная нестабильность частоты зависит от чувствительности наблюдаемой частоты перехода к окружающим условиям и рабочим параметрам. Эта нестабильность может происходить из-за флуктуаций: окружающего магнитного поля, разницы потенциалов на электродах, числа удерживаемых ионов, температуры и давления буферного газа.
Опыт показывает, что эти параметры могут с достаточной степенью контролироваться для того чтобы обеспечить нестабильность частоты порядка 10^{-15} для $10^{4}<\tau<10^{6}$ с [6.107, 6.108, 6.109].

Теоретический и экспериментальный анализ различных частотных сдвигов дает оценку неточности стандарта частоты на ионах ртути $^{199}Hg^{+}$ порядка нескольких единиц на 10^{-13} [6.107, 6.108, 6.109].

(d) Текущие разработки

Ловушка разделяется на две электрически изолированные коллинеарные части и ионы могут перемещаться из одной части в другую при применении соответствующего потенциала постоянного тока [6.109]. Подготовка атомных состояний завершается в одной части, а взаимодействие с СВЧ

полем в другой. Эта конфигурация устраняет некоторые технические препятствия и ожидается улучшение характеристик.

6.7.5. Стандарт частоты на ионах иттербия

Ион $^{171}Yb^{+}$ также был использован для создания стандарта частоты. Он тоже имеет ядерный спин равный 1/2, а частота его сверхтонкого перехода в основном состоянии близка к 12,6 ГГц. Оптическая накачка осуществляется на линии D_1 на длине волны 369,5 нм. Стандарт создавался различными способами. В одном случае использовалась трехмерная радиочастотная ловушка, а лазерный источник получался из лазера на красителе с удвоением частоты [6.112]. В другом случае радиочастотная ловушка была линейной, а источником света накачки являлся титан-сапфировый лазер [6.113].

Измеренная нестабильность частоты этих стандартов близка к $10^{-13}\tau^{-1/2}$ для $100 < \tau < 3000$ с.

Частота сверхтонкого перехода иона иттербия $^{171}Yb^{+}$ в основном состоянии в измерена с точностью до 2 миллигерц [6.113].

6.7.6. Стандарт частоты на ионе ртути с лазерным охлаждением

Ионы можно охладить лазерным излучением также как и атомы [6.114]. Фактически экспериментальные методы для ионов проще, чем для нейтральных атомов, потому что ионы удерживаются вблизи оси или центра ловушки. Первое радиационное охлаждение удержанных ионов было осуществлено в 1978 году в национальном бюро стандартов, переименованным в национальный институт стандартов и технологии (США).

Механизм охлаждения можно понять на уровне интуитивного мышления. Из-за периодических компонент в движении ионов их оптический спектр поглощения имеет боковые компоненты точно также как и СВЧ спектр поглощения, показанный на рис.6.36. Поэтому ион может поглощать фотоны при облучении его лазерным пучком с частотой излучения ниже частоты его оптического перехода. Далее ион излучает флуоресцентный фотон, средняя частота которого равна частоте оптического перехода. Каждый цикл поглощения – флуоресценции отбирает кинетическую энергию иона. При этом достаточно одного лазера и остаточная тепловая энергия, соответствующая температуре около 10 мК, является обычным результатом. Остаточный эффект Доплера второго порядка при таких температурах приводит к сдвигу частоты порядка 10^{-17} или 10^{-18}.

Такие низкие температуры могут достигаться, если число удержанных ионов мало. В этом случае ионы могут локализоваться на оси ловушки, где они образуют линейный кристалл из нескольких ионов. Так как ионы располагаются вдоль узловых линий радиочастотного поля, то

квадратичный эффект Штарка становится пренебрежимо малым. К сожалению, стабильность частоты также уменьшается с уменьшением числа ионов. Это компенсируется уменьшением ширины сверхтонкой резонансной линии благодаря увеличению до 100 секунд времени T между двумя взаимодействиями во временном методе Рэмси. Сдвиг частоты из-за столкновений с остаточными газами может быть уменьшен до минимально возможной величины путем разрежения и применения криогенной техники.

Согласно только что описанным принципам [6.110] был создан и работает стандарт частоты на ионах ртути с лазерным охлаждением в NIST. Когерентное излучение на длине волны 194 нм реализуется с помощью цепи синтеза оптических частот [6.111]. Оно используется для охлаждения ионов и также для создания разности населенностей между сверхтонкими подуровнями. Ионы удерживаются в линейной радиочастотной ловушке длиной 4 мм с расстоянием между электродами около 1 мм. По опубликованным рабочим условиям набор из семи холодных ионов образует линейный кристалл вдоль оси ловушки. Ловушка расположена в объеме, охлаждаемом до температуры жидкого гелия. Часовой переход реализуется с помощью временного метода Рэмси. Во время одного элементарного цикла работы ионы сначала в течение 300 мс охлаждаются. Затем они подвергаются оптической накачке для заселения уровня $F = 0, m_F = 0$ основного состояния. Свет выключается и осуществляется СВЧ взаимодействие по временному методу Рэмси. Для этого применяются два импульса на частоте 40,5... ГГц с длительностью по 250 мс и временным разделением между ними до 100 секунд. Затем оптическим способом детектируется часовой переход. Управление частотой кварцевого генератора очень похоже на управление, используемое в атомных фонтанах. Нестабильность частоты и неточность достигают $3{,}3\times10^{-13}\tau^{-1/2}$ и $3{,}5\times10^{-15}$ соответственно [6.110]. Улучшение характеристик ожидается путем увеличения числа удержанных ионов или их подготовкой в специальных квантовых состояниях [6.115, 6.116].

6.8. Другие атомные стандарты

Применение переходов в сверхтонкой структуре не являются непременным требованием для создания опорных атомных стандартов. Другие переходы также могут быть приемлемы при условии, что они отвечают требуемому качеству.

6.8.1. Стандарт частоты на пучке магния

Щелочно-земельные атомы не имеют сверхтонкой структуры в основном состоянии. Однако существуют интересные переходы между уровнями их возбужденных состояний. Например, в магнии частота одного из пере-

ходов равна 601 ГГц и время жизни уровня таково, что можно получить добротность 10^{10}. На основе этого перехода был создан пучковый стандарт частоты с оптимистическими результатами [6.117]. Для улучшения свойств прибора может потребоваться более эффективная подготовка необходимых атомных состояний и охлаждение атомов магния.

6.8.2. Оптические стандарты частоты

Очень много стандартов частоты было разработано или находится в состоянии исследований в оптической части спектра. Они используют атомные или молекулярные переходы между основным и возбужденным состояниями или между двумя возбужденными состояниями. Источником света для наблюдения таких переходов всегда является лазер, частота которого подстраивается по частоте выбранного перехода. Поэтому эти стандарты являются пассивными стандартами частоты. Опорными являются переходы в следующих атомах, молекулах и ионах, например: I_2, CH_4, CO_2, OsO_4, Rb, Ca, Hg^+, Yb^+, Ba^+, Sr^+, Ca^+, In^+. Здесь применяются всевозможные непрерывные лазеры, включая He-Ne, CO_2, полупроводниковые, легированные Nd:YAG, титан-сапфировые и т.д.

Были получены значительные результаты. На частоте 29 ТГц (10 мкм) CO_2 лазер, стабилизированный по переходу в молекуле четырехокиси осмия OsO_4, имеет нестабильность частоты ниже 10^{-14} за 10 с $< \tau <$ 10^4 с и долговременную (за 1 год) воспроизводимость $1{,}3\times10^{-13}$ с потенциальной неточностью около 10^{-14} [6.118]. Похожие результаты были получены на 88 ТГц (3.39 мкм) с He-Ne лазером, стабилизированным по поглощению в молекуле метана [6.119].

6.8.3. Оптический стандарт частоты на одиночном ионе в покое

Техника удержания ионов вероятно будет играть ключевую роль в разработке исключительно высококачественных оптических стандартов частоты. Еще в 1973 году Демелт выдвинул идею использования *одиночного иона в покое* [6.120].

Рассмотрим ион, в котором могут быть два перехода с совершенно различными свойствами. Один переход связывается с уровнем основного состояния за очень короткое время жизни, порядка 10 нс. Это означает, что около 10^8 циклов поглощения – флуоресценции могут произойти за одну секунду и таким образом ион может быть очень эффективно охлажден. Другой переход относится к переходам, имеющим очень малую вероятность. Он связывает основное состояние с метастабильным состоянием, время жизни которого около 1 с. Соответствующий резонанс по этой причине проявляет присущую ему ширину порядка 1 Гц. Это служит основой создания оптической опор-

ной частоты с добротностью около 10^{14} и 10^{15}. Так как охлаждающий переход обеспечивает около 10^8 флуоресцентных фотонов в секунду, то наличие иона в ловушке легко обнаруживается. И действительно это было экспериментально продемонстрировано в ряде лабораторий.

Одиночный ион пребывает на оси линейной ловушки Пауля с исключительно малой кинетической энергией. Таким способом достигается хорошее приближение к идее одиночного атома (или иона) в покое. При такой ситуации причины, вызывающие нежелательные сдвиги частоты уменьшаются до абсолютного минимума.

Возможность использования иона индия $^{115}In^+$ была рассмотрена в работах [6.121, 6.122]. Соответствующие уровни энергии показаны на рис.6.39. Уровень $^3P_0$ является долгоживущим. Этот ион не имеет электронного углового момента ни в основном состоянии $^1S_0$, ни в метастабильном состоянии $^3P_0$. Поэтому зеемановский сдвиг частоты является исключительно малым. Более того, частота опорного перехода не сдвигается из-за эффекта Штарка и поэтому она полностью безразлична к присутствию удерживающего электрического поля. Разработка двух источников излучений на длинах волн 231 и 236 нм кажется достижимой путем удвоения частоты излучения полупроводникового лазера.

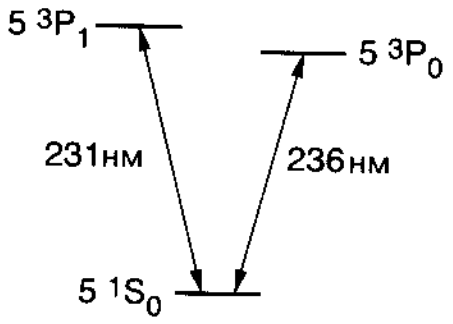


Рис. 6.39. Уровни энергий иона индия выбранные для построения оптического стандарта частоты.

Опорные переходы могут детектироваться так называемым *методом полки*, работающим следующим образом [6.120]. Если ион находится в основном состоянии $^1S_0$, он детектируется в результате флуоресценции на длине волны 231 нм, происходящей после облучения лазерным охлаждающим пучком. С другой стороны, если частота излучения на 236 нм настроена на частоту опорного перехода, то ион попадает в состояние $^3P_0$ и флуоресценция исчезает.

Ожидается отличная стабильность частоты такого стандарта, потому что линия имеет очень высокую добротность и хорошее отношение сигал/шум для детектирования перехода. Неточность может достигать 10^{-17}, потому что сдвиги, действующие на частоту, значительно уменьшены.

Убедительные экспериментальные результаты для исключения эффектов остаточных столкновений были недавно получены в NIST при использовании одиночного иона $^{199}Hg^{+}$, охлаждаемого лазером в криогенной линейной радиочастотной ловушке [6.123]. Охлаждение иона достигается путем индуцирования перехода на 194 нм между основным состоянием $^{2}S_{1/2}$ и первым короткоживущим возбужденным состоянием $^{2}P_{1/2}$. Опорный переход на 282 нм происходит между основным состоянием и метастабильным состоянием $^{2}D_{5/2}$, имеющим время жизни 90 мс. Применяется метод полки и детектирование перехода на 282 нм совершается путем наблюдения света флуоресценции на 194 нм. Была получена ширина линии 6,7 Гц на переходе на 282 нм ($1{,}06\times10^{15}$ Гц). Соответствующая добротность составляет $1{,}6\times10^{14}$. Это наибольшая добротность, когда либо достигнутая в СВЧ или оптической спектроскопии. Этот результат доказывает потенциальные возможности оптических стандартов частоты успешно конкурировать со стандартами частоты, основанными на СВЧ переходах.

Следует, однако, заметить, что для того чтобы наблюдать резонанс в оптическом диапазоне с такой малой шириной как 1 Гц сам лазерный источник должен иметь спектральную ширину ниже этого значения. Для создания таких источников необходимо проведение специальных исследований.

6.8.4. Оптические стандарты для измерения времени

Частота излучения на длине волны, например, 0,6 мкм равна 500×10^{12} Гц. Это примерно в 5000 раз выше, чем самые высокие частоты (~100 ГГц), используемые для телекоммуникации и радиолокации. Хотя ряд специализированных лабораторий и способен осуществлять сличение СВЧ стандартов с такими оптическими стандартами с помощью техники синтеза частот, тем не менее экспериментальные установки являются искусными, громоздким и очень дорогостоящими. На время написания книги еще не было открыто практических способов использования оптических стандартов частоты для производства надежной и продолжительной шкалы времени.

С целью улучшения ситуации продолжаются и развиваются исследования в этом направлении. Новые результаты открывают путь для существенно более простого способа сличения частот СВЧ и оптического диапазонов [6.124, 6.125]. Они получены путем применения лазеров с сихронизацией мод, которые равномерно излучают очень узкие (несколько фемтосекунд) и очень интенсивные импульсы. Периодическое излучение создает гребенку частот в оптическом диапазоне. Скорость повторения импульсов и, следовательно, частота каждой компоненты гребенки может быть синхронизирована по СВЧ стандарту частоты.

Таблица 6.2. Практическая информация по атомным стандартам частоты и времени. Цифровые значения указывают порядок величины. Они соответствуют разным уровням по характеристикам и по степени разработок.

Тип стандарта	Объем ($дм^3$)	Масса (кг)	Цена (10^3 евро)
Лабораторный цезий	500-3000	500-1000	1000
Промышленный цезий	10-30	10-30	50-80
Активный водородный мазер	150-250	75-250	150-450
Пассивный водородный мазер	20-80	20-60	75-150
Рубидий	0,2-1	0,4-1,5	3-8
Ионы ртути	10-250	10-250	150-2000

Таблица 6.3.Типовые неточности для различных атомных стандартов на 2000 год

Тип стандарта	Неточность
Лабораторный цезий: фонтан холодных атомов	1×10^{-15}
Лабораторный цезий: непрерывный тепловой пучок	$6 \times 10^{-15} - 10^{-13}$
Промышленный цезий	$1{,}5 \times 10^{-12} - 3 \times 10^{-11}$
Водородный мазер	$(2 - 5) \times 10^{-12}$
Удержанные ионы ртути	$3{,}5 \times 10^{-15} - 2 \times 10^{-13}$
Рубидий	Не имеет смысла (необходима калибровка)

6.9. Заключение

В этой главе мы описали передний фронт состояний дел по вопросам атомного времени и стандартов частоты. Таблица 6.2 дает некоторые детали, относящиеся к объему, массе и стоимости таких приборов.

На рис.6.11 показан обзор по нестабильностям на сегодняшний день. Другие важные характеристики, такие как неточность, как итог представлены в таблице 6.3.

Мы также упомянули современные тенденции в разработках. Можно надеяться, что через некоторое время будут создаваться стандарты частоты с нестабильностью 10^{-16} на интервале τ =1 день и с неточностью, приближающейся к 10^{-16}. Существуют также возможности дальнейших улучшений. Обращаясь к рис. 4.3, на котором схематически показано качество лучших разработанных часов, мы, таким образом, можем полностью подтвердить оптимизм в нашей экстраполяции на время после 2000 года.

Глава 7

Измерения атомного времени

В предыдущих главах была описана научная и техническая работа, которая служит фундаментом для измерения атомного времени. В главе 4 показано, как такие измерения постепенно вытеснили астрономические измерения. В настоящей главе, после напоминания определения атомной секунды и определения основных шкал времени, основанных на атомной секунде, мы будем рассматривать пути, какими атомное время нашло всеобщее применение. Мы опишем все ступени, необходимые для установления и передачи мировых шкал времени, чтобы они могли использоваться и для синхронизации, и как способ воспроизведения единицы времени. Мы также обсудим шкалы времени, подходящие для динамических задач.

7.1. Определения

7.1.1. Определение секунды

Давайте припомним определение, принятое 13-й генеральной конференцией по мерам и весам (CGPM) в 1967 году:

Секунда – это длительность 9 192 631 770 периодов излучения, соответствующего переходу между двумя сверхтонкими уровнями основного состояния атома цезия 133.

Как объяснялось в главе 3, секунду следует рассматривать как единицу собственного времени вдоль мировой линии, относящейся к атому цезия. Так как атомы движутся по отношению к прибору, который наблюдает их частоту, т.е. по отношению к часам, то должны вводиться поправки, учитывающие это движение, чтобы можно было оценить собственное время часов. На современном уровне точности единственным релятивистским эффектом, который необходимо принимать во внимание, является эффект Доплера второго порядка – эффект специальной теории относительности. Однако скоро может прийти время, когда все часы необходимо будет трактовать в рамках ОТО, для того чтобы обеспечить соб-

ственное время в заранее оговоренной точке, в такой, например, как конкретный разъем.

Если часы находятся в непосредственной близости друг с другом и они неподвижны относительно наблюдателя, то ими можно пользоваться как и в классической физике при условии, что мы принимаем принцип эквивалентности Эйнштейна. Если они находятся на некотором расстоянии, например, на спутнике, и они передают данные с помощью радио, то тогда эти данные должны быть обработаны так, как указано в главе 5.

7.1.2. Международное атомное время и его связь с геоцентрическим и барицентрическим координатными временами

Как объяснялось в главе 3 (раздел 3.3.2e), международное атомное время TAI можно рассматривать как среднее время по показаниям часов. Часы в этом случае работают в соответствии с определением секунды. Предполагается, что они неподвижны относительно Земли и расположены на вращающемся геоиде. Эти определения устанавливают *единицу шкалы* TAI, но все же вся система градуировки должна как-то настраиваться совместно. На эту проблему ссылаются как на *установление начала*. Задача была завершена путем присвоения одной и той же даты в TAI и во всемирном времени (UT) некоторому фиктивному событию, происшедшему 1 января 1958 года в 0 часов UT. (Если более точно, то использовалась форма UT2 всемирного времени, определенная в разделе 8.1.3.). UT основано на вращении Земли и не может быть измерено с такой же точностью как TAI. Более того, в последующем признавалось, что были допущены систематические ошибки при определении UT в 1958 году, и что значения UT должны были корректироваться на несколько миллисекунд. Однако эти поправки не были применены к TAI. Поэтому мы можем сказать, что начало TAI просто фиксируется самим своим появлением с 1 января 1958 года и что она приблизительно совпадает с UT.

В главе 3 мы определили геоцентрическое координатное время. Единица шкалы для этого идеального времени получена теоретически из определения секунды. Однако и в этом случае надо установить начало. Как раз это и сделал IAU в 1991 г. в своей резолюции A4 путем определения частного геоцентрического координатного времени TCG, такого, что

$$TCG - TAI = 32{,}184 \text{ с, 1 января 1977 года в 0 часов TAI.} \qquad (7.1)$$

Появляющееся здесь число 32,184 должно по соглашению трактоваться как точное число на эту дату. Отметим, однако, что существует неопределенность около 1 мкс в показаниях TAI на данную эпоху, так что можно сказать, что TCG выдерживается в пределах этой величины. На практике это не имеет значения.

Мы также видели в главе 3, что другой вариант геоцентрического координатного времени был определен через единицу шкалы с длительностью, очень близкой к длительности (собственной) секунды на вращающемся геоиде. Этим вариантом является земное время (TT), связанное с TCG через выражение разности в секундах

$$TCG - TT = L_G \times 86\,400 \times \Delta D, \tag{7.2}$$

где ΔD – есть время, накопившееся в TAI сутках с 1 января 1977 года 0 часов TAI, и L_G – имеет согласованное значение[1]

$$L_G = 6.\,969\,290\,134 \times 10^{-10}. \tag{7.3}$$

Таким образом, TAI с отстройкой 32,184 с, является реализацией TT. Поэтому мы можем написать

$$TT(TAI) = TAI + 32{,}184 \text{ с}, \tag{7.4}$$

применяя общепринятые обозначения, когда $U(x)$ представляет реализацию теоретической величины U, промаркированной символом x. Следовательно, TT(TAI) должно читаться как «реализация TT на основе TAI».

Читатель может поинтересоваться откуда берется разница в 32,184 между TT и TAI. В 1958 году было закреплено положение TAI по отношению к UT, а не к эфемеридному времени (ET), потому что последнее было определено довольно неточно. Однако позже астрономам захотелось заменить TT на ET. Все попытки или сдвинуть TAI, или изменить аргумент в астрономических эфемеридах оказались безуспешными. Число 32,184 представляет наилучшую оценку отклонения в секундах между ET и TAI на 1 января 1977 года.

Рассмотрим теперь барицентрическое координатное время, упомянутое в разделе 3.3.2b. Оно используется при изучении планетарного движения и траекторий космических зондов. Слово «барицентрический» означает «совмещенный с центром масс Солнечной Системы». Так же как и для TCG единица шкалы барицентрического координатного времени основана на (собственной) секунде. Начало одного из частных барицентрических координатных времен, обозначаемого TCB, выбрано так, что TCB имеет тоже показание, что и TCG на 1 января 1977 года в 0 часов TAI в центре масс Земли. Существенно уточнить «в центре масс Земли», потому что соотношение между этими двумя координатными временами является четырехмерным, с точностью до c^{-2} это соотношение дается следующим выражением[2].

---

[1] Соотношение (7.2) является приблизительным согласно решениям, принятым по TAI в 2000 году. ΔD должно отсчитываться в сутках TCG с 1 января 1977 года в 0 часов TAI. Разность составляет около 10^{-18} в относительной частоте. Во времени оно составляет 3 нс за столетие.

[2] Более точное выражение приводится в Рекомендации MAC B1-5 (2000).

$$TCB - TCG = \frac{1}{c^2}\left\{\int_{t_0}^{t}\left[\frac{\upsilon_E^2}{2} + U_p(\mathbf{x}_E)\right]dt + \mathbf{v}_E \cdot (\mathbf{x} - \mathbf{x}_E)\right\}, \qquad (7.5)$$

где $\mathbf{x}_E$ и $\mathbf{v}_E$ обозначают барицентрическое положение и скорость центра масс Земли, параметр $\mathbf{x}$ является барицентрическим положением наблюдателя и U_p это ньютоновский гравитационный потенциал всех тел Солнечной Системы за исключением самой Земли. В интеграле $t = TCB$ и t_0 выбирается в соответствии с началом координат, установленном ранее. На поверхности Земли последний член достигает максимального значения 2,1 мкс.

Для земного наблюдателя интеграл в (7.5) приводит к приращению, пропорциональному времени со скоростью около 0,5 с за год, к которому надо также добавить периодические члены. Наибольший из них происходит от эллиптической формы орбиты Земли. Этот периодический член имеет годовой период, а его амплитуда равна 1,7 мс. Для того чтобы достичь точности в 100 нс необходимо учесть более сотни периодических членов в [7.1].

К 1970-м годам появилась необходимость применения релятивистской модели для вычисления эфемерид тел в Солнечной Системе. В 1976 году IAU (MAC) определил временной аргумент в эфемеридах путем проведения преобразования единицы шкалы в барицентрическом координатном времени так, чтобы остались только периодические члены в ее разности с TT. В 1979 году новая шкала времени получила название динамическое барицентрическое время TDB. В это время некоторые астрономы высказывали сомнения в справедливости ОТО, поэтому все основные положения этой теории и ее терминологии тщательно избегались. Соотношение между TDB и TCB выражается в секундах через

$$TCB - TDB = L_B \times 86\,400 \times \Delta D, \qquad (7.6)$$

где ΔD имеет то же самое значение, что и в (7.2), а L_B – в настоящее время оценивается как

$$L_B = 1{,}550\,519\,768 \times 10^{-8} \pm 2 \times 10^{-17}. \qquad (7.7)$$

Как и TT, носившее в период между 1979 и 1991 годами название TDT (земное динамическое время), так и TCB и TDB стали основываться на TAI с помощью соотношений (7.2), (7.5) и (7.6). Использование TDB вместо TCB приводит к теоретическим усложнениям. Однако был выполнен большой объем работы чтобы принять TDB за временной аргумент, особенно в конструкциях эфемерид JPL (JPL – лаборатория реактивного движения), которые используются для планетарных исследований. По этой причине трудно сделать желательный переход к TCB.

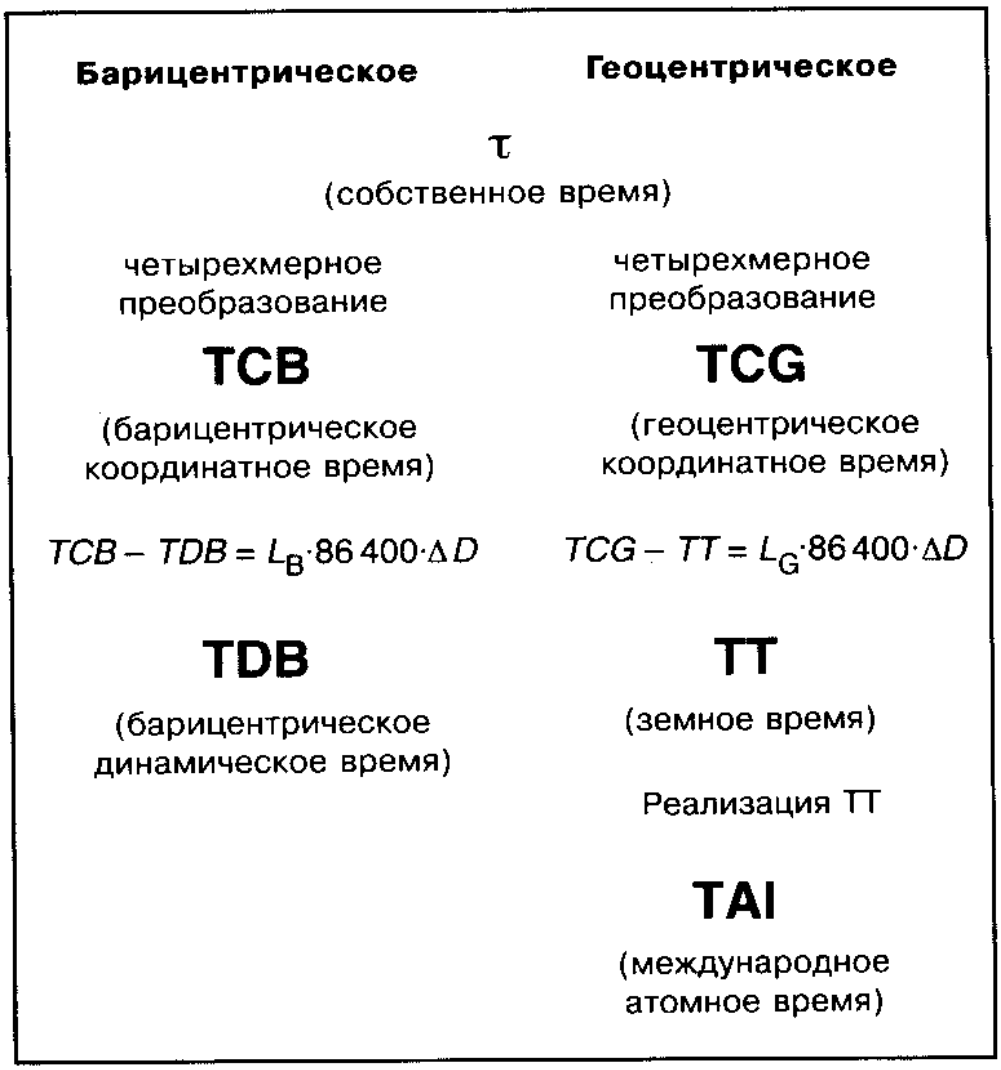


Рис. 7.1. Соотношения между координатными шкалами времени, начиная с собственного времени наблюдателя τ. Барицентрические и геоцентрические шкалы связаны четырехмерными преобразованиями, такими, что TAI может быть использована для реализации всех теоретических времен. Величины L_B, L_G и ΔD – определены в тексте.

На рис. 7.1 просуммированы соотношения между различными координатными шкалами времени.

7.2. Установление международного атомного времени

7.2.1. Алгоритмы шкалы времени

Если доступно несколько атомных часов, то их показания могут использоваться для установления единой шкалы времени. Она конечно более надежна, чем шкалы времени, основанные на вкладе каждых часов, взятых по отдельности, и можно надеяться, что она имеет лучшие характеристики, чем любая шкала отдельных часов.

Если часы достаточно локализованы в том смысле, что они не разбросаны очень широко, они могут быть использованы для того чтобы установить собственное время их лабораторий без необходимости учета любых релятивистских эффектов кроме эффекта Доплера второго порядка, вытекающего из специальной теории относительности. Если часы распределены по стране или по всему миру, то обычно конструируют реализацию земного времени TT.

Это обычно происходит, например, когда устанавливается TAI. В этом случае собственное время τ_C каждых часов C должно быть преобразовано в приближение t_C времени TT, прежде чем данные могут быть объединены надлежащим образом для установления локальной шкалы времени. Это преобразование вводит произвольные дополнительные константы, но мы увидим, что их присутствие не приводит к препятствиям. Для того чтобы упростить обозначения, t_C будет на протяжении этой главы писаться как t_C.

Прежде чем установить алгоритм для некоторой шкалы времени, которую мы намерены сконструировать, должен быть составлен, по крайней мере, следующий перечень вопросов:

- На каком периоде усреднения мы хотим оптимизировать стабильность частоты?
- Хотим ли мы иметь определенную точность частоты и если так, то что будет приемлемым компромиссом в случае, если это требование влияет на стабильность?
- Какой период должен быть между данными по сличению часов (период сбора данных)?
- Какой интервал длительности должен быть между данными, с которыми вычисляемые шкалы становятся доступными, и какой интервал до последней даты, для которой эта шкала была вычислена?
- Как часто должны обновляться вычисления?

Различные используемые промежутки времени показаны на рис. 7.2.

Ответы на эти вопросы взаимозависимы. Например, если мы добиваемся долговременной стабильности, мы можем принять бóльшие интервалы между сличениями часов по сравнению с интервалами, необходимыми для кратковременной стабильности. Тем не менее в диапазоне различных случаев вопрос остается открытым. В некоторых применениях шкала должна быть получена почти в реальном масштабе времени, в то время как для TAI является приемлемым двухмесячный интервал (в настоящее время 30 суток). Для изучения пульсаров, у которых предполагается высочайшая возможная добротность, никто не будет жаловаться на необходимость ожидания в течение целого года. Поэтому невозможно описать типовой алгоритм. Кроме того алгоритмы подвержены техничес-

ким веяниям, таким как, например, изобретение новых часов или новые методы сличения времени. Тем не менее все алгоритмы проявляют некоторые общие свойства, хотя они до некоторой степени могут быть спрятаны с помощью математического формализма. Эти свойства могут быть проявлены путем испытания с помощью алгоритма, установленного МБВ в 1973 г. и все еще используемого, с некоторыми улучшениями приспособленным к современным техническим достижениям [7.2, 7.3], в МБМВ для установления TAI.

7.2.2. Алгоритм для международного атомного времени

(a) Общая организация

Международное атомное время в настоящее время основано на 200 атомных часах и стандартах частоты из примерно шестидесяти различных лабораторий, сличаемых через GPS и российский GLONASS, а также методом дуплексной спутниковой передачи времени и частоты (TWSTFT) (см. раздел 5.6.3).

Алгоритм установлен таким способом, чтобы обеспечить точность и долговременную стабильность частоты на временах от одного месяца или более. Благодаря долговременной стабильности становится возможным установить достаточно большое время для публикации TAI (от одного до двух месяцев). При этом действительно разность времени между локальными часами и TAI может быть экстраполирована так, чтобы получить соответствующее приближение к TAI в реальном масштабе времени.

TAI устанавливается в два приема. *Алгоритм стабильности* дает нам так называемую *свободную атомную шкалу* (EAL), т.е. свободную среднюю шкалу, стабильность которой оптимальна для соответствующих периодов выборки данных. Затем долговременная система автоподстройки, основанная на частоте первичных стандартов, создает TAI путем использования зависящих от времени поправок $[TAI - EAL](TAI)$ по отношению к EAL. Это так называемое *управление* стремится обеспечить точность частоты TAI.

График показан ниже (см. рис. 7.2).

- Период вычисления: один месяц. Дата t_0 находится посреди месяца m.
- Скорость сбора данных: – каждые 5 суток (в 0 часов UTC для данных, оканчивающихся на 4-й или 9-й день, если дни представлены в модифицированной юлианской дате (MJD). Раньше, до 1995 года, данные собирались каждые 10 суток.
- Последние данные по публикациям: t_3 – в конце месяца m–1.
- Почти не привлекаются другие данные (т.е. $t_2 - t_3 = 0$).

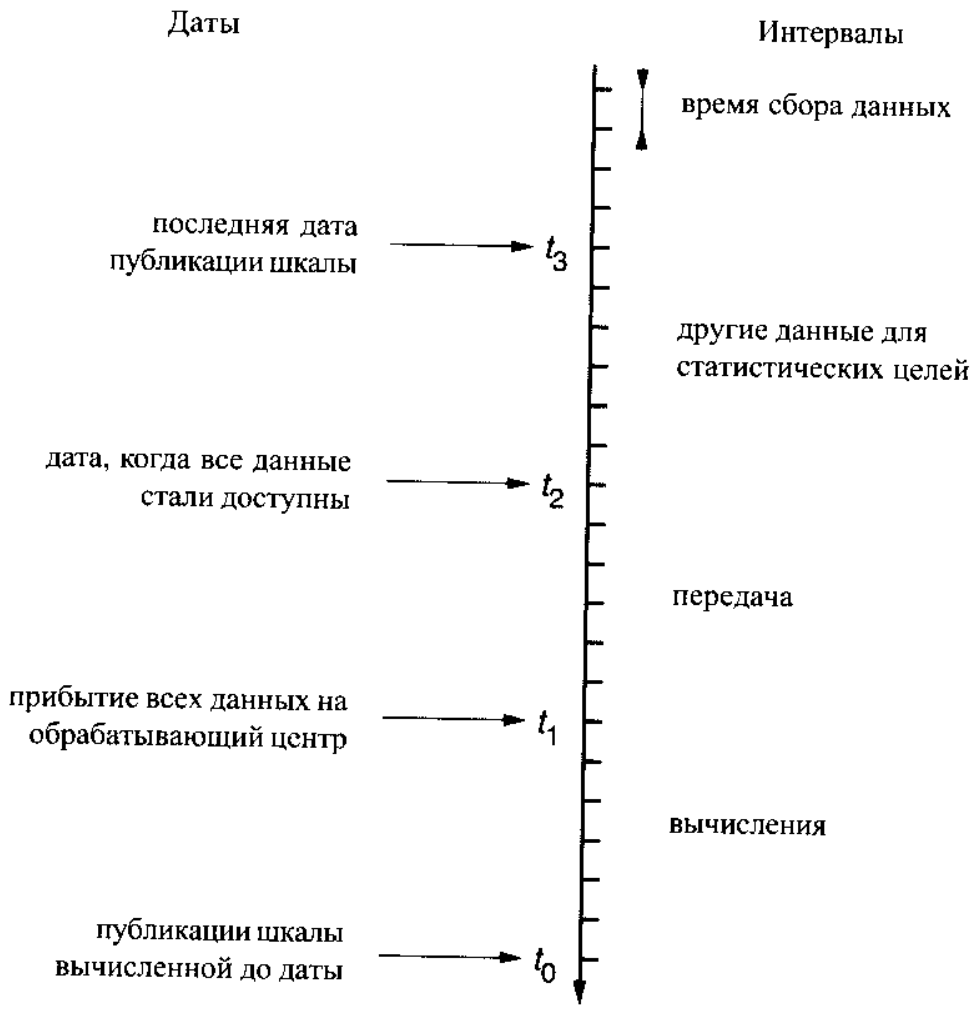


Рис. 7.2. График проведения алгоритмов шкалы времени. Данные подразумевают либо разность показаний часов, либо информацию, которую можно использовать для вычисления таких разностей, (например, данные приема сигналов GPS).

- Для передачи требуется несколько суток (включая подготовку данных в участвующих лабораториях). Это приводит к определенному уровню ограничений на часть различных лабораторий и на BIPM.

(b) Алгоритм стабильности

Начнем с рассмотрения воображаемой ситуации, когда все участвующие часы I (I = J, K и т.д.) считываются одновременно согласно соглашению о координатной одновременности, на дату t в геоцентрическом координатном времени (TCG, TT или TAI). Показания часов $\tau_I(t)$ преобразуются в $t_I(t)$ путем введения произвольных констант интегрирования. Обозначив $\{\}(t)$ взвешенное среднее по ансамблю часов на дату t, EAL можно определить через

$$[EAL](t) = \{t_I\}(t), \tag{7.8}$$

присваивая вес в соответствии с выбранными критериями стабильности.

Определение (7.8) может отвечать требуемым целям столько долго, сколько все часы могут работать, пока не изменятся их веса и не введутся новые часы. Этому можно подвести итог, заключающийся в том, что ансамбль часов должен оставаться неизменным. Если ансамбль каким-либо образом изменяется, то ясно, что это вызовет появление скачка времени, так как не все часы имеют одно и то же показание времени, и это приведет также к скачку частоты, так как не все часы имеют одну и ту же частоту. Для избежания такой ситуации определение (7.8) заменяется на

$$[EAL](t) = \{t_l\}(t) + A + B(t - t_0), \tag{7.9}$$

где t_0 какое-то произвольно выбранное начало, A и B константы, которые при изменении ансамбля часов изменяются так, чтобы поддержать непрерывность времени и частоты. Пусть t_m будет датой, когда были совершены некоторые изменения. Обозначим старые значения A и B через $A_$ и $B_$ и новые значения через A_+ и B_+. Требования непрерывности времени выражаются через

$$A_+ = A_- + \{t_l\}_-(t_m) - \{t_l\}_+(t_m) - (B_+ - B_-)(t_m - t_0), \tag{7.10}$$

ансамбль усредняет $\{t_l\}_-(t_m)$ и $\{t_l\}_+(t_m)$, которые вычисляются в момент времени t_m для старого и нового ансамбля соответственно. Рассмотрим случай, когда одни часы остановились. Тогда t_m является последней датой, для которой часы все еще давали данные. Поэтому оно включается в $\{t_l\}_-(t_m)$, но не в $\{t_l\}_+(t_m)$. До тех пор, пока вопрос касается непрерывности частоты, это приводит к

$$B_+ = B_- + \left\{\frac{dt_l}{dt}\right\}_- - \left\{\frac{dt_l}{dt}\right\}_+, \tag{7.11}$$

где производные вычисляются во время t_m.

В то же время это есть прямой и строгий способ вычисления A_+ если мы нашли B_+, но фактически из (7.11) невозможно получить B_+ по тем простым причинам, что мы не сможем оценить мгновенные значения частот. (Мы можем заменить начало, выбрав $t_0 = t_m$. Тогда B_S не нужно для вычисления A_+.) Мы должны работать со средними значениями частот. До даты t_m должна быть выбрана длительность, на которой вычисляются эти средние частоты. После даты t_m данные недоступны, если мы хотим создать шкалу в реальном или близком к реальному времени. Следовательно, мы должны *предсказать* частоты каждых часов на основе их предыдущего поведения.

Тонкая проблема возникает при рассмотрении (7.11), потому что в рассмотрение включаются производные по отношению к неизвестному идеальному времени t. Однако нам необходима только разность между

двумя такими производными. Это означает, что t можно заменить на физически реализуемую шкалу, при условии, что ее частота достаточно стабильна и лежит достаточно близко к частоте t. На практике EAL сама может выполнять эту задачу. Следовательно, используя прошлые разности частот каждых часов по отношению к EAL, мы пытаемся предсказать будущее значение этих разностей.

Предсказания частоты и взвешивание формируют центральную часть алгоритмов шкалы времени. Сюда же должна быть добавлена другая важная задача, а именно обнаружение аномального поведения каких-либо часов, которое ведет к необходимости исключения этих часов из ансамбля. Это серьезная статистическая проблема. Тип решения, приведенный в действие, в конечном счете влияет на качество производимой шкалы.

Теперь мы преобразуем наши уравнения так, чтобы они были применимы к реальным измерениям. Этими измерениями являются разности собственных времен между часами, полученные методами, описанными в главе 5. Напомним, что эти разности в показаниях относятся к одной и той же дате t, но так, что они изменяются слабо со временем, и поэтому они могут быть датированы опорной шкалой θ, для которой остается большая свобода выбора. На практике используется UTC. Начнем с разности собственного времени двух часов J и K и вычислим разность координатных времен t_J и t_K, выбирая постоянные интегрирования произвольным образом. Это дает

$$[t_J - t_K](\theta) = \delta_{JK}(\theta), \tag{7.12}$$

где δ_{JK} – измеренное значение. Положив

$$x_i(\theta) = [EAL - t_i](\theta) \tag{7.13}$$

уравнения (7.9) и (7.12) переходят в

$$\{x_i\}(\theta) = A + B(\theta - \theta_0), \tag{7.14}$$

$$x_J(\theta) - x_K(\theta) = -\delta_{JK}(\theta), \tag{7.15}$$

где коэффициент t–t_0 заменяется на $\theta - \theta_0$, так как B исключительно мало (порядка 10^{-13}). Сеть сличения времени организована так, чтобы система формировалась по (7.14) и уравнения типа (7.15), написанные для пары часов, могли бы быть решены точно, потому что обработка избыточных связей приносит с собой больше проблем, чем преимуществ. Если ансамбль часов изменяется, то A и B переустанавливаются через

$$A_+ = A_- - \{x_i\}_-(\theta_m) + \{x_i\}_+(\theta_m) - (B_+ - B_-)(\theta_m - \theta_0) \tag{7.16}$$

и

$$B_+ = B_- - \left\{ \frac{dx_I}{dt} \right\}_- + \left\{ \frac{dx_I}{dt} \right\}_+, \tag{7.17}$$

где, как и прежде, t заменяется производной по EAL.

Полученное решение для x_I преобразуется к EAL – τ_I с помощью теоретических соотношений между τ_I и t_I при тех же значениях постоянных интегрирования, что и для обратного преобразования. Тогда EAL становится доступной в месте расположения участвующих часов через поправки к их показаниям.

При выполнении вычислений мы используем тот факт, что все часы, участвующие в конструировании EAL, установлены неподвижно на Земле. Для перехода от их собственного времени к координатному времени необходимо учитывать только постоянный сдвиг частоты. Этот сдвиг используется не совсем точно. Он содержится в поправочном члене B. Из чего следует, что алгоритм стабильности производит непосредственно EAL – τ_I. Однако если мы хотим объединить спутниковые часы, то необходимо их полное описание. Прежде упомянутое упрощение, в котором θ_0 сделано совпадающим с θ_m приводит к изменению θ_0 каждый раз, когда изменяется ансамбль.

(c) Управление частотой

Только что описанный алгоритм стабильности оптимизирует стабильность на периоде усреднения, который был установлен путем выбора способа предсказания частот. В своем применении к EAL мы увидим, что оптимальное значение находится где-то между 30-м и 60-м днями. Однако алгоритм не гарантирует нас от долговременных флуктуаций. Пошаговая настройка B_S может даже привести к неограниченному дрейфу частоты.

Первая проблема, которую надо решить, состоит в том, чтобы оценить частоту EAL используя данные первичных стандартов. В этом обсуждении все частоты приводятся к уровню вращающегося геоида.

Первичные стандарты обеспечивают оценку среднего значения частоты EAL между датами $\theta_i - \theta_f$. С течением времени, когда калибровка становится частью прошлого, EAL остается памятью этой калибровки благодаря стабильности своей частоты. Тем не менее эта память не идеальная, и поэтому должен быть добавлен вклад от нестабильности частоты EAL к внутренне присущим нестабильностям стандарта.

Для этого был разработан *фильтр,* дающий оптимальное значение для средней частоты EAL на данном интервале путем присвоения коэффициентов (с суммой 1) по всем возможным калибровкам независимо от того, когда они были, до того или после. Слово «фильтр» предпочтительнее, чем вес, потому что некоторые коэффициенты могут быть отрицательным. Фильтр учитывает следующие факторы:

- статистическую модель для частотного шума в EAL, выражающуюся в сумме белого фазового шума (вклад сличений времени), белого и фликкерного частотного шума, а также шума из-за случайного блуждания частоты (вклад часов);
- случайную и систематическую неопределенности в калибровке частоты;
- начальную и конечную даты калибровки и период оценки.

Теория по этим вопросам достаточно сложна и здесь обсуждаться не будет. Обсуждение можно найти в [7.4].

Управление шкалой состоит в нахождении функции $P(\theta)$, позволяющей нам вывести TAI из EAL через

$$TAI(\theta) = EAL(\theta) + P(\theta), \tag{7.18}$$

таким образом, что частота TAI остается близкой к частоте точного фильтра, и на ней не сказывается нестабильность фильтра по отношению к EAL. Например, если частота TAI изменяется сразу же как только происходит новая калибровка, то скачки частоты будут неприемлемы. Чтобы избежать такой ситуации нестабильность частоты $P(\theta)$ устанавливается так, чтобы она была меньше, чем оцениваемая нестабильность частоты EAL.

(d) Применение

Здесь представлено в высшей степени схематическое описание применения алгоритма, так как за это время был достигнут значительный прогресс и в разработке атомных часов и в сличениях времени, и поэтому алгоритм BIPM должен иметь определенный уровень гибкости в своем применении. Наше обсуждение относится к ситуации конца 2000 года. Однако начальная оценка роли, которую играет TAI, была выполнена в BIH еще в 1973 году и регулярно подтверждалась международными организациями. По этой причине основные принципы алгоритма с тех пор не изменялись.

Когда проводятся вычисления, то целое число секунд опускается. Они вводятся опять после получения результатов. Таким образом, мы можем рассматривать следующие уравнения одинаково применимыми как к TAI так и к UTC.

Сеть сличения времени. Обозначим через k лабораторию, дающую вклад в TAI. Каждая лаборатория k имеет *ведущие часы,* которые обеспечивают приближение UTC(k) к всемирному координированному времени UTC. Эти часы служат в качестве опоры для всех процедур сбора местных данных и их обработки. В частности измеряется разность времени $UTC(k) - \tau_C$ для всех часов C в лаборатории k. BIPM использует только значения, измеряемые каждые 5 суток в *стандартные даты.* Они определяются

модифицированной Юлианской датой, оканчивающейся на 4-е и 9-е сутки в 0 часов UTC.

Разнесенные сличения времени относятся к UTC(k). Почти непрерывно BIPM оценивает стандартные данные путем усреднения или фильтрования. Рис. 7.3 показывает сеть линий связи по передаче времени. Несколько лабораторий действуют как центры для линий связи до 1000 км путем применения GPS и GLONASS. Эти центры связаны вместе с помощью лучшей доступной техники, а именно TWSTFT и GPS с использованием в последнего случае точных эфемерид, вычисляемых ретроспективно, и измеренных значений ионосферных поправок. В некоторых странах существует определенная иерархия, в которой центральная лаборатория обеспечивает связь с рядом национальных лабораторий. Эта иерархия, например, долго сохранялась во Франции.

Алгоритм стабильности. Алгоритм стабильности, называемый Алгос, вычисляет значения *EAL-UTC*(k) в стандартные даты за один пакет. (Продолжительность одного пакета обычно 30 суток, а иногда бывает 25 или 35 суток). Совершенно произвольно эти даты начала и окончания пакета были выбраны как ансамбль модифицированных дат, т.е. ранее определенных θ_m. Следовательно, вес часов удерживался неизменным на каждом месячном интервале. Вес часов равен нулю, если часы дают незавершенные данные, или они остановились или были введены в ансамбль после начала интервала.

В настоящее время (2000 г.), Алгос обрабатывает данные более чем 200 часов, главным образом промышленных цезиевых часов, имеющих отличную долговременную стабильность. Так как здесь уместна только стабильность, то вклад могут давать и другие часы. Сейчас в формировании TAI участвуют около сорока водородных мазеров и их число остается стабильным.

Ансамбль усредняется через взвешенные значения. Вес часов C пропорционален обратной величине вариации $\sigma_C^2(12,\tau)$ из 12 средних значений частоты, взятых за каждый месяц по отношению к EAL[3]. То, что была выбрана эта вариация, а не обычная двухвыборочная вариация $\sigma_{y,C}^2(\tau)$, сделано для того, чтобы уменьшить вес часов, имеющих дрейф частоты или, как это иногда отмечалось, годовые флуктуации частоты. В модели частотных фликкерных шумов, которые подходят для рассматриваемых значений τ

$$\sigma_C^2(12,\tau) = 2\sigma_{y,C}^2(\tau). \tag{7.19}$$

[3] В долговременных вариациях τ представляет длительность, за которую вычисляется средняя частота, т.е. период усреднения, как это обсуждено в разделе 5.2.3.

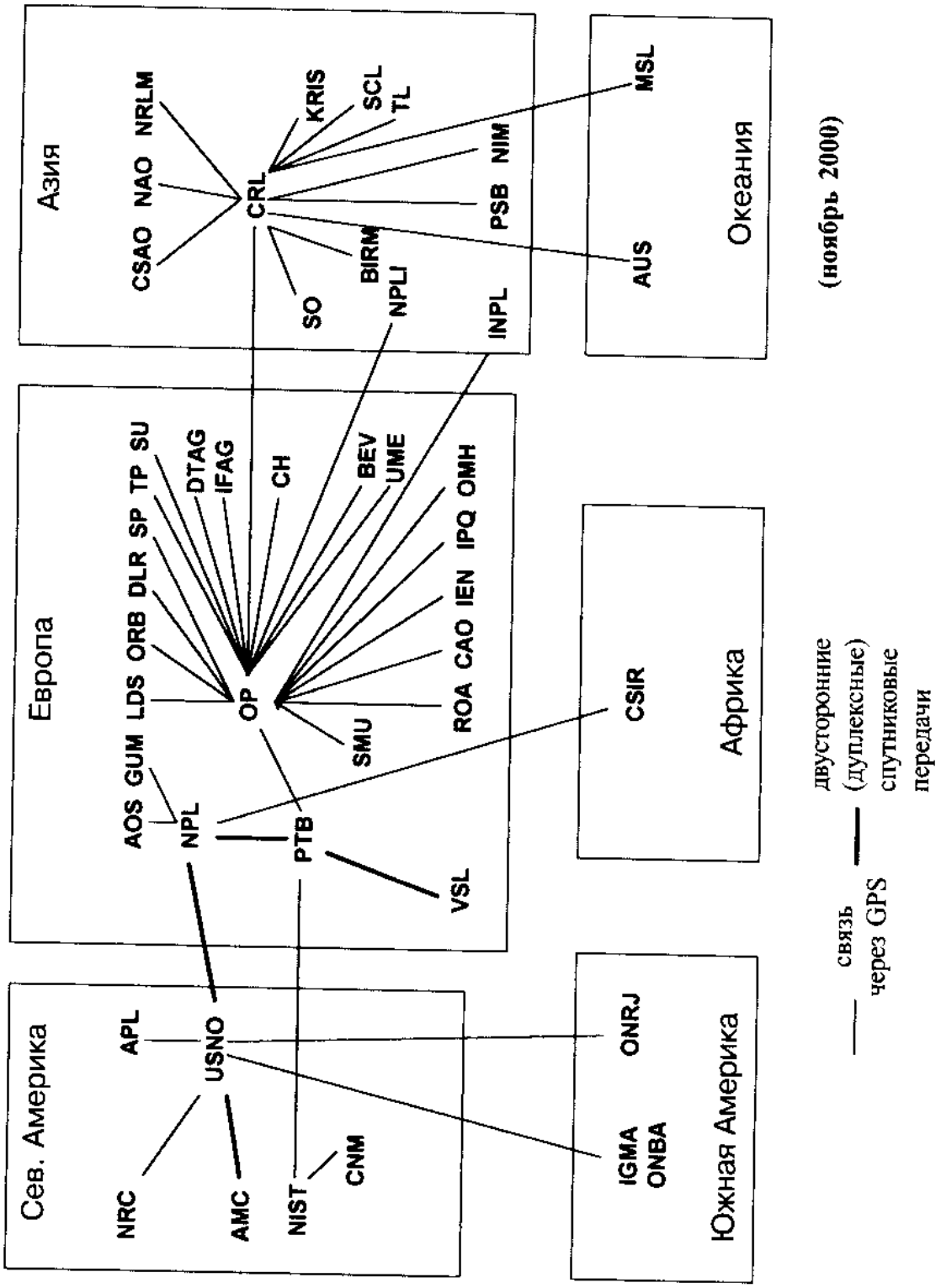


Рис. 7.3. Линии передачи времени для установления TAI (на ноябрь 2000 года). Акронимы даны в приложении 1.

Когда флуктуации частоты часов оцениваются по отношению к EAL, в которую они и дают вклад, то некоторые часы последовательно усиливают свой статистический вес и обычно доминируют в ансамбле. Эту проблему можно решить путем применения верхнего предела взвешивания. Этот предел изменяется несколько раз для того, чтобы соответствовать улучшенным часам. С января 1998 года правило таково, что никакие

часы не могут получить вес, превышающий 0,7% от общего веса. Однако, вместе с возрастающей ролью, которую играют новейшие промышленные часы, это правило становится неудовлетворительным, потому что оно не достаточно сильно осуществляет отбор среди этих часов. Относительный верхний предел L в виде L/N в настоящее время находится в стадии изучении (ноябрь 2000), где N – общее число часов и L может быть между 2 и 3. Были также предприняты теоретические исследования для определения оптимального верхнего предела.

Для того чтобы обнаружить аномальное поведение часов делаются сличения их средней частоты по отношению к EAL за последний месяц, т.е. за месяц, для которого вычисляется TAI, а также их среднее значение частоты по отношению к EAL за последние 11 месяцев. Если вариации отличаются больше чем в три раза по отношению к вариациям, ожидаемым от случайных флуктуаций, то вес этих часов устанавливается равным нулю. Все часы проходят этот тест. Вычисления повторяются каждый раз, когда исключаются какие-либо часы. Применяются четыре итерации.

Частоты предсказываются из простого продолжения наблюдения частоты по отношению к EAL за предыдущий месяц. Многие попытки исследования реальных данных для получения лучших предсказаний другим способом не привели к успеху.

Любые новые часы остаются под наблюдением по крайней мере четыре месяца прежде чем эффективно начинают участвовать в TAI. За этот период они включаются в расчеты с нулевым весом. Если число частотных выборок n меньше 12, то $\sigma_C^2(12,\tau)$ оценивается по $\sigma_C^2(n,\tau)$ в предположении фликерного частотного шума.

С алгоритмом Алгос возможно применение различного взвешивания и правил прогнозирования в соответствии с типом часов. Например, было бы резонно принять во внимание дрейф частоты в водородных мазерах. До сих пор эта возможность не была внедрена.

Как пример рассмотрим лаборатории и часы, принимающие участие в формировании шкалы за период с 28 августа по 28 сентября 2000 года. BIPM получил данные от 221 часов (179 цезиевых часов и 42 водородных мазеров) из 39 лабораторий и национальных организаций, которые могут включать в себя несколько лабораторий. 121 часам был приписан максимальный вес, включая 98 последних промышленных часов, 4 первичных цезиевых часов, построенных в исследовательских центрах и 17 водородных мазеров. Нулевые веса были приписаны 44 главным образом новым часам в их испытательные периоды. Наконец 56 остальных часов получили промежуточные веса, обычно очень небольшие.

Часы, которым были приписаны максимальные веса, дали 85% вклада и, как упоминалось ранее, это очень большая пропорция.

Управление частотой. В 1969 году BIH, опираясь на навигационную систему Лоран-С, имело возможность применять не метод средних частот от каждых часов при вычислении TAI (известной тогда как TA(BIH)), а заменить его общим усреднением по времени. Однако при этом терялся доступ к индивидуальным часам. TAI было средним от атомных шкал времени, устанавливаемых небольшим числом развитых национальных центров времени. Она вычислялась по алгоритму, похожему на Алгос [7.5]. Первоначально поправки B в определяющем уравнении типа (7.9) были равны нулю для трех участвующих центров времени, а именно: военно-морской обсерватории (USNO) США, физико-технического института (PTB), Германия, и национальной комиссии по времени, Франция, в Парижской обсерватории. Не было попыток ни соединить частоты с предыдущими значениями TAI, ни настроить их с данными от первичных стандартов, так как эти данные были слишком разрозненные и слишком неопределенные.

В 1973 году, когда был организован прямой доступ к индивидуальным данным часов, BIH внедрило алгоритм Алгос. Были приняты предосторожности, чтобы с помощью должным образом настроенного первоначального значения B исключить любые скачки частоты. TAI производилась непосредственно Алгосом, иначе говоря, она как раз равнялась EAL. Однако первичные стандарты совершенствовались благодаря большому техническому прогрессу и сразу же после начала генерации TAI было показано, что ее частота выше нормы на 10×10^{-13}.

Эта ошибка частоты была скорректирована 1 января 1977 года путем скачка частоты TAI равного в точности 10×10^{-13} , согласно резолюции, принятой MAC в 1976 году.

После этого стало возможным ввести управление поведением частоты, как это уже предлагалось в 1974 году консультативным комитетом по определению секунды (CCDS). Предложение было отложено потому, что частотная ошибка, которую было необходимо компенсировать, была слишком велика. Поэтому сначала функция $P(\theta)$ была последовательностью линейных функций со скачками частоты 2×10^{-14}, происходящими когда это было необходимо на интервалах более двух месяцев. Величина нарушения непрерывности была уменьшена в 1996 году до нескольких единиц, умноженных на 10^{-15}, а скачки могли быть выполнены на месячных интервалах. Рис. 7.4. показывает двухмесячные средние значения длительности единицы шкалы TAI в секундах на вращающемся геоиде. Отстройка около 2×10^{-14} перед 1996 годом произошла из-за отсутствия поправок на излучение черного тела (см. раздел 6.4.1e). Решение, внести эту поправку ко всем первичным стандартам частоты, было принято в мае 1996 года. Отстройка частоты постепенно уменьшалась путем управления частотой.

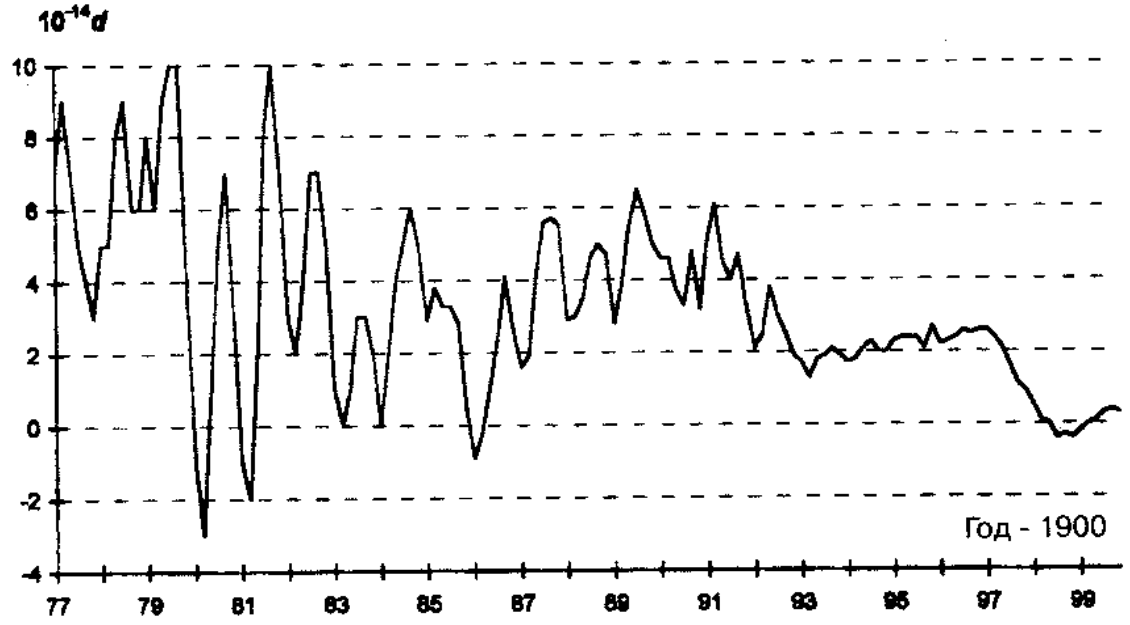


Рис. 7.4. Длительность (1+*d*) секунд единицы шкалы TAI на вращающемся геоиде (двухмесячные значения)

В то время как Алгос оптимизирует стабильность за один или два месяца, управление поведением частоты гарантирует стабильность на много больших периодах. Оно может рассматриваться как способ нового взвешивания. Если возрастет период выборки для всех часов, дающих вклад в EAL, то каждые часы влияют на TAI все меньше и меньше, в то время как первичные стандарты становятся все более и более значимыми. Для стабильности на периодах более 5 лет все взвешивания выполняются первичными стандартами.

Форма результатов. После применения слежения за частотой TAI становится известной из значений *TAI – UTC*(*k*) за стандартные даты. Фактически BIPM каждый месяц публикует значения *UTC – UTC*(*k*) в циркуляре T, доступном с помощью электронной почты. В качестве примера представлен рис. 7.5. Кроме того, путем слежения за GPS и GLONASS спутниками для сличения времени, значения *UTC – GPS Времени* и *UTC – GLONASS Времени* обеспечиваются с неопределенностью, близкой к неопределенности *UTC – UTC*(*k*). Центры времени *k* и GPS и GLONASS спутники образуют сеть точек первичного доступа к TAI и UTC.

Важно отметить, что TAI и UTC становятся определяемыми после публикации. Иногда допускаются небольшие поправки (достигающие нескольких наносекунд), но никогда не допускается общий пересмотр. Сейчас даже эти небольшие поправки стараются исключить. Отсюда следует, что если ошибки из какого бы то ни было источника не были обнаружены, то они навсегда включаются в TAI. Выявление и исключение этих ошибок является ключевой задачей Алгоса и программ пред-

варительной обработки, на которую выделяется только несколько суток.

1 – Всемирное координированное время UTC. Вычисленные значения UTC – UTC(k). (с 1 января 1999 в 0 часов UTC, TAI – UTC =32 с)

Дата 2000 0 часов UTC		сентябрь 27	октябрь 2	октябрь 7	октябрь 12
MJD		51814	51819	51824	51829
Лаборатория k		UTC-UTC(k) (единица равна одной наносекунде)			
AOS	(Borowiec)	-2757	-2793	-2736	-2659
APL	(Laurel)	6676	6716	6740	6771
AUS	(Sydney)	284	271	249	253
BEV	(Wien)	-383	-372	-371	-371
BIRM	(Beijing)	222	213	200	194
CAO	(Cagliari)	-3306	-3284	-3292	-3290
CH	(Bern)	-95	-115	-139	-142
CNM	(Queretaro)	-157	-161	-163	-171
CRL	(Tokyo)	-44	-50	-53	-50
CSAO	(Lintong)	-23	-26	-8	2
CSIR	(Pretoria)	-526	-591	-646	-704
DLR	(Oberpfaffenhofen)	-13618	-13790	-13974	-14149
DTAG	(Darmstadt)	-695	-701	-681	-681
GUM	(Warszawa)	-759	-767	-770	-765
IEN	(Torino)	59	55	58	55
IFAG	(Wettzell)	-3186	-3213	-3236	-3247
IGMA	(Buenos Aires)	119	112	108	104
INPL	(Jerusalem) (1)	-81	-88	-97	-105
IPQ	(Monte de Caparica)	4734	4787	4842	4891
JATC	(Lintong)	-4130	-4171	-4211	-4257
KRIS	(Taejon)	13	13	2	26
LDS	(Leeds)	-	-	-	-
MSL	(Lower Hutt)	10612	10694	10839	10949
NAO	(Mizusawa)	-2316	-2292	-2269	-2229
NIM	(Beijing)	-2709	-2717	-2722	-2721
NIST	(Boulder)	-7	-8	-8	-13
NPL	(Teddington)	1	2	4	3
NPLI	(New-Delhi)	7953	7981	-	-
NRC	(Ottawa)	17	20	24	25
NRLM	(Tsukuba)	3514	3546	3563	3597
OMH	(Budapest)	4993	5044	5054	5051
ONBA	(Buenos Aires)	17	-47	-52	-55
ONRJ	(Rio de Janeiro)	2904	2912	2922	2934
OP	(Paris)	-39	-35	-39	-38
ORB	(Bruxelles)	31	37	41	36
PSB	(Singapore)	1999	2023	2028	2062
PTB	(Braunschweig)	11	10	18	20
ROA	(San Fernando)	-35	-38	-41	-43
SCL	(Hong Kong)	-11	-3	-13	-3
SMU	(Bratislava)	-3655	-3665	-3679	-3690
SO	(Shanghai)	-	-	-	-
SP	(Boras)	-442	-423	-412	-395
SU	(Moskva)	84	88	89	90
TL	(Chung-Li)	-4	-47	-57	-68
TP	(Praha)	98	107	114	118
UME	(Gebze-Kocaeli)	-600	-712	-203	-243
USNO	(Washington DC) (MC)	-3	-3	-2	-5
VSL	(Delft)	8	14	11	11

Рис. 7.5. Первая страница циркуляра T Международного бюро мер и весов (BIPM).

7.3. Свойства TAI и UTC

7.3.1. Надежность

Надежность гарантируется международной основой TAI. Представим на момент, что бы случилось, если бы кто-то нашел способ (практически это только шутка) остановить все атомные часы хотя бы на короткое время. Это бы вызвало такой огромный беспорядок в мировых делах, что всем было бы не до TAI. Более того, если бы его пришлось устанавливать опять, то фаза TAI могла бы быть восстановлена в пределах нескольких десятых микросекунды из наблюдений быстро вращающихся пульсаров (см. главу 8).

7.3.2. Точность показаний

Неопределенность показаний TAI, использующих значения $UTC - UTC(k)$, является одной из неопределенностей первичных доступов лабораторий k и приблизительно равна неопределенности, вводимой этой лабораторией в сличения времени, используемые для формирования TAI. Она представляет наименьшую возможную неопределенность, если датировать события по TAI в рассматриваемой лаборатории. В лабораториях, оборудованных лучшими приборами, эта неопределенность оценивалась (в 2000 году) в десяток наносекунд с учетом неопределенностей задержек в приборах.

Доступ, разрешенный к TAI путем приема сигналов GPS и использования значений *UTC – GPS Времени,* приводит к неопределенностям того же порядка в лабораториях, обладающих наилучшем оборудованием, и при усреднении большого массива данных сличений, по крайней мере за одни сутки. То же самое можно сказать и про GLONASS.

7.3.3. Стабильность частоты

Значения ординат на рис. 7.4 представляют после изменения знака разность нормализованных частот между TAI и группой первичных стандартов (на геоиде). Можно видеть годовые флуктуации. Их общая амплитуда была близка к 10^{-13} на период с 1977 по 1982 годы, а потом постепенно снизилась до 10^{-14} в 1993-1995 годах, прежде чем как-бы исчезнуть насовсем. Из-за чего это происходит, из-за TAI или из-за первичных стандартов? Если амплитуда большая, то кажется, что флуктуации происходят, главным образом, из-за влияния окружающих условий на промышленные цезиевые часы с доминирующим фактором – влажностью. Это слабое место было постепенно исключено или улучшением часов или их лучшей защитой. Крошечное значение флуктуаций остававшихся в 1995 году могло все еще быть вызванным влиянием окружающей

среды во время выполнения сличений. Уже упоминалась температурная чувствительность некоторых GPS приемников. Здесь может так же сказываться неадекватное моделирование рефракции, в частности, тропосферной рефракции, которая никогда не измеряется. Похоже на то, что первичные стандарты исключаются из обнаруживаемых годовых вариаций.

Стабильность TAI может быть оценена из сличений с другими атомными шкалами, поддерживаемыми рядом метрологических центров времени. Нестабильности на рис. 7.6 была вычислена для времени выборки τ в диапазоне от 10 до 160 суток. Фактически это нестабильность EAL в 1999-2000 годах [7.6], но для этих значений τ она незначительно отличается от нестабильности TAI. Для больших времен τ нестабильность оценивается в предположении, что текущие свойства часов и первичных стандартов остаются теми же самыми бесконечно долго. Она выходят на уровень около 5×10^{-15} в результате управления поведением частоты. Точность частоты налагает верхний предел на сверхдолговременную стабильность.

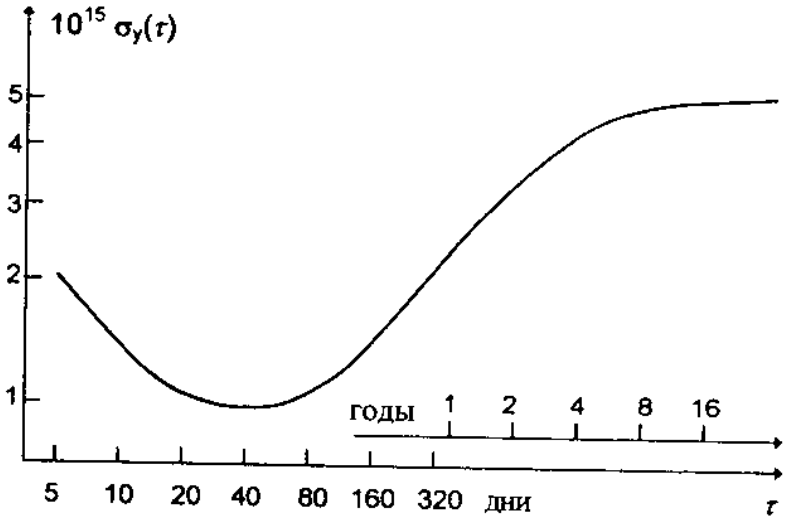


Рис. 7.6. Нестабильность частоты TAI. (см. текст для объяснения).

7.3.4. Точность частоты

Пусть d_{TAI} будет относительным отклонением единицы шкалы TAI от единицы шкалы TT (секунды СИ на вращающемся геоиде), т.е. относительная девиация частоты TAI с противоположным знаком ($d_{TAI} = -y_{TAI}$). Среднемесячное значение d_{TAI} обычно меньше $\pm\ 10^{-14}$. Его величину BIPM рассчитывает ретроспективно вместе с его стандартной неопределенностью u. Последняя должна оцениваться с величайшей осторожностью, принимая во внимание неопределенности в первичных стандартах частоты и в их линиях связи с TAI. В конце 2000 года значение u лежало в диапазоне $(2–3)\times10^{-15}$. Например, на интервале от 28 августа до 27 сентября 2000 года, $d_{TAI} = 7{,}3 \times 10^{-15}$ и $u = 2{,}2 \times 10^{-15}$.

7.4. Мировая организация измерений времени

Всемирная организация измерения времени постепенно совершенствовалась по мере прогресса в атомных часах, стандартах и шкалах времени. Однажды эта организация достигла определенной зрелости, и это было признано и установлено соответствующими международными организациями. Глава 4 дает обзор эволюционных фаз. В этой главе мы опишем настоящее положение организаций, которые должны бы появиться чтобы представлять будущее, по крайней мере, на следующие несколько лет или даже на десятилетие и более.

7.4.1. Всемирное координированное время

Как объяснялось в разделе 4.5, всемирное координированное время UTC в его настоящей форме есть ни что иное, как международное атомное время, скорректированное на целое число секунд для того, чтобы приблизительно соответствовать всемирному времени UT1, связанному с вращением Земли.

С 1 января 1972 года UTC реализовывалось согласно инструкциям в рекомендациях ITU-R TF 460-4 Международного союза электросвязи (ITU). Эти рекомендации имеют корни в рекомендации 460, принятой в 1970 году международным консультативным комитетом радио (CCIR), комитетом в ITU, который после этого исчез. Давайте теперь обрисуем ряд директив и объясним, как они внедрены в практику.

(a) Скачущие секунды

Скачек на целую секунду, ITU определяет как «введение одной положительной или отрицательной скачущей секунды», которая должна быть введена в конце UTC месяца, предпочтительнее в конце декабря или в конце июня, в ином случае в конце марта или в конце сентября. Оговорено, что положительная скачущая секунда должна начинаться в 23 часа 59 минут 60 секунд и оканчиваться в 0 часов 0 минут следующего дня. В таком случае нет неоднозначности, если события датируются таким образом. Однако двузначность появляется в других системах, например, когда используются доли дня. Если была введена положительная скачущая секунда, то дата N 000 005 79 может означать или сутки ($N-1$) в 23 часа 59 минут 60,5 секунд или сутки N в 0 часов, 0 минут 0,5 секунд. Эта неоднозначность не возникает в случае отрицательной скачущей секунды. Однако вероятность того, что появится необходимость в отрицательной скачущей секунде, практически равна нулю.

Даты для скачущей секунды устанавливаются международной службой вращения Земли IERS и объявляются, по крайней мере, за 8 недель до ее введения.

Система UTC работает хорошо. На 1 января 1972 года *TAI – UTC* была равна 10 с. Положительные скачущие секунды всегда вводились в конце декабря или в конце июня со скоростью от 1 секунды в год и до 1 секунды за 2,5 года. С 1 января 1999 года и до даты, которая еще не определена, *TAI – UTC* = 32 с. Соотношение между TAI и UTC можно найти в ежегодном отчете IERS и секции времени BIPM.

(b) Сигналы времени. Коды для всемирного времени UT1

Излучения сигналов времени соответствуют как можно ближе к UTC. Рекомендации ITU устанавливают допуск в 1 мс. На практике расхождение много меньше чем 1 мс. Рекомендуется так же, чтобы несущая частота была настроена по частоте TAI с относительной отстройкой частоты меньше 10^{-10} и чтобы соотношения между сигналами времени и фазой несущей были известны (т.е. если несущая частота 10 МГц, то следующий импульс должен появляться через каждые 10^7 периодов).

Рекомендуется, чтобы сигналы времени несли звуковую кодовую информацию о разности *UTI – UTC*. Эта информация обозначается величиной *DUTI* и представляется в форме целых кратных от 0,1 с, значение которых появляется путем маркирования определенного числа секундных импульсов либо их уширением, либо расщеплением, либо другим методом. На рис. 7.7 показано, как работает код. Значения *DUTI* и даты их применения обеспечиваются IERS за месяц вперед до их применения и они одинаковы для всех излучений.

Некоторые сигналы времени несут также закодированную информацию *dUTI*, дающую значения *UTI – UTC* до ближайшей 0,02 с, но это неофициально. Даже кодированная информация *DUTI* используется очень редко и некоторые радиостанции ее больше не передают.



Рис. 7.7. Код для *DUTI = UTI – UTC*, передаваемый вместе с сигналами времени. Секундный импульс подчеркивается уширением, расщеплением или другим способом. Для отрицательных значений *DUTI* соответствующее число импульсов отмечается от 9-й секунды.

Ежегодный отчет секции времени BIPM предоставляет также список международных сигналов времени и их основные характеристики.

(c) Применение UTC

Всемирное координированное время используется как практическая опора по всему миру, его использование, как и применение акронима UTC, одинакового на всех языках, рекомендовано международными организациями, включая с 1971 года генеральную конференцию по мерам и весам (CGPM) и ITU.

В разных странах по всему миру местное время соединено с UTC через целые числа часов. Иногда эта связь устанавливается законом, как во Франции (декрет от 9 августа 1978 года). Официальные организации иногда предпочитают использовать национальное приближение к UTC, как это имеет место в Германии. Существуют страны, в которых UTC официально не признано, хотя оно используется фактически, так как отсутствуют другие доступные шкалы времени.

Давайте только напомним, что все еще широко распространенное использование акронима GMT некорректно, если мы намерены ссылаться на UTC, а также и в случае выражения времени для всеобщего применения. Эта практика часто критиковалась особенно международным астрономическим союзом IAU. Строго говоря GMT должно быть заменено на UTC. Однако для повседневного применения, где неточности достигают более 1 секунды и не имеют реального значения, UT может быть приемлемым названием.

7.4.2. Локальное представление UTC и независимые локальные атомные времена

Как упоминалось в разделе 7.2.2 приближение UTC(k) к UTC поддерживается в центре времени k. Это обычно реализуется атомными часами, оборудованными приборами, которые добавляют линейные поправки, осуществляемые подстройкой только двух параметров. В других случаях эти физические приборы заменяются вычислениями, допускающими в принципе более сложные корректировки, например с учетом квадратичных членов. В обоих случаях все усилия обычно направляются на то, чтобы удержать UTC(k) как можно ближе к UTC. В 1993 году консультативный комитет по определению секунды рекомендовал, чтобы лаборатории, хранящие время, представляли информацию, необходимую для облегчения синхронизации с UTC в реальном времени по возможности в

пределах 100 нс. Один из способов как этого достичь – это обеспечить $|UTC - UTC(k)| < 100$ нс.

Ежемесячный цикл BIPM для выработки и распространения $UTC - UTC(k)$ требует от лабораторий одно или двухмесячной экстраполяции для того, чтобы подстроить UTC(k) к UTC. Чтобы этого добиться лаборатория k может основывать свои вычисления на экстраполяции тех своих часов, которые имеют лучшую стабильность при времени выборки около одного месяца. Однако если лаборатория имеет несколько атомных часов, то лучше сконструировать на локальном уровне шкалу времени, известную как *независимое локальное атомное время*. Такая шкала времени, обозначенная как TA(k), может быть установлена более быстро, чем TAI/UTC, потому что локальные сличения времени чрезвычайно точны. Это обеспечивает более серьезные основания для экстраполяции UTC(k).

В 2000 году около пятнадцати лабораторий производили TA(k) различными методами. Рассмотрим различные примеры. Так в военно-морской обсерватории США шкала TA(USNO) основана на десятках промышленных часов и на водородных мазерах. В Германии TA(PTB) непосредственно производится одним из первичных стандартов PTB. Во Франции TA(F) вычисляется в лаборатории первичного времени и частоты в парижской обсерватории, используя данные около двадцати промышленных часов, распространенных по стране. BIPM публикует значения $TAI - TA(k)$ так же, как и значения $UTC - UTC(k)$ в своем циркуляре T.

TA(k) не независимо от TAI (или UTC), так как одни и те же часы участвуют в обоих предприятиях. Но они независимы одни от других и это делает возможным вывести их нестабильность из взаимной нестабильности каждой пары шкал времени (используя так называемый *метод треугольки*, упомянутый в разделе 5.2.8 или в действительности *метод n-уголки*).

В 1995 году только две лаборатории в Соединенных Штатах Америки поддерживали свои UTC(k) в пределах 100 нс с UTC, а именно национальный институт стандартов и технологии (NIST) и уже упоминавшаяся ранее USNO. В 1996 году в список были добавлены национальная физическая лаборатория (NPL) в Великобритании и парижская обсерватория (OP). На 27 сентября 2000 года из 46 $UTC - UTC(k)$ 20 лабораторий были опубликованы BIPM с абсолютными значениями $UTC - UTC(k)$ меньше, чем 100 нс. В таблице 7.1 с января по сентябрь 2000 года показано качество синхронизации для четырех цитированных выше лабораторий, и для GPS.

Таблица 7.1. Синхронизация между *UTC* и *UTC(k)* для четырех лабораторий времени и для GPS за период с января по сентябрь 2000 года

Лаборатория или система	Предельные значения *UTC – UTC(k)* (нс)	Среднее значения *UTC – UTC(k)* (нс)	Среднее значения *UTC – UTC(k)* (нс)
GPS	–39	–7	13
NIST	28	15	17
NPL	80	20	29
OP	–49	–31	31
USNO	–20	2	9

7.4.3. Другие формы атомного времени

TAI является международной опорой по соглашению. Однако ничто не мешает другим шкалам времени быть установленными так, чтобы использовать данные со всего мира, если в этом есть преимущества для каких-либо специальных проектов. Можно подумать что могли бы быть реализованы более стабильные и более точные шкалы времени, чем TAI, если бы были возможны более длительные публикации и устанавливались бы по соглашению поправки на основе самой последней полученной информации. Например эффект излучения черного тела на частоту первичных стандартов был принят в расчете в однородной форме для TAI только в сентябре 1995 года, и не было сделано никаких ретроспективных поправок для улучшения прошлых значений TAI.

В главе 8 мы увидим, что некоторые недостатки TAI очень критичны при изучении пульсаров, в котором используется измерение времен милисекундных пульсаров. В осознании необходимости в улучшенной шкале времени для такого изучения BIPM установило датированную шкалу времени, называемую TT(BIPMxx), где xx представляет две последних цифры года, в котором она была произведена. Хотя эта шкала основана на EAL, данные от первичных стандартов используются в ней не таким же способом как для TAI. Ежегодные флуктуации, упомянутые раньше, исключаются [7.7]. Последовательные версии TT(BIPMxx) не являются прямым продолжением, так как они могут отличаться по всей длине шкалы времени от их начальных дат на 1 января 1977 года, которое одно и то же для всех шкал. TT(BIPMxx) представляется в форме разности с TAI на стандартные даты (каждые десять суток, каждые пять суток) по обращению к BIPM. Равное нулю на 1 января 1977 года расхождение между TT(BIPM96) и TAI + 32,184 секунд, может достигать 25 мкс.

7.5. Распространение времени и частоты

7.5.1. Частота

(a) Излучение эталонных частот

Несущие частоты излучений специализированных сигналов времени и стандартных частот основываются на частоте TAI со много большей точностью, чем это было предложено в рекомендациях ITU. Наибольшие объявленные неопределенности около 10^{-11}, хотя в некоторых случаях они снижаются до 10^{-12}. BIPM предоставляет список этих излучений в ежегодном отчете секции времени.

Некоторые другие радиосигналы, используемые для различных целей, имеют очень точные значения частот. Лоран-С излучает на частоте 100 кГц в пределах $\pm 10^{-12}$. Некоторые передающие станции также излучают очень точные частоты, среди них Allouis во Франции на частоте 162 кГц с неопределенностью в пределах $\pm 2 \times 10^{-12}$.

Эти сигналы удовлетворяют большинству требований к опорным частотам. Они легко доступны с использованием самых обычных приемников.

(b) Доступ к первичным стандартам частоты

Для фундаментальной метрологии, предмета этой книги, часто полезно иметь доступ к частоте удаленного первичного стандарта частоты E. Когда необходима максимально возможная точность, используются сличения времени, как это объясняется в разделе 5.6.4a. Если A – локальный генератор, $y_E - y_A$ получается через (5.67) и (5.68).

Если время не торопит, $y_E - y_A$ может быть получено без какого бы то ни было запроса у первичной лаборатории, находящейся в E, просто путем консультаций с публикациями BIPM. Они дают местный доступ к TAI (т.е. через прием сигналов GPS). Они так же обеспечивают через обозначение d относительную разность между длительностью единицы шкалы TAI и собственной секундой, реализуемой стандартом E на геоиде. Мы используем здесь обозначения d_{TAI} для этой величины, которая является средней на интервале (TAI_1, TAI_2). Следовательно

$$[\langle y_E \rangle - \langle y_{TAI} \rangle]\,|_{geoid} = d_{TAI,E}, \tag{7.20}$$

где $\langle\rangle$ представляет усреднение за соответствующий интервал. Для часов пользователя A с собственным временем τ_A мы измеряем

$$N_A = \frac{[\tau_A - TAI](TAI_2) - [\tau_A - TAI](TAI_1)}{TAI_2 - TAI_1}, \tag{7.21}$$

затем из (5.73)

$$\langle y_A \rangle - \langle y_{TAI} \rangle = N_A + \langle \underline{h}(TAI) \rangle . \qquad (7.22)$$

Функция $\underline{h}$ дается выражением (3.30) или (3.32) и ее среднее значение берется вдоль мировой линии А между TAI_1 и TAI_2. Следовательно

$$\langle y_A - y_E \rangle = N_A - d_{TAI,E} + \langle \underline{h}(TAI) \rangle . \qquad (7.23)$$

Для фиксированных часов на Земле мы находим

$$\langle y_A - y_E \rangle = N_A - d_{TAI,E} - 1{,}091 \times 10^{-13} h_A / km , \qquad (7.24)$$

где h_A – высота пользователя над геоидом в километрах.

Либо путем прямых сличений времени, либо через публикации BIPM неопределенности, вносимые в оценку первичных стандартов частоты, могут быть понижены меньше чем до 10^{-14}. Однако такие неопределенности слишком велики для новейших стандартов, в которых неопределенность частоты достигает 10^{-15}. Проблема увеличения удаленного доступа к частотам этих стандартов все еще не решена, за исключением лабораторий, оборудованных системами дуплексной спутниковой передачи времени и частоты TWSTFT (см. раздел 5.6.3(3)).

7.5.2. Время

GPS Время связано с TAI через

$$[TAI - GPS\ Time] = 19\text{с} + C_0 , \qquad (7.25)$$

где расхождение в синхронизации C_0 обычно поддерживается в пределах ± 100нс. С марта 1995 года до настоящего времени (май 2002 года) оно оставалось внутри этих границ и систематически уменьшалось. В настоящее время оно часто лежит в пределах меньше 10 нс. Значения C_0 публикуются на каждые сутки в 0 часов UTC в циркуляре BIPM T. Неопределенность составляет несколько наносекунд. Следовательно, доступ к TAI и UTC, хотя и отличается одним или двумя месяцами в реальном времени, возможен везде в пределах ±10 нс при использовании промышленных приемников времени. Для того чтобы достичь такой точности дневное усреднение должно быть выполнено на большом количестве наблюдений спутников. В дополнение координаты антенны должны быть известны в пределах одного – двух метров в системах WGS84 или ITRF. GPS приемники времени могут дать эти координаты, но с недостаточной точностью.

Другой очень точный способ доступа к UTC или TAI, опять отсроченный одним или двумя месяцами, заключается в том, чтобы установить связь с лабораторией *k*, часы которой принимают участие в формировании TAI и для которой публикуется $UTC - UTC(k)$. Неопределенности

могут быть несколько ниже 10 нс. Этот метод порождает проблему линий передачи времени на короткие расстояния. Может использоваться GPS (некоторые ошибки уменьшаются с расстоянием), в других случаях дальнейшие возможности представляются кабельной связью или прямым телевидением.

В настоящее время (2002 год) доступ через GLONASS имеет незначительно уступающее качество по сравнению с доступом через GPS из-за меньшего количества спутников. Положив

$$[UTC - GLONASS\ \ Time] = C_1, \tag{7.26}$$

где значения C_1 лежат в пределах нескольких сотен наносекунд. Они публикуются BIPM на каждые сутки с неопределенностью около 10 нс.

В реальном времени мы можем использовать либо значения UTC, передаваемые GPS (±100 нс), либо связь с лабораторией, где $UTC - UTC(k)$ могут быть экстраполированы с возможностью достижения неопределенности, достигающей величины около ±100 нс.

Если не требуется большая точность, то существует множество доступов к шкале UTC или к официальному времени, такому, как говорящие часы или радиопередачи типа шести точек. Некоторые кодированные излучения могут быть использованы для показа времени. Конечно не определено, что радиосигналы времени, направленные главным образом для навигаторов, будут способны устоять против возможностей, предлагаемых другими развивающимися средствами распространения времени и определения положения, а именно GPS, GLONASS или их гражданскими партнерами.

7.6. Итоги и перспективы

Текущая мировая организация хранения атомного времени была установлена в начале 1970 годов. Со временем выбор определился тремя главными подходами.

- Существует много цезиевых атомных часов, выполненных промышленным способом и обладающих отличной долговременной стабильностью, но не имеющих точности и кратковременной стабильности. Систематические сдвиги разности частот преследуют их на протяжении всей их жизни.
- Первичные стандарты частоты работают только на случайной основе и остаются сомнения относительно их заявленной точности.
- Чтобы уменьшить неопределенность сличений частоты через сличения времени необходимо производить усреднения на одном-двух месяцах.

В начале 1980 годов был поднят вопрос об организации способов измерения атомного времени. Первичный стандарт NRC Cs V в Канадском национальном исследовательском совете работал непрерывно (как *первичные часы*) с 1976 года. В 1978 году к нему присоединились немецкие первичные часы PTB Cs1 и затем в 1980 году еще трое других первичных часов NRC, а именно NRC Cs VI A, B и C. Между этими стандартами было хорошее совпадение частоты, а непревзойденная точность PTB Cs1 ($\pm 3\times10^{-14}$) была подтверждена в 1986 году часами PTB Cs2, которые оказались даже более точными ($\pm 1{,}5\times10^{-14}$). Эти разработки сделали резонной постановку вопроса, должна ли TAI основываться непосредственно на таких часах и только на таких часах. Вычисления TAI из показаний часов многих лабораторий тем не менее продолжались, так как это гарантировало надежность результатов. Однако были другие причины, а именно желание, чтобы TAI оставалось коллективным предприятием, включающим значительное число лабораторий и больших и маленьких. Число стран, присоединившихся к этому клубу, к 2000 году выросло до тридцати. Этот дух сотрудничества в области измерения времени ни разу не провалился. Именно через это сотрудничество стала возможной тесная синхронизация с UTC, включая синхронизацию GPS, которая не нужна системе для работы, но которая служит просто нуждам мирового сообщества. Этот же дух сделал возможным установление сети сличений времени, которая сформировала фундамент для всего сооружения.

Текущие исследования вероятно приведут даже к более точным и стабильным первичным стандартам с неопределенностями, уменьшенными по крайней мере в 10 раз. Но с 1994 года большинство национальных лабораторий времени сами были оборудованы новыми промышленными цезиевыми часами, у которых долговременная нестабильность также была уменьшена примерно в 10 раз по сравнению с предыдущими часами. Таким образом, эти приборы также прогрессируют с той же скоростью. Точность сличений через GPS сделала возможным уменьшение времени выборки, по которому вычисляется TAI, с двух месяцев до одного.

Несмотря на эти измерения и на давление прогресса первоначальная организация TAI 1970 года продолжает работать. Она часто переоценивается консультативным комитетом по времени и частоте, но до сих пор заключения комитета всегда оставляли все как есть.

В заключение давайте напомним, что время в TAI и UTC доступно везде в пределах сотни наносекунд в реальном времени и в пределах около 10 наносекунд за один или два месяца. Взятая как сущность секунда, т.е. длительность одной (единственной) секунды, может быть реализована с относительной ошибкой около $\pm 5 \times 10^{-14}$ через стандарты на холодных атомах цезия благодаря их кратковременной нестабильности. В то же время эти стандарты могут выразить через секунды длительность нескольких часов в пределах 10^{-15} (относительное значение). Через доступ к TAI секунда может быть получена везде в пределах нескольких единиц от $\pm 10^{-15}$ на усреднении в несколько суток.

Глава 8

Астрономические времена

Хотя астрономические времена больше не являются лучшей мерой времени, они продолжают играть важную роль в современных исследованиях. Всемирное время UT свидетельствует о вращении Земли и используется для установления всемирного координированного времени UTC, являющегося практической базой единого времени для всего мира. Эфемеридное время ET являлось лучшим представлением динамического времени до наступления в 1955 году атомного времени, и оно все еще используется для обработки астрономических наблюдений, выполненных до этой даты. И, наконец, существует некоторая надежда, что *пульсарное время* может проявить лучшую долговременную стабильность, чем атомное время. В данной главе мы объясним, как эти времена определяются, реализуются и применяются.

8.1. Всемирное время

8.1.1. Концептуальные определения

Историческое рассмотрение развития, приведшего к всемирному времени UT, представляет нам серию метаморфоз истинного солнечного времени (см. раздел 4.2). Проще дать основанное на вращении Земли концептуальное определение, в котором проявляется современная роль всемирного времени.

Существуют различные формы всемирного времени. Одна из них, непосредственно связанная с вращением Земли, называется UT1. Другие формы будут определены в разделе 8.1.3.

В разделе 3.3.2b мы определили геоцентрические координатные системы. Одна из них является невращающейся в пространстве (*небесная система*), а другая вращается вместе с Землей (*земная система).* Вращение земной системы по отношению к небесной системе является *вращением Земли.* Мы покажем, как эти системы реализуются в международной опорной небесной системе ICRF, обозначаемой здесь через (C), и международной опорной земной системе ITRF, обозначаемой здесь (T). Должно быть описано вращение системы (T) по отношению к системе (C). Для того чтобы сделать

Глава 8

Астрономические времена

Хотя астрономические времена больше не являются лучшей мерой времени, они продолжают играть важную роль в современных исследованиях. Всемирное время UT свидетельствует о вращении Земли и используется для установления всемирного координированного времени UTC, являющегося практической базой единого времени для всего мира. Эфемеридное время ET являлось лучшим представлением динамического времени до наступления в 1955 году атомного времени, и оно все еще используется для обработки астрономических наблюдений, выполненных до этой даты. И, наконец, существует некоторая надежда, что *пульсарное время* может проявить лучшую долговременную стабильность, чем атомное время. В данной главе мы объясним, как эти времена определяются, реализуются и применяются.

8.1. Всемирное время

8.1.1. Концептуальные определения

Историческое рассмотрение развития, приведшего к всемирному времени UT, представляет нам серию метаморфоз истинного солнечного времени (см. раздел 4.2). Проще дать основанное на вращении Земли концептуальное определение, в котором проявляется современная роль всемирного времени.

Существуют различные формы всемирного времени. Одна из них, непосредственно связанная с вращением Земли, называется UT1. Другие формы будут определены в разделе 8.1.3.

В разделе 3.3.2b мы определили геоцентрические координатные системы. Одна из них является невращающейся в пространстве (*небесная система*), а другая вращается вместе с Землей (*земная система*). Вращение земной системы по отношению к небесной системе является *вращением Земли*. Мы покажем, как эти системы реализуются в международной опорной небесной системе ICRF, обозначаемой здесь через (C), и международной опорной земной системе ITRF, обозначаемой здесь (T). Должно быть описано вращение системы (T) по отношению к системе (C). Для того чтобы сделать

это описание, мы можем обратиться к классической кинематике. Время t понимается здесь как абсолютное и реализуется международным атомным временем TAI с пренебрежимо малой неопределенностью.

Хотя для ориентации (T) по отношению к (C) достаточно трех функций времени (три угла Эйлера), предпочтительнее представление из пяти параметров, включающих направление оси вращения (I). Ось перемещается по отношению и к (T) и к (C).[1] Используются следующие параметры.

- Две угловые координаты, представляющие положение (I) в (C), описывают *прецессию-нутацию*. Наблюдаемые значения обычно представляются в виде отклонений от теоретической модели прецессии-нутации выбранной по соглашению. В настоящее время отклонения от теоретической модели составляют около 0,01". Они будут уменьшены до 0,001", когда новая модель, принятая в 2000 году, будет внедрена в практику.
- Две угловые координаты (I) в (T), описывающие *движение земных полюсов*. Они представляются в виде отклонений между направлением оси (I), которая должна быть ориентирована на север, и оси X [3] системы (T), обозначаемой через Z в публикациях международной службы вращения земли IERS: $x(t)$ измеряется вдоль начального меридиана, $y(t)$ – перпендикулярно меридиану на запад. Значения $x(t)$ и $y(t)$ не превышают 1", что соответствует 30 м на поверхности Земли.
- Завершает представление угол $A(t)$, описываемый вокруг оси вращения (I). Положение оси (I) это то положение, которое она занимает на дату t.

Угол A может быть определен, как мы увидим, двумя способами. UT1 выражается в виде функции от A через математические соотношения в которых числовые параметры выбраны так, что какой бы ни был сделан выбор, получались бы те же самые значения UT1. Эти соотношения установлены так, чтобы удовлетворить следующим условиям.

(1) Производная UT1 по времени должна быть пропорциональна модулю (изменяющегося) вектора вращения. Тогда, в предположении равномерного вращения, UT1 выполняет свою историческую роль в качестве равномерного времени и его нынешняя роль состоит в простом представлении вращения Земли.

(2) Константа пропорциональности, подразумеваемая в (1), и фаза выбраны так, чтобы и через тысячелетия Солнце проходило через начальный меридиан (Гринвичский меридиан) в среднем в 12 часов UT1.

---

[1] Мгновенная ось вращения (IAR) имеет суточное и внутри-суточное движения и в (T) и в (C). Хотя амплитуды движения меньше чем 0,02", они создают практические неудобства. IAR заменяется близкой к ней осью, не имеющей этих движений. Эта ось впервые была определена международным астрономическим союзом IAU в 1976 году и затем ее определение было улучшено в 2000 году. Соответствующие полюса называются звездным эфемеридным полюсом CEP и звездным переходным полюсом (Celestial Intermediate Pole) CIP соответственно. В дальнейшем неявно подразумевается CIP вместе с его экватором.

Строго говоря, условие (2) несовместимо с условием (1), потому что даже если вращение Земли равномерное, то средняя продолжительность истинных солнечных суток очень медленно изменяется. Однако предпочтение было отдано условию (1), а условие (2) выполняется только приблизительно. Происходящий из-за этого приближения сдвиг между UT1 и чередованием дня и ночи ничтожно мал и составляет меньше одной секунды за несколько столетий.

Если вектор вращения земли $\boldsymbol{\omega}(t)$ и его длина $\omega(t)$, то мы имеем

$$\frac{\mathrm{d}(UT1)}{\mathrm{d}t} = \frac{\omega(t)}{\omega_0}. \tag{8.1}$$

где ω_0 – постоянная угловая скорость, соответствующая $\mathrm{d}(UT1) = \mathrm{d}t$, значение которой выбрано так, чтобы удовлетворить условию (2). В интегральной форме это означает, что

$$[UT1](t) = [UT1](t_0) + \frac{1}{\omega_0}\int_{t_0}^{t} \omega(t)\mathrm{d}t, \tag{8.2}$$

где постоянная интегрирования также выбирается так чтобы удовлетворить условию (2).

Уравнение (8.2) выражает концептуальное определение UT1, в то время как соотношения между двумя видами A и UT1, представленные в разделе 8.1.2, дают практические определения.

Давайте рассмотрим некоторые полезные соотношения. Так как значения UT1 публикуются в виде поправок к TAI и являются функциями даты TAI (или UTC) даты, то иногда бывает удобно написать (8.2) в виде:

$$[UT1-TAI](t) = [UT1-TAI](t_0) + \int_{t_0}^{t} \frac{\Delta\omega(t)}{\omega_0}\mathrm{d}t, \tag{8.3}$$

где $\Delta\omega = \omega - \omega_0$.

Величина $\omega(t)$ может быть выведена из публикуемых значений $UT1 - TAI$ через

$$\omega(t) = \left[1 + \frac{\mathrm{d}(UT1-TAI)}{\mathrm{d}(TAI)}\right]\omega_0. \tag{8.4}$$

Вместо ω_0 обычно используют величину $D(t)$, называемую *длительностью суток*, которая означает длительность в секундах TAI, соответствующую возрастанию на 24 часа в UT1. Тогда

$$D(t) = \frac{\omega_0}{\omega(t)} \times 86400\,\mathrm{c}. \tag{8.5}$$

8.1.2. Практическое определение UT1

Теперь мы остановимся на деликатном вопросе определения направлений, между которыми угол $A(t)$ измеряется двумя способами.

(а) Использование невращающегося начала

Недавно была найдена такая форма $A_S(t)$ представления $A(t)$, иногда называемая *звездным углом (stellar angle),* в соответствии с которой этот угол линейно связан с UT1 [8.1, 8.2].

Следуя этому методу, обычно используемому астрономами, мы представим направления прямыми линиями, исходящими из некоторой произвольно выбранной точки О. Эти линии пересекают сферу единичного радиуса (*небесная сфера*) в точках, которые могут быть использованы для того чтобы представить эти направления. Экватор является большим кругом на этой сфере и $A_S(t)$ представляется дугой на экваторе на дату t.

Начнем с определения начала $A_S(t)$ в (Т). Для упрощения предположим что ось X^3 соответствует положению P_0 полюса на дату t_0. Пусть Ω_0 будет началом счета долгот в (Т), т.е. точкой, представляющей X^1. На дату t полюс передвинется к Р вдоль траектории С (см. рис.8.1). Экватор, соответствующий полюсу Р, пересечет траекторию экватора, связанного с P_0 в точке N (*восходящий узел*). Для того чтобы установить положение *мгновенного начала счета долгот* Ω закладывается следующее условие: прямоугольная триада, связанная с ОР и ОΩ, не имеет мгновенной компоненты вращения вокруг ОР при движении Р. Если Р выражается через

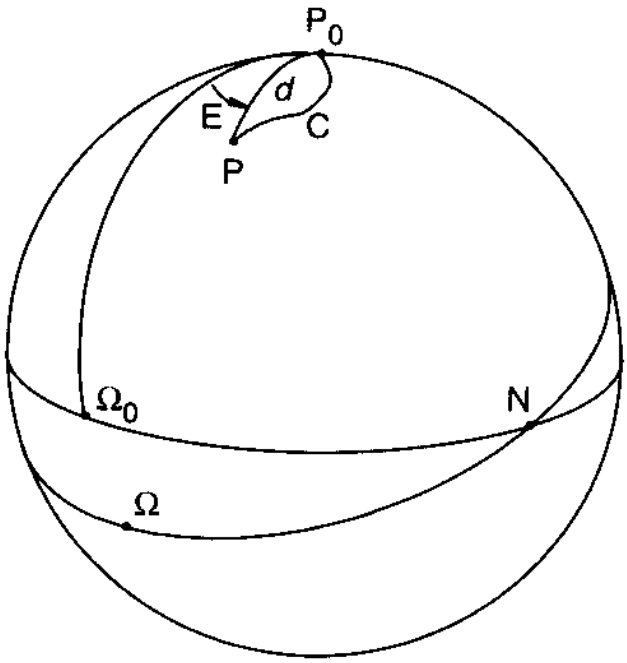


Рис. 8.1. Определение начала Ω мгновенной долготы на движущемся экваторе.

полярные координаты, как показано на рис.8.1, то легко показать, что величина $s = \Omega N - \Omega_0 N$ определяется из

$$s = \int_C (\cos d - 1)\mathrm{d}E\,. \tag{8.6}$$

Положение Ω зависит, таким образом, от истории вращения Земли – это вызывает некоторое беспокойство, но является неизбежным выводом. При этом s мало, потому что d никогда не превышает $1''$. Это удачное обстоятельство, и на практике s всегда принимается равным нулю. Ситуация совершенно отличается в системе (C).

При истолковании в системе (C) та же проблема приводит нас к определению начала на движущемся экваторе, которое называется *невращающимся началом* σ. Однако движение P в (C) включает в себя вековой член (прецессия) с величиной около$20''$ в год и периодические члены (нутация) с общей амплитудой около $20''$, так что s может иметь значительную величину. Например, для прецессии мы находим

$$s_P = 36{,}28''(t - t_0)^3 - 0{,}04''(t - t_0)^4 + \ldots, \tag{8.7}$$

где $t - t_0$ измеряется в тысячелетиях. Полное выражение для s, включающее прецессию и нутацию и справедливое вплоть до 2100 года до нескольких единиц $10^{-6''}$, дано в работе [8.3].

Установив начало Ω и σ, мы можем определить $A_S(t)$ по модулю 2π как дугу $\Omega\sigma$. Так как по определению невращающего начала каждая компонента вращения вокруг полярной оси равна нулю, то $A_S(t)$ представляет непосредственно интеграл в выражении (8.2). Значения UT1 пропорциональны A_S. Полезно вывести соотношения между UT1 и A_S из выражения A_S на 0 часов UT1 каждых суток. Пусть d_u' будет числом UT1 суток, прошедших с 1 января 2000 года от 12 часов UT1 (d_u' положительное или отрицательное целое число плюс или минус 0,5). Тогда, устанавливая $T_u' = d_u'/36\,525$ (выражаемое в юлианских столетиях по 36 525 суток) и принимая в качестве угловой единицы одну секунду времени, т.е.$15''$ (обычная практика для астрономов), мы имеем:

Для A_S в 0 часов

$$\mathrm{UT1} = 24110{,}54841 + 8639877{,}31738T_u'\,. \tag{8.8}$$

Это уравнение устанавливается путем применения значения ω_0

$$\omega_0 = 7{,}292115146706 \times 10^{-15}\ \text{радс}^{-1}\,. \tag{8.9}$$

Чтобы получить A_S в какой-то момент времени, d_u должно вычисляться таким же способом как и d_u', но в сутках UT1 и долях суток для этого момента вместе с соответствующим T_u путем добавления 86400×36525 секунд к коэффициенту при T_u'. Например, для A_S в оборотах

$$A_S = 0{,}779057273264 + 1{,}0027378119113 54 d_u . \qquad (8.10)$$

На своей генеральной ассамблее в 2000 году IAU рекомендовал использовать связанное с небесным переходным полюсом (CIP) невращающееся начало, которое называется небесным эфемеридным началом в (C) и земным эфемеридным началом в (T). Внедрение этой рекомендации должно начаться в 2003 году.

(b) Использование равноденствия

Хотя звездный угол A_S и начинает применяться на практике, давайте вернемся к разработкам, основанным на традиционном начале в (C). Этим традиционным началом является *весеннее равноденствие* Υ, пересечение экватора и эклиптики. Величина $\Omega\Upsilon$ является *гринвичским звездным временем* GST. Оно связано с UT1 много более сложным образом, чем звездный угол. Оно включает периодические члены, которые следует вычесть, чтобы определить *гринвичское среднее звездное время* GMST. Затем, согласно решению IAU [8.4, 8.5] и предыдущему определению T_u' в секундах времени, мы находим

$$GMST \quad \text{на} \quad 0\text{h} \quad \text{UT1} = 24110{,}54841 + 8640184{,}812866 T_u' + 0{,}093104 T_u'^2 - 6{,}2\times 10^{-6} T_u'^3 \qquad (8.11)$$

Соотношение в произвольное время выводится таким же образом, как и для A_S.

8.1.3. Другие виды всемирного времени

Иногда требуется оценить всемирное время в таких случаях, когда не известно мгновенное положение полюсов Земли. Для этого на время оценки принимается, что $x = 0 = y$, и это приводит нас к виду времени UT0. Разность *UT1 – UT0* дается через

$$UT1 - UT0 = -(x\sin L + y\cos L)\operatorname{tg}\phi , \qquad (8.12)$$

где ϕ и L – широта и долгота (положительная в направлении на восток) наблюдателя, а x и y координаты полюса. На средних широтах эта разность меньше нескольких сотых секунды.

Существует приближенно-равномерный вид всемирного времени, обозначаемый как UT2, в котором введены поправки на сезонные неравномерности. Разность между *UT2 – UT1* колеблется в пределах ±30 мс. UT2 представляет вращение Земли не больше, чем это делает UT0. Только UT1 имеет физическое значение.

Также определен вид UT1R, в котором в UT1 введены поправки на зональные приливные эффекты, воздействующие на полярный момент инерции и имеющие периоды меньше 35 суток. Поправки составляют по

величине менее 2 мс. Наконец UT1R' имеет поправки на эффекты всех приливов, основной член имеет амплитуду 0,16 с и период 18,6 лет.

8.1.4. Техника измерений

Примерно до 1970 года наши знания о вращении Земли основывались целиком на оптических наблюдениях видимого движения звезд по отношению к земным опорным направлениям. После примерно десятилетнего переходного периода оптические методы были заменены методами, используемыми в технике космической геодезии и интерферометрии.

Ряд этих методов сосуществует. Они обеспечивают дополнительные данные для установления небесных и земных опорных координатных систем и мониторинга пяти параметров, точно устанавливающих вращение Земли. Они также являются конкурирующими в некоторых областях, где они приводят к результатам, имеющим сравнимое качество. В следующем кратком обозрении мы увидим, что все эти методы основываются на время-частотной технике.

(а) Радиоинтерферометры со сверхдлинной базой (РСДБ)

РСДБ сначала были разработаны для изучения структуры небесных радиоисточников. С начала 1980-х годов они были использованы для измерения направления этих источников по отношению к земным базисам.

Метод измерения включает в себя прием радиоизлучения квазаров в сантиметровом и дециметровом диапазонах длин волн с использованием двух антенн, разнесенных на несколько сотен или несколько тысяч километров. Эти антенны определяют базис. Характеристики излучения записываются на магнитную ленту в виде функции времени, которое определяется часами, расположенными у каждой антенны. Давайте предположим, что часы синхронизированы. Тогда можно сравнивать магнитные записи и искать корреляцию между сигналами, принятыми двумя антеннами. Таким образом может быть определена временная задержка между этими сигналами и, если расстояние между антеннами известно, то также может быть определен и угол между базисной линией и направлением на источник (рис.8.2).

Когда внедрялся этот метод, часы еще не могли быть синхронизированы достаточно точно внешними методами, такими как описаны в главе 5. Поэтому часы оставались независимыми и на протяжении измерений, объединявшихся в единый пакет (часто 24 часовой), сличения времени должны были моделироваться, например, полиномом второго порядка по времени. Параметры модели, такие как коэффициенты этого полинома, оставались неизвестными и должны были выводиться из самих РСДБ наблюдений. Поэтому было существенно важным использование часов,

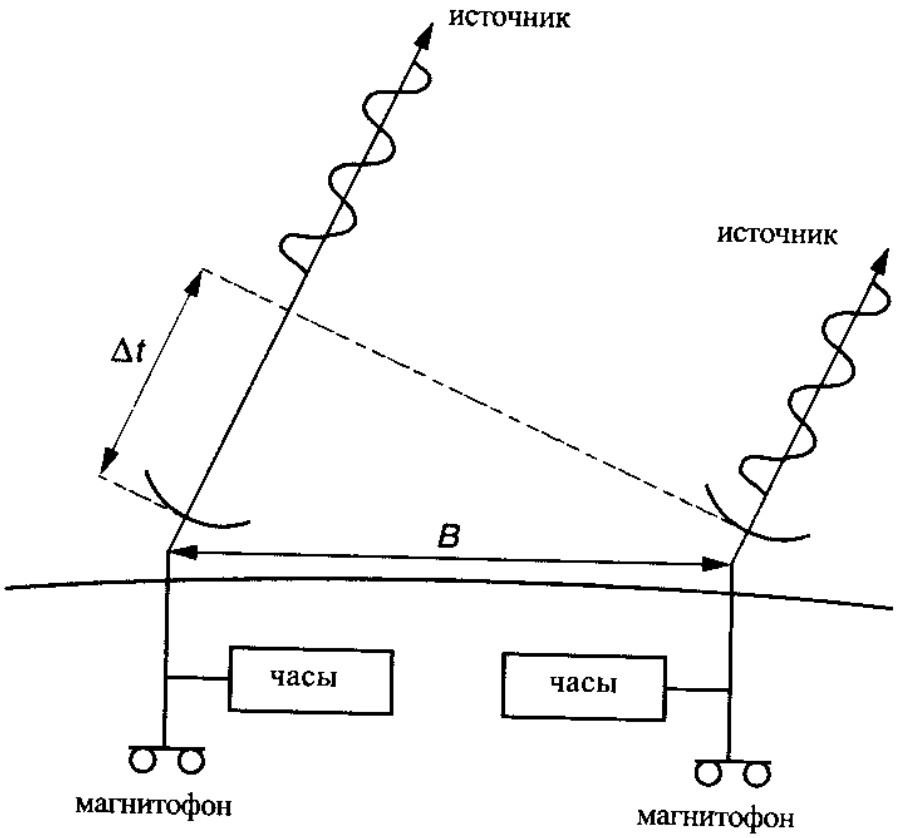


Рис. 8.2. Астрометрическое применение РСДБ. Путем измерения временной задержки Δt, образованной из-за геометрии, мы можем найти ориентацию базисного вектора B по отношению к направлению на источник.

максимально стабильных по частоте за периоды выборки, длящиеся до периодов соответствующих пакетов наблюдений, так чтобы модель их временной разности содержала минимум параметров. Отсюда становится ясным, почему применяются водородные мазеры и это действительно одно из их основных применений.

Ни координаты антенн, ни положение источника также неизвестны априори, или известны с недостаточной точностью. Если анализируются наблюдения нескольких источников в глобальном масштабе с использованием нескольких базовых линий, то можно отыскать различные неизвестные параметры. РСДБ может обеспечить все пять параметров вращения Земли для конструирования и небесных и земных координатных опорных систем. Но земная опорная система может быть получена только с точностью до параллельного переноса. Положение в центре масс Земли не может быть определено.

Применение РСДБ к астрометрии и геодезии координируется международной службой РСДБ (IVS). Мы уже упоминали в разделе 5.6.3c, что как побочный результат мы могли бы иметь сличения часов, если бы была измерима инструментальная задержка РСДБ.

(b) Динамические методы

В РСДБ природа непосредственно приводит нас к невращающейся небесной системе координат. Таким образом расчищается путь для геометрии. Для методов, которые здесь рассматриваются, наблюдения относятся к объектам, находящимся на орбитах вокруг Земли, а именно к искусственным спутникам и Луне. Невращающаяся опорная система координат реализуется через динамическое изучение движения таких объектов. Для обнаружения существенных особенностей этих так называемых динамических методов давайте примем в высшей степени схематический подход. Представим, что Земля сферическая, однородная и изолированная в пространстве, и что спутник вращается по круговой орбите. Плоскость орбиты имеет фиксированную ориентацию в пространстве. Далее, орбитальный период спутника является постоянным и может быть измерен с большой точностью путем выполнения наблюдений в различных точках земной поверхности настолько долго, насколько это необходимо. Затем мы можем установить геоцентрические оси в орбитальной плоскости и, таким образом будет установлена невращающаяся система.

На практике динамические методы очень сложны из-за необходимости учета значительных возмущений. Эти возмущения включают в себя эффекты соседних небесных тел, неравномерности в форме и во вращении геоида, остаточные эффекты трения, давление солнечного излучения и рефракция в земной атмосфере. Все соответствующие проблемы решены, и динамические методы теперь достигают точности, сравнимой с РСДБ везде, где применимы оба метода.

Динамические методы могут определить движение полюсов Земли, обеспечить данные относительно прецессии-нутации и реализовать земную опорную систему координат, связанную с центром масс Земли (последнее РСДБ сделать не может). Однако остаточные члены, не включенные в поправки для вышеуказанных возмущений, приводят к дрейфу значения UT1. Этот дрейф становится более выраженным при увеличении возмущений, действующих на спутники. Для Луны он очень мал, но не пренебрежим. По этой причине такие методы очень хороши для изучения кратковременных вариаций UT1. Они могут быть распределены по категориям в соответствии с наблюдаемым объектом и способом его наблюдения.

Координаты полюсов впервые были получены в 1967 году путем слежения с использованием доплеровского эффекта за спутниками спутниковой системы Navsat. Улучшенной модификацией этой техники является система DORIS, разработанная французским национальным центром по изучению космоса (CNES) и предназначенная для слежения за спутниками Spot, а также за франко-американскими океанографическими спутниками TOPEX-POSEIDON.

Для исключительно точных применений был запущен ряд сферических спутников во Франции (Starlette, 1975 г.) в США (Lageos. 1976 г.), затем в Японии и в России. Эти спутники покрыты уголковыми отражателями и сделаны очень плотными для того, чтобы уменьшить эффекты возмущения негравитационного происхождения. Расстояние до них измеряется лазерными дальномерами и неопределенности в измерении расстояния исчисляются величиной порядка одного сантиметра. Кроме непосредственного вклада этих спутников в измерение вращения Земли и создание земной системы координат они играют важнейшую роль в отображении геоида. Это в свою очередь является главным вкладом в геофизику и во всю ту технику и технологию, которая базируется на точной орбитографии искусственных спутников Земли. Жаль, что лазерная телеметрия обходится так дорого и так чувствительна к условиям облачности, потому что это очень точный метод, который может быть использован также и для калибровок. Международная служба лазерной дальнометрии (ILRS) обеспечивает координацию этой деятельности.

Динамический метод, включающий в себя наблюдение за двадцатью четырьмя GPS спутниками, кажется наиболее эффективным. Координация и анализ наблюдений являются задачей международной службы GPS (IGS). Эта служба обеспечивает важные результаты по движению земных полюсов и геодезических координат.

Три отражающие панели были установлены на Луне в 1969 астронавтами Аполло и еще две (сделанные французами), советскими беспилотными Луноходами. Одна панель никогда больше не обнаруживалась, но остальные четыре используются для лазерной дальнометрии. Лунная лазерная дальнометрия также достигла сантиметровой точности на этой исключительной технике, разработанной только во Франции и в США, где и получают регулярно данные наблюдений. По этой причине метод лазерной локации Луны привел не к глобальной оценке вращения Земли, а только к некоторым полезным величинам в UT1. Он также дал больший вклад в изучение системы Земля-Луна.

(c) Координация и результаты

Результаты, полученные каждым из технических центров (IVS, DORIS, ILRS и IGS) передаются в IERS, где они обрабатываются. IERS среди других результатов публикует также окончательные значения пяти параметров вращения Земли, настоящие, ежемесячные и за год [8.6, 8.7]. В 2000 году неопределенность в ориентации (I) при усреднении за 5 суток была около 0,00005" по отношению к (T), т.е. ± 1,5 мм для положения полюса вращения на поверхности Земли и 0,0002" по отношению к (C). Неопределенности в $UT1 - TAI$, усредненные также за пять суток, составляют около 10 мкс.

8.1.5. Неравномерности вращения Земли

Если Землю рассматривать как твердое тело, то параметры ее вращения становятся предсказуемыми как только путем измерений установлено небольшое число начальных условий. Согласно этому земные полюса должны бы выполнять круговые движения с периодом 305 дней, а их амплитуду и фазу достаточно было бы измерить только однажды. UT1 изменялось бы линейно со временем. Лунно-солнечная прецессия-нутация из-за влияния Луны и Солнца на земное экваториальное вздутие конечно является трудной проблемой теоретической астрономии, но она решена с неопределенностями значительно меньшими, чем остаточные неопределенности в наблюдениях. Необходимо было бы наблюдать только за одним параметром, а именно за динамически сплюснутой формой Земли.

Расхождения между наблюдениями и теорией вращения твердой Земли происходят из-за перемещений жидких компонент на планете, т.е. атмосферы, океанов и ядра по отношению к литосфере, а также из-за упругих и вязких деформаций так называемых твердых компонент. Эти перемещения и деформации могут привести, как мы увидим, к потере вращательной энергии. Даже если основной механизм выяснен, все равно невозможно сделать количественное прогнозирование их влияния на вращение Земли. Мы ограничимся обсуждением UT1. Читатель, заинтересовавшийся вышеперечисленными геофизическими явлениями, может обратиться к [8.8, 8.9].

(а) Вековое замедление вращения Земли

Оценки наиболее древних доступных длительностей суток D относятся ко времени примерно 400 миллионов лет назад. Суточные, лунные и годовые циклы записаны различными природными явлениями. Среди таких явлений укажем прежде всего рост кораллов и двустворчатых моллюсков и их укладку в отложениях. Продолжительность года оставалась практически неизменной, в то время как изменения лунного месяца, выраженные в равномерном времени, довольно хорошо известны. Путем счета суток в году и в лунном месяце можно определить, что длительность суток возрастает в среднем на 2 мс за сто лет.

Это значение согласуется с данными, которые можно получить из наблюдений затмений с древних времен. Появление затмения происходит под действие орбитальных движений и поэтому зависит от равномерного эфемеридного времени (см. раздел 8.2). Однако место, в котором будет наблюдаться затмение, зависит от вращения Земли. Так как размер этого географического сдвига иногда вызывает изумление, то мы обрисуем способ как его вычислить.

Например, предположим, что D возрастает на 2 мс за сто лет. Из (8.5).

$$\frac{\mathrm{d}\omega}{\mathrm{d}t} = -5{,}35\times10^{-22}\ \text{рад}\,\text{с}^{-2}. \tag{8.13}$$

Интегрирование в (8.2) выполняется на интервале времени, продолжающемся последние 2000 лет, т.е. от $t = -2000$ лет, до $t_0 = 0$ (настоящее время) в два приема. Для первого из этих параметров, мы положим $\omega = \omega_0$, а для второго:

$$\omega = \omega_0 + \frac{\mathrm{d}\omega}{\mathrm{d}t}\,t.$$

Получив разность, обнаружим, что с начала христианской эры накопилось отставание во времени на 4,1 часа, что соответствует сдвигу по долготе на 61°. Однако сдвиг узкой полосы, в которой солнечное затмение в эту эпоху было бы полным, смещается на восток только на 14° по отношению к тому смещению, которое было получено из расчета, в котором скорость UT1 имела бы современное значение. Причина уменьшения сдвига заключается в том, что вращение Земли оказывает возмущение на орбитальное движение Луны. Общий сдвиг тем не менее оказывается достаточно большим чтобы извлечь информацию о вращении Земли путем изучения архивных документов.

Вековое замедление вращения Земли приписывается главным образом океанским приливам. Вращательная энергия теряется в мелководных морях. Существует также перенос энергии на орбитальное движение Луны, которая в результате отдаляется от Земли на 4 см в год. Недавние исследования обнаружили значительные потери энергии в глубине океанов [8.10].

(b) Десятилетние флуктуации

Примерно на последних трех столетиях разность между UT1 и эфемеридным временем может быть восстановлена. На рис.8.3 показаны соответствующие значения длительности суток D. Флуктуации достаточно велики чтобы спрятать вековые изменения. Детали этих флуктуаций становятся все более и более ясно различимыми. В качестве много лучшей опоры с 1955 года стала доступной TAI, однако этот период еще слишком мал для изучения вариаций с типичными периодами в несколько десятков лет (вкладка на рис.8.3). Происхождение этих флуктуаций все еще недостаточно понято, но почти не осталось сомнений в том, что они должны происходить из-за взаимодействия между жидким ядром Земли и ее твердой мантией. Могут давать свой вклад и океаны. Они остаются совершенно непредсказуемыми. Наблюдалось, что иногда они приводили к относительным флуктуациям D и следовательно к флуктуациям длительности секунды UT1 с величиной, достигающей около 10^{-7}.

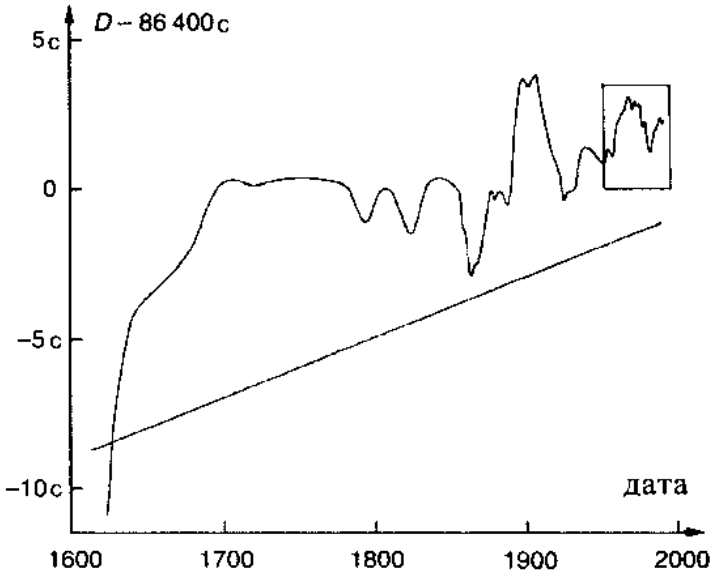


Рис. 8.3. Флуктуации длительности суток D, измеренного сначала в секундах эфемеридного времени до 1955 года (из [8.11]), а потом в секундах атомного времени (вставка). Наклонная прямая линия представляет вековое замедление вращения Земли.

(c) Кратковременные флуктуации

На рис.8.4 и 8.5 показаны флуктуации в UT2 и D с наилучшим достижимым разрешением во времени. Эти флуктуации в существенной мере обязаны движениям атмосферы и закону сохранения углового момента. При движении атмосферы происходит и обмен ее углового момента с моментом твердой Земли. На рис. 8.5 видна годовая компонента флуктуаций. На рис. 8.4 основная часть этой компоненты исключена.

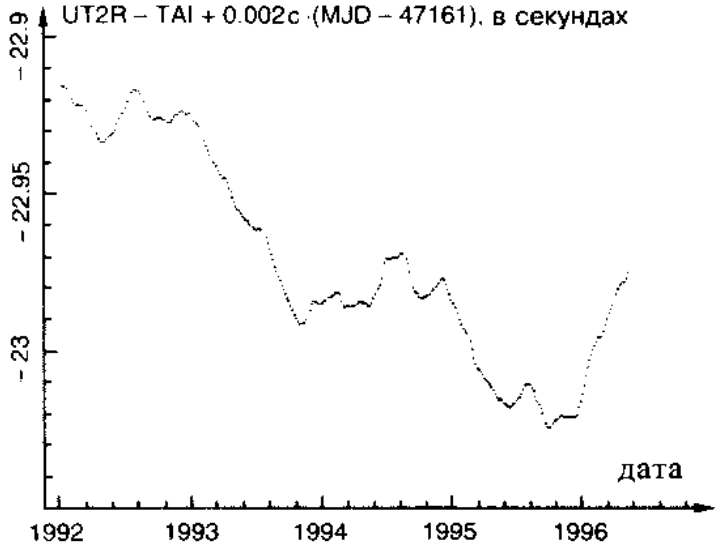


Рис. 8.4. Наблюденные значения всемирного времени как линейной функции времени после вычитания годовых вариаций, а также периодических членов из-за приливов. Детали рисунка являются реальными флуктуациями, а не ошибками наблюдений. (Рисунок любезно представлен IERS).



Рис. 8.5. Флуктуации длительности суток D в (атомных) миллисекундах после введения поправок на эффекты кратковременных зональных приливов. (Рисунок любезно представлен IERS).

(d) Частотные характеристики неравномерности всемирного времени

Нестабильность частоты всемирного времени показана приблизительно на рис. 8.7 в конце этой главы. На очень больших интервалах времени в ней доминирует вековое замедление вращения Земли.

8.1.6. Применение всемирного времени

UT1 используется для подстройки всемирного координированного времени, которому требуется не очень большая точность. Если необходима предельно высокая точность, то для ее достижения используется всемирное время, которое обеспечивает связь между небесной и земной системами координат, как функцию от TAI (см. раздел 3.3.2b). Всемирное время используется в комбинации с другими четырьмя параметрами вращения Земли. Эти параметры имеют также важное значение для геофизики. Хотя вычленять факторы, дающие вклад в описание вращения Земли, не очень хорошо, мы все-таки добавим краткие комментарии к каждому фактору.

Невращающаяся небесная система координат служит для установления динамических моделей движения тел в Солнечной Системе путем их наблюдения относительно этой системы координат, будь они естественного или искусственного происхождения. Модели обеспечивают *эфемериды* для данных объектов, позволяющие предсказать положение этих объектов. Конечно, практически существование системы координат опирающейся в радиодиапазоне на РСДБ не всегда помогает. Но она была расширена и на видимый диапазон путем установления углов и угловых

скоростей системы координат, образованной 100 000-ми звезд в программе астрометрического спутника HIPPARCOS [8.12, 8.13, 8.14, 8.15]. Это достигалось путем наблюдения звезд, которые излучают и в радио и в видимом диапазоне спектра. Эфемериды планет и их спутников, а также орбитография космических зондов имеют решающее значение для исследования солнечной системы. К тому же астрофизики нашли, что очень полезно знать собственные (угловые) движения звезд по отношению к невращающейся системе координат.

Очень точная равномерная земная система координат была установлена только недавно. В 1984 году международное бюро времени начало публиковать координаты, которые оно приняло для станций, участвующих в наблюдении вращения Земли, и таким образом создавать первые очертания того, что стало затем ITRF. Национальные и региональные геодезические сети, которые могли отличаться друг от друга до более чем 100 м, стали участвовать в подстройке с целью приближения к ITRF. Эта унификация оказалась тем более полезной в том плане, что GPS основана на системе, очень близкой к ITRF (с разницей в дециметры) и, следовательно, GPS обеспечивает координаты в этой системе. Сантиметровая точность ITRF обеспечивает опору для изучения тектонических деформаций там, где они связаны с крупномасштабным движением тектонических плит или локальным движением в критических районах.

Что касается параметров вращения Земли, то их важность в геофизике становится ясной из их непредсказуемости. Их изменения не всегда ведут к однозначности геофизической модели. Наоборот, они создают основу для отказа от моделей, предсказания которых слишком неточны. Эти параметры имеют также геометрическую важность при преобразованиях координат между небесными и земными системами. Например наблюдение космического зонда с помощью РСДБ дает его направление по отношению к Земле. Для того чтобы обеспечить встречу зонда с другой планетой необходимо вводить поправки в его траекторию, а это требует знания его положения в небесной системе. В этом примере параметры вращения Земли должны быть известны в реальном времени, что возможно только на основе их прогнозов. Этот пример показывает, почему хорошие значения для этих параметров необходимы немедленно, если неопределенности прогнозов должны быть минимизированы. IERS имеет срочную службу (в США), отвечающую за прогнозирование параметров вращения Земли.

8.2. Эфемеридное время

8.2.1. Определение

Концептуальное определение эфемеридного времени ET было дано в главе 4. Это вид классического динамического времени. Единица и начало ET определяются по соглашению путем принятия числового выражения для *геометрической средней долготы* Солнца, установленной Ньюкомбом, а именно:

$$L = 279^\circ 41' 48{,}04'' + 129602768{,}13'' T + 1{,}089'' T^2, \qquad (8.14)$$

где T считается в юлианских столетиях, состоящих из 36 525 *эфемеридных суток*. Начало T находится в начале 1900 года в момент, когда геометрическая средняя долгота Солнца приняла значение 279°41′48,04". Этот момент времени датировался как 0 января 1900 года ровно в 12 часов ET (решение IAU от 1958 года).

Для тех читателей, которые не знакомы с астрономической терминологией, сделаем краткий обзор. Долгота Солнца измеряется вдоль эклиптики от точки весеннего равноденствия. Последняя является пересечением эклиптики и экватора и она двигается в пространстве из-за движения этих двух плоскостей. Ее движение в эклиптике традиционно описывается двумя типами членов, именно: теми, которые представлены в виде полиномиальной функции от времени (*вековые члены)*, и теми, которые представляются точно периодическими функциями. Прилагательное «среднее» относится к тому факту, что периодические члены исключены. Более того, движение Солнца в эклиптике, в том виде как оно видно с Земли, включает в себя периодические компоненты из-за эксцентриситета земной орбиты вместе с возмущениями от других планет и Луны и (очень маленьких) вековых компонент. В средней долготе остаются только вековые члены.

Что касается прилагательного «геометрический», то оно относится к проблеме *аберрации*. Мы видим Солнце в положении, которое оно занимало 8 минут назад, время, необходимое свету чтобы достичь Земли. Это *кажущееся положение* Солнца. В геометрическое положение вносится поправка на эту задержку.

Выражение (8.14) содержит определение секунды эфемеридного времени, которая фактически основана на скорости в долготе $T = 0$. Официальное определение, данное в разделе 4.3.3, было принято так, чтобы показать секунду как часть длительности, но оно обращается к самому эфемеридному времени, а оно не определено. Значение этой части длительности находится путем вычисления длительности тропического года в секундах из (8.14) при увеличении L на 360°, если подставить $T = 0$ в член, содержащий T^2.

8.2.2. Установление эфемеридного времени

Эфемеридное время было сконструировано и использовалось прежде, чем появилось атомное время, или, по крайней мере, прежде, чем оно начало широко применяться среди астрономов. Поэтому в данном контексте мы должны поставить себя в это положение.

Выражение (8.14) можно рассматривать как начальное определение ET. Соответствующее установление времени состоит в измерении долготы Солнца в момент времени, который наблюдатель датирует датой τ_C по своим часам C. Из измерений долготы ET выводится через (8.14) и мы таким образом получаем $ET - \tau_C$. Так как неточность определений UT1 много меньше чем ET, то результат был дан в универсальным виде путем принятия

$$\Delta T = ET - UT1 \text{ на дату наблюдения.} \qquad (8.15)$$

Медленность изменений и отсутствие точности в ΔT допускают большую свободу в выборе опорной шкалы для датирования этих измерений.

Для того чтобы измерить долготу Солнца, оно должно быть расположено относительно звезд на небосводе. Ясно, что это не может быть сделано непосредственно. Что, однако, было возможно, так это датировать прохождение Солнца и звезд через меридиан в UT1 и использовать угловую скорость Земли для выведения разности между их прямыми восхождениями. Измерение зенитных расстояний во время этих прохождений дает разность в наклонениях, необходимую для того, чтобы оценить долготу Солнца. Необходимо множество поправок. Одна из них – аберрация, как мы уже видели, но есть еще поправка на параллакс, так как долгота в (8.14) является геоцентрической, а также поправки на другие астрономические явления. Фактически это начальное определение ET очень трудное и также очень неточное. Если мы предположим что вероятная неопределенность в L равна 0,5″, то соответствующая неопределенность в ET из наблюдений Солнца равна 10 с. Если мы надеемся измерить интервал времени в ET путем двух наблюдений Солнца с относительной неопределенностью 10^{-8}, то эти наблюдения должны проводиться в течение более 50 лет.

Поэтому должны были быть исследованы другие методы. Астрономы, установившие геоцентрические эфемериды, дающие угловое положение и расстояние до тел в солнечной системе, выбрали единицу и начало их временных параметров так, чтобы они удовлетворяли условиям, выраженным в (8.14). Если это однажды сделано, то тогда все эфемериды имеют ET в своих аргументах. В принципе наблюдение любого из этих тел дает доступ к ET с возможным увеличением точности. Однако, тем не менее остается тот факт, что этот способ нахождения ET зависит от способа калибровки с помощью солнечных наблюдений.

Чем быстрее орбитальное движение наблюдаемого тела, тем точнее определяется ET. В этом отношении Луна является наиболее подходящим объектом. К сожалению, движение Луны возмущается Земным вращением (см. раздел 8.1.5a). Ее средняя геоцентрическая угловая скорость (или *среднее движение)* постепенно уменьшается так, что это можно установить только путем наблюдений. Изменения в среднем движении Луны можно получить только в том случае, если сначала ET была бы определена через наблюдения Солнца и планет. После выполнения этих тщательных калибровок, требующих десятилетий скучной работы, Луна могла бы использоваться для практического получения ET. Луна является чем-то вроде вторичных часов, которые дают возможность получения отсчетов с большей точностью, но при этом возникает риск появления систематических ошибок, происходящих от каких либо ошибок калибровки, возрастающих со временем.

Чтобы гарантировать единую реализацию ET IAU рекомендовал применение специальных лунных эфемерид в обозначении ET*j*, где индекс *j* указывает, что выбрано. (С 1960 по 1984 годы использовались эфемеридные версии с j =0, 1, 2).

Движение Луны относительно звездного фона было измерено через покрытия звезд Луной, или путем наблюдений, или с помощью фотографических методов. Широкомасштабная кампания наблюдений была организована в конце 1950 годов. Несмотря на значительные приложенные усилия, эта техника все же привела к неопределенностям в несколько сотых секунды на годовом усреднении $\Delta Tj = ETj - UT1$. Когда была установлена TAI, определения ET стали не нужны и было проще принять, что

$$ET - TAI = 32{,}184\text{с}, \tag{8.16}$$

в котором договоренное значение константы основано на наблюдениях.

8.2.3. Трудности и текущие решения

Кроме неопределенностей, упомянутых в разделе 8.2.2, в ET могут быть большие систематические ошибки. Оно очень чувствительно к способу каким была реализована небесная координатная система, а также к значениям, принятым для различных астрономических констант. Например, каталоги, в которых приводятся положения и собственные движения звезд и реализация невращающейся геоцентрической координатной системы (ранее формировалась из системы (C), обсуждаемой в разделе 8.1.1) должны были пересматриваться несколько раз из-за улучшающихся измерений, а также потому, что неопределенности в собственном движении со временем теряют точность. Каждый пересмотр приводит к скачкам в ET. Чтобы проиллюстрировать роль астрономических констант, мы можем, например, рассмотреть влияние аберрационных констант, связанных со

средней долготой в (8.14), на наблюдаемую видимую долготу. Изменение в рекомендованном значении, принятом IAU в 1968 году, привело к скачку в ET на 0,6 с.

Чтобы преодолеть трудности, присущие определению ET через солнечную долготу, IAU в 1994 году просто рекомендовал, чтобы реализация земного времени TT, определенная в разделе 7.1.2, была распространена назад во времени на период до 1955 года через динамику солнечной системы. И действительно, единица длительности шкалы TT равна длительности атомной секунды на вращающемся геоиде и она была определена таким образом, чтобы в наилучшей степени быть согласованной с определением эфемеридной секунды. Более того начало TT было выбрано так, чтобы получить неразрывность с ET. Следовательно замена ET на TT не приводит к какой-либо большой разрывности в конструировании эфемерид.

В этом современном применении концепции эфемеридного времени, допустимом для релятивистской динамики, единица шкалы установлена современным атомным определением секунды и начало связано с международной атомной шкалой. Каждая эфемерида k тем не менее ведет к различным приближения TT(k) к TT, и это должно быть четко сформулировано (Резолюция C7, IAU, 1994 год).

В разделе 4.3.4 мы упомянули историческую роль, сыгранную ET. На рис. 8.7 даны некоторые идеи по ее частотной стабильности.

8.3. Пульсарное время

Пульсары наблюдались впервые в 1967 году [8.16]. Эти объекты посылают радиоимпульсы с регулярными интервалами. Предполагается, что они являются звездами, состоящими главным образом из нейтронов в состоянии предельного сжатия, излучающими радиопучок в фиксированном направлении по отношению к материи звезд. Импульсы излучения, получаемые здесь на Земле, свидетельствуют о вращении пульсаров.

Все пульсары, открытые между 1967 и 1982 годами имеют период вращения порядка одной секунды. На частоту, с которой их импульсы достигают Земли, действует изменение в расстоянии между пульсаром и наблюдателем, в частности, действует орбитальное движение Земли. После введения поправок и при использовании в качестве опоры атомной шкалы времени, было получено, что относительная нестабильность частоты этих импульсов лежит в диапазоне от 10^{-10} до 10^{-11}. Однако в 1982 году астрономы добавили в свою коллекцию небесных объектов другую категорию пульсаров. Это были *миллисекундные пульсары* с периодом в несколько миллисекунд. Диаметры этих объектов оцениваются величиной около 20 км и их массы составляют около 1,4 массы Солнца. Первый

пульсар был открыт Д.С. Бейкером и его группой с применением исключительно большой фиксированной радиоантенны трехсотметрового диаметра в Аресибо (Пуэрто Рико) [8.17]. Этот пульсар PSR1937+21 имеет период 1,6 мс. Форма его импульсов показана на рис.8.6. Вскоре было осознано, что прибытие импульса от PSR1937+21 может быть датировано относительно TAI с неопределенностью в несколько десятых микросекунды. Более того, если сразу же учесть его постепенное замедление вращения, то пульсар демонстрирует долговременную стабильность частоты по крайней мере такую же хорошую, как у лучших атомных часов этого времени (нестабильность порядка 10^{-14} или меньше за один год). Было последовательно открыто несколько десятков других миллисекундных пульсаров, причем некоторые из них являются компонентами сложных звездных систем. Была разработана международная программа по установлению времени прибытия импульсов пульсаров. Наблюдения выполнялись на частотах порядка 1 ГГц с применением больших радиоантенн. Собираемая мощность очень мала и составляет порядка 10^{-16} Ватт с наибольших антенн.[2] Для этой программы не требовалось непрерывных наблюдений. Стабильность частоты и пульсаров и TAI такова, что наблюдения могут прерываться на периоды от несколько недель до нескольких месяцев без риска потерять хотя бы один период.

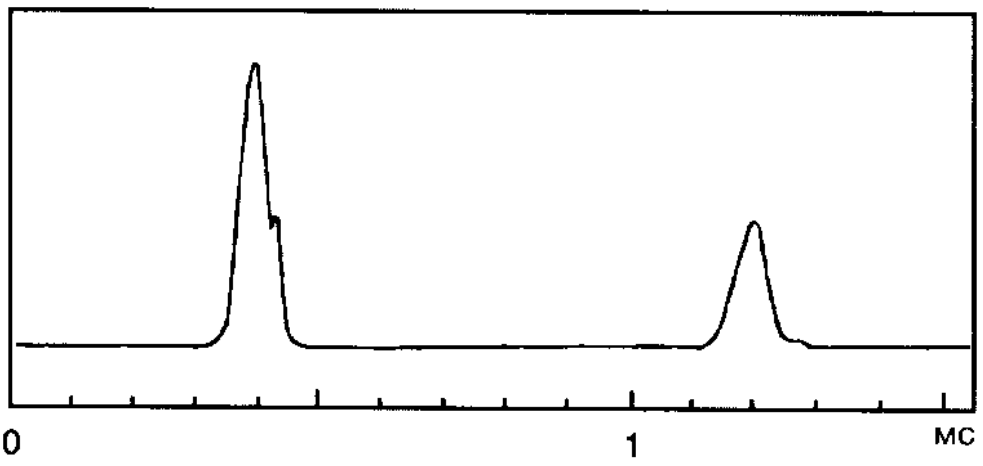


Рис. 8.6. Средняя форма импульсов от пульсара PSR1937+21. Из наличия промежуточного импульса предполагается, что магнитная ось пульсара перпендикулярна к оси вращения.

Появились предположения, что миллисекундные пульсары вернут измерения времени назад, в область астрономии. Однако при этом забыли эффект замедления вращения, который довольно сильно ограничивает

[2] В этой области используется единица потока Янски, где 1 Jy = 10^{-26} Вт $Гц^{-1}м^{-2}$. На типичных наблюдаемых частота получены значения от 10 до 100 mJy.

полезность измерения времени через вращение пульсаров. Например в случае с PSR1937+21 период P возрастает на $10{,}5\times10^{-20}$ с за секунду, или в относительном значении

$$\frac{1}{P}\frac{\mathrm{d}P}{\mathrm{d}t} = 7\times10^{-17}\,\mathrm{c}^{-1}. \tag{8.17}$$

Относительное замедление вращения оказывается в десять раз большим чем замедление вращения Земли. Хотя предполагают, что $\mathrm{d}P/\mathrm{d}t$ является константой, ее значение не может быть известно априори с достаточной точностью чтобы конкурировать со стабильностью частоты атомных часов. Существуют и другие проблемы. Фаза вращения пульсара должна передаваться на Землю, и в связи с этим появляются неопределенности из-за относительного движения наблюдателя и источника. Например, из-за орбитального движения Земли наблюдения зависят от положения пульсара, и его положение не может быть определено достаточно точно. К тому же пульсары имеют свое собственное движение, если они находятся в двойной системе. Могут быть также флуктуации времени распространения, если сигнал проходит через ионизованную среду в пространстве. По этим причинам вращение пульсара не может ни служить основой для определения единицы времени ни использоваться в качестве часов для реализации шкалы времени, например, используемой в теории TCG.

Как следствие, из-за замедления вращения и нестабильности, появляющейся при оценке изменений во времени распространения, мы должны пересмотреть что означает нестабильность частоты пульсара. В дальнейшем *внутренняя нестабильность* будет означать гипотетическую нестабильность после введения идеальных поправок к константе $\mathrm{d}P/\mathrm{d}t$, если бы она наблюдалась в системе координат, движущихся вместе с пульсаром. Именно эта внутренняя нестабильность рассматривается как меньшая, чем нестабильность атомных часов.

Мы можем далее представить возможность калибровки пульсарных часов с помощью атомных часов путем определения дрейфа их частоты на интервале времени длительностью I. Мы могли бы использовать пульсары для улучшения стабильности частоты TAI в ретроспективе на периодах выборки короче чем I путем формирования смешанной шкалы времени [8.18]. К сожалению, случайные неопределенности времени прибытия импульса требуют использования очень длительных интервалов I. Например, если атомная шкала имеет нестабильность 10^{-14} за один год, то с помощью пульсара, излучающего наиболее подходящие импульсы, датируемые в пределах 0,3 мкс, мы можем надеяться на улучшенную стабильность смешанной шкалы, если время выборки лежит между одним годом и I. Недавний прогресс в атом-

ных часах означал бы необходимость расширения *l* до нескольких десятилетий, что кажется концом этой притягательной идеи. *Пульсарное время* тем не менее рассматривается. Можно было бы сконструировать алгоритмы для наблюдений группы пульсаров и сличения их с атомными часами на больших интервалах времени, что могло бы привести к новым открытиям [8.18].

Изучался также и совершенно другой вклад миллисекундных пульсаров в измерения времени. Путем измерения эффекта Доплера измерение времени этих пульсаров дает положение Земли на ее орбите и, следовательно, ведет к новой мере эфемеридного времени. Здесь можно получить много большую точность отсчетов времени по сравнению с наблюдениями, основанными на определении (8.14) эфемеридного времени через долготу Солнца. Нестабильность частоты также улучшается на периодах выборки меньше шести месяцев. Но в плане долговременной нестабильности ничто не улучшается по сравнению с традиционными методами [8.18].

Третий метод измерения времени с помощью пульсаров заключается в мониторинге орбитального движения двойных пульсаров путем наблюдения эффекта Доплера. Теория такого движения исключительно сложна, так как гравитационное поле очень сильное и гравитационные волны производят значительные эффекты. Однако теория может быть развита. Остается вопрос, – может ли такая динамическая шкала, основанная на этом движении, конкурировать с атомным временем. В исследованиях [8.18] авторы приходят к следующему заключению. Открытие, наблюдение и интерпретация двойных пульсаров могли бы произойти раньше появления атомных часов (так как было бы достаточно иметь кварцевые часы для этих целей) и в этом случае время двойных пульсаров могло бы быть хорошим кандидатом для замены эфемеридного времени. Однако эта мера времени не могла бы занять место атомного времени, имеющего лучшую точность и стабильность.

Таким образом, кажется вероятным, что миллисекундные пульсары не дадут нам ни единицу времени, ни время, которое мы отсчитываем с наших часов. Тем не менее до недавнего времени нестабильность международного атомного времени в какой-то мере ограничивала качество научных результатов, которые могли бы получены путем измерения пульсарного времени. Действительно это было подтверждено ретроспективными вычислениями более точных и более стабильных шкал, чем шкала TAI, как это обсуждалось в разделе 7.4.3. Новые стандарты частоты могут компенсировать неидеальности атомной шкалы до ничтожности, но только время скажет об этом.

Измерение времени пульсаров является наиболее востребованным в применении их в атомных шкалах времени при рассмотрении точности и

долговременной стабильности (на интервалах шесть месяцев или более). Оно приводит к новым результатам в астрометрии, динамике солнечной системы и в самой теории гравитации. Измерение времени двойных пульсаров косвенно подтверждает существование гравитационных волн, предсказываемых теорией относительности, и таким образом усиливает доверие к этой теории [8.19, 8.20]. За эти исследования в 1993 году была присуждена нобелевская премия по физике Р.А.Хулсу и Дж.Х.Тейлору.

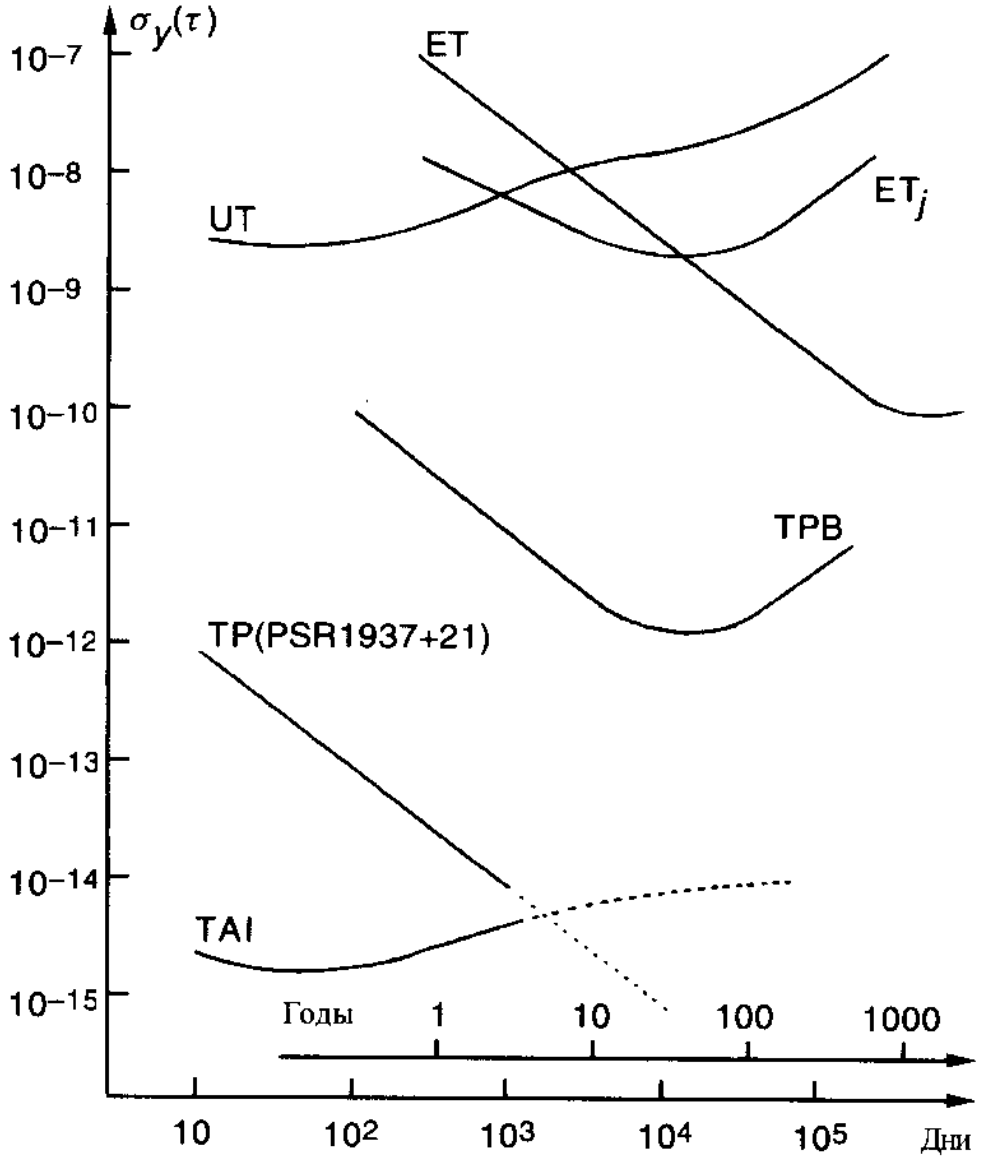


Рис. 8.7. Нестабильности астрономических шкал и TAI: ET и ETj - эфемеридные шкалы времени установленные наблюдениями Солнца и Луны соответственно; UT – всемирное время; TP(PSR1937+21) – время, основанное на вращении поименованного пульсара; TPB – время, основанное на орбитальном движении двойных пульсаров (оптимальный случай). Что касается TAI, то ее свойства спроектированы с 1998 года в будущее. Этот рисунок основан на деталях из [8.18] и [8.21].

Нестабильность частоты во времени, связанная с пульсарами, показана на рис. 8.7. Этот рисунок дает довольно качественный обзор проблемы. Для астрономических времен он повышается только на порядок величины. Для атомного времени он предполагает, что характеристики стандартов остаются на неизменном уровне.

Глава 9

Применение сверхточного времени и частоты

Измерения времени являются присущими везде, где вопрос касается частоты или времени часов. Обычные наручные часы, использующиеся для организации нашей повседневной деятельности, являются точным инструментом. Нестабильность кварцевых наручных часов может быть меньше 10^{-7} за один день. Некоторые часы автоматически подстраиваются по сигналам времени, связанным с лучшими цезиевыми стандартами. Таким образом, мы можем носить часы, точность частоты которых остается в пределах 10^{-14} при условии усреднения за несколько миллионов лет. В наших собственных домах мы окружены сигналами времени и частоты: таймерами, компьютерными часами, радиоприемниками, беспроволочными телефонами и т.д. Многие системы и службы зависят от времени: телекоммуникации, телевидение, энергоснабжение, определение местоположения.

Было бы интересно поупражняться, выбрав относительную нестабильность частоты, например 10^{-10} и определить, какая деятельность стала бы невозможной, если бы не было стандартов частоты, способных перешагнуть за этот предел. Ясно, что глобальная система определения положения и массовое ее применение не могли бы существовать. Не существовало бы также международное атомное время, так как астрономы удерживали бы контроль мировых шкал времени. Космическая геодезия была бы недостаточно точна, чтобы давать вклад в геофизику. Должны бы предприниматься некоторые меры на различных уровнях точности. К сожалению мы должны признать, что мы неспособны обеспечить исчерпывающее рассмотрение различных случаев. Мы можем только пригласить читателя проследить историческую эволюцию техники, в частности, космической техники и телекоммуникации, где измерение времени является ключевым элементом.

В качестве примера мы приведем ряд применений измерения времени с использованием лучших стандартов, либо самых точных, либо самых стабильных на различных временах выборки, или тех стандартов, которые обеспечивают хороший компромисс между точностью и стабильнос-

тью с одной стороны и надежностью с другой стороны. В действительности если мы хотим получить надежные и удобные инструменты, то в большинстве применений мы должны допустить незначительное их ухудшение по отношению к качеству лабораторных стандартов. Для таких применений производятся тысячами промышленные рубидиевые и цезиевые часы и иногда даже используются водородные мазеры.

Какова цель применения этих инструментов? Их основное применение сконцентрировано в трех различных направлениях. Первое, измерение интервалов времени и частот, например, в измерениях расстояния с помощью радаров или лазерных дальномеров. Второе возникает в системах, основанных на синхронизме часов, как в случае глобальной системы определения положения, или на равности частот (синхронизация). Последнее имеет место, если шкала времени используется для представления времени в некоторых физических теориях, например в моделировании орбит спутников и планет. Однако этот способ дробления измерений времени на разные категории не может служить основой для систематического представления предмета обсуждения, так как различные категории применения стремятся перекрыть друг друга. Поэтому мы проиллюстрируем наше обсуждение выбором некоторых практических и научных тем, в которых характерным является измерение времени.

9.1. Фундаментальные исследования

9.1.1. Метрология

Секунда является лучшей реализацией основных единиц международной системы единиц СИ. Поэтому совершенно естественно определять другие основные единицы в терминах секунды используя только физические законы и фундаментальные физические константы, которые вовлекаются в их определение. Если мы примем постулат, что такие константы действительно существуют (т.е. эйнштейновский принцип эквивалентности), то вся точность в реализации секунды, таким образом, в принципе, передается и на другие единицы. Таким шагом, сделанным в 1983 году, было определение метра, основанное на скорости света. Из семи основных единиц международной системы СИ (см. приложение 3), в настоящее время только три остаются независимыми, а именно: секунда, килограмм и кельвин. Например определение ампера основано на понятии силы в [м кг $с^{-2}$] и расстояния в [м] между двумя проводниками. Это определение, в конце концов выражается через секунду и килограмм.

Будет достигнут значительный прогресс, если килограмм удастся определить в терминах секунды. Нынешнее определение основано на массе конкретного объекта, цилиндра из платино-иридиевого сплава, содержащегося в международном бюро мер и весов (МБМВ). В этой единице

едва ли реализуется неопределенность лучше, чем 10^{-8}. Возможно удастся его заменить через электромагнитные или электростатические силы. Тогда все меры кроме термодинамических будут связаны с секундой.

Процесс переноса измерения величины на измерение времени происходит и в напряжении с использованием эффекта Джозефсона, и в сопротивлении с использованием квантового эффекта Холла. Однако эти квантовые эффекты не могут служить основой для нового определения единиц, используемых в электричестве, по крайней мере не сейчас, потому что в этом случае будет потеряна связь с механикой: механический ватт должен оставаться равным электрическому ватту. В настоящее время ватт используются для поддержания вольта и ома, а их значения все еще связанны с определением ампера.

Эффект Джозефсона происходит в двух сверхпроводниках, разделенных очень тонким изолирующим барьером. Если их подвергнуть СВЧ облучению с частотой ν, то на вольтамперной характеристике перехода появляются ступеньки, разделенные характерным значением напряжения V_n в приложенной разности потенциалов. Они теоретически определяются следующим выражением:

$$V_n = n\frac{h}{2e}\nu, \tag{9.1}$$

где n – целое число, h – постоянная Планка и e – заряд электрона.

Неопределенность в измерении отношения $h/2e$, приводимого в этом выражении, меньше, чем неопределенность в независимых экспериментальных определениях h и e по отдельности. Следуя международной кампании измерений в 1988 году по соглашению была принята величина, обозначаемая K_J и равная $h/2e$. Ее значение равно:

$$K_J = 483597{,}9\ ГГц/В^{-1}. \tag{9.2}$$

Это обеспечивает стабильную опору для разности потенциалов, привязанных к измерениям частоты.

Квантовый эффект Холла, известный также как эффект фон Клитцинга, происходит в случае, если тонкий полупроводниковый слой охлаждается до примерно 1 К и помещается в перпендикулярное слою магнитное поле порядка 15 Т. Напряжение Холла V_t, перпендикулярное направлению протекающего тока I, ведет себя таким образом, что сопротивление $R = V_t/I$ совпадает с теоретическим:

$$R = \frac{1}{n}\frac{h}{e^2}, \tag{9.3}$$

где n – опять целое число. Экспериментальное определение квантового сопротивления Холла дает для h/e^2 более точное значение, чем то, кото-

рое получается из независимых определений h и e, и в 1988 году по соглашению для этого отношения была принята величина R_K, а именно:

$$R_K = 25812{,}807\Omega . \tag{9.4}$$

С начала 1990 годов стала возможной посылка одиночных электронов один за другим через потенциальный барьер. Поэтому становится реальным создание электрического тока, величина которого будет определяться частотой сигнала, контролирующего открытие барьера. Затем станет возможной проверка взаимного соответствия методов определения электрических единиц, так как величина тока также связывается с частотой через квантовые эффекты Холла и Джозефсона.

Согласованные значения K_J и R_K подразумевают определение электрических единиц, которые не соответствуют строго официальной системе единиц СИ. Но они имеют преимущество в том, что они обеспечивают однородное представление вольта и ома с много большей воспроизводимостью, чем та, которая может быть получена из абсолютных измерений, основанных на определении ампера.

9.1.2. Измерение физических констант

Некоторые константы могут быть определены просто путем измерения частоты физического явления и их значения могут поэтому быть полученными с очень малой неточностью.

Примером является постоянная Ридберга. Она оценивается с неопределенностью в пределах $7{,}7\times10^{-12}$. Это результат выдающейся работы, выполненной группами из высшей технической школы в Париже (Франция) и в институте Макса Планка в Гаршинге (Германия). Используя методы спектроскопии сверхвысокого разрешения они измерили частоты ряда оптических переходов атома водорода. Полученное значение постоянной Ридберга было [9.1]:

$$R_\infty = 109737{,}31568550 см^{-1} . \tag{9.5}$$

Фактор Ланде g для электрона близкий к 2 был измерен в университете Вашингтона в Сиэтле с помощью удержания одиночного электрона в ловушке Пенинга. Это было получено путем наложения однородного магнитного поля на неоднородное электрическое поле. Значение g равно удвоенному отношению ларморовской частоты прецессии электрона к его циклотронной частоте. Полученное значение с неопределенностью 4×10^{-12} равно [9.2]:

$$\frac{g}{2} = 1{,}001159652188 . \tag{9.6}$$

Используя ту же технику для позитрона было получено то же самое значение g с неопределенностью 2×10^{-12}. Этот результат в настоящее время является наиболее точным подтверждением гипотезы, что античастицы имеют такие же свойства, как и соответствующие им частицы.

Путем сличения частот циклотронного резонанса протона и антипротона в пеннинговской ловушке было показано, что их инертные массы в пределах 4×10^{-8} равны [9.3]. С помощью аналогичной техники с неопределенностью до 3×10^{-9} было измерено отношение массы протона к массе электрона [9.4].

Как можно видеть из (9.6) значение фактора Ланде для электрона отличается от 2 на очень малую величину. Отклонение от этого значения зависит от эффектов квантовой электродинамики и может быть выражено через постоянную тонкой структуры α. Это одна из наиболее важных констант атомной и квантовой физики. Недавний теоретический прогресс сделал возможным использование измерения g чтобы в пределах 4×10^{-9}получить [9.5]

$$1/\alpha = 137{,}03599944\ . \tag{9.7}$$

9.1.3. Воздействие на атомную физику

(a) Проверка линейности квантовой механики

Основным уравнением квантовой механики является линейное уравнение Шредингера (6.27). Однако теоретически было показано, что может существовать слабая нелинейность [9.6]. Частоты переходов между двумя энергетическими уровнями зависели бы тогда от населенности этих уровней и вносили бы в эти переходы незначительный ангармонизм. Были выполнены экспериментальные проверки, одна с ионами бериллия Be^+ в радиочастотной ловушке [9.7] и другая с атомами водорода в водородном мазере [9.8]. В обоих случаях в пределах экспериментальных неопределенностей было найдено, что частоты переходов между соответствующими сверхтонкими подуровнями не зависят от их населенности. Так как благодаря очень малой нестабильности частоты используемых в проверке приборов эти неопределенности исключительно малы, мы можем заключить, что нелинейное обобщение уравнения Шредингера не является необходимым.

(b) Понимание свойств атомов и молекул

Этот вклад очень важен и разнообразен и детальное описание выходило бы за рамки этой книги.

Техника спектроскопии сверхвысокого разрешения, развитая и используемая в атомных или молекулярных стандартах частоты, сделала возможным измерение частот огромного числа переходов с очень малой нео-

пределенностью. Примерами являются частоты переходов атома водорода и его изотопов, изотопов рубидия с массой 85 и 87, и многих ионов с одним валентным электроном. Аналогично частоты атомных и молекулярных переходов в инфракрасном и видимом диапазонах известны теперь с неопределенностью до нескольких единиц на 10^{-12}.

Эти измерения, вместе с исследованиями эффектов от магнитного и электрического полей, значительно продвинули наши знания атомных констант, таких например, как ядерное и гиромагнитное отношение и межатомная связь в молекулах.

Лучшее понимание межатомных потенциалов было получено путем измерения сечения спин-обменных столкновений и сечения спин-обменных сдвигов частоты атома водорода в диапазоне 1–373 К. Производится изучение и способ сравнения теоретических и экспериментальных значений этих сечений для атомов рубидия и цезия с малыми относительными скоростями. Эти исследования обусловлены применением в часах охлаждаемых лазером атомов.

Наблюдался неожиданный вклад в сдвиг частоты сверхтонкого перехода атомов рубидия, когда они помещались в сильное (около 5 Т) магнитное поле [9.9]. Результат предполагает, что энергия магнитного взаимодействия между ядром и электронным облаком еще недостаточно хорошо известна и необходимы оценки дальнейших членов теории [9.10]. Эти члены могут также оказывать влияние на значения сдвигов частоты из-за квадратичного эффекта Зеемана, величина которого должна приниматься во внимание в атомных стандартах частоты и времени. Этот вопрос мог бы быть исследован в экспериментах с цезиевым пучковым стандартом [9.11]. Этот стандарт был бы необычным в том смысле, что в качестве опорного использовался бы переход

$$F = 3,\ m_F = -1 \leftrightarrow F = 4,\ m_F = -1.$$

Его частота имеет экстремум в зависимости от магнитного поля в области около 82 мТ. Точные измерения частоты этого перехода, близкой к 8,901 ГГц, обеспечили бы проверку формулы Брайта-Раби (6.4), содержащую все необходимые члены для вычисления зеемановского сдвига частоты.

9.1.4. Строение пространства-времени и гравитация

Большинство проверок строения пространства времени и гравитации используют измерения времени. Исчерпывающее обсуждение важности этих проверок дано в [9.12, 9.13]. Мы рассмотрим несколько примеров экспериментов, в которых атомные стандарты времени играли решающую роль. Некоторые проверки непосредственно основаны на стабильности или точности стандартов времени [9.14]. В других случаях эти стандарты являлись существенными дополнительными инструментами.

(а) Непосредственное использование атомных стандартов времени

Для проверки предсказаний общей теории относительности был организован ряд экспериментов с перевозимыми часами. В частности, в 1972 году Хафеле и Китинг из военно-морской обсерватории США организовали перевозку четырех цезиевых часов вокруг света на коммерческих авиалиниях. В поездках в западном направлении они обнаружили прибавление времени на 237 ± 7нс, а при поездках в восточном направлении потерю времени на 59 ± 10 нс. Теоретические расчеты давали прибавление на 275 ± 21 нс и потерю на 40 ± 23 нс соответственно. Ухудшенная неопределенность возникает из-за недостаточного знания траектории полетов часов [9.15]. Этот эксперимент рассматривался как проверка сдвига частоты из-за гравитации, однако, он был не очень информативен, потому что Паунд, Ребка и Снидер уже выполнили эту проверку с лучшей точностью с помощью эффекта Мессбаура [9.16, 9.17]. Однако вместе с другими аналогичными экспериментами это был убедительный аргумент в необходимости понимания измерений времени в рамках общей теории относительности.

Дальнейшее измерение гравитационного сдвига частоты было выполнено Вессо и Левиным в 1976 году [9.18]. Водородный мазер с помощью зонда был поднят на высоту 10 000 км. Во время полета его фаза непрерывно сличалась с фазой наземного водородного мазера с помощью трех линий связи, исключающих эффект Доплера. Было найдено, что в пределах экспериментальной неопределенности, а именно с относительной неопределенностью 7×10^{-5}, наблюдаемый сдвиг частоты согласуется с предсказаниями общей теории относительности. В таком же направлении планируются более точные эксперименты либо со спутниками с очень вытянутой орбитой вокруг Земли, либо с космическим зондом, летящим к Солнцу. Эти эксперименты интерпретируются как проверки эйнштейновского принципа эквивалентности (EEP) [9.12].

Путем сличения основанных на переходе в сверхтонкой структуре цезиевых часов, и расположенными рядом часов на магнии, основанных на переходе в тонкой структуре, Годоне и др. [9.19] оказались способны установить предел эволюции во времени произведения $g_p(m_e/m_p)$, где g_p – гиромагнитное отношение протона, m_e и m_p массы электрона и протона соответственно. Этот (относительный) предел равен $5{,}4\times10^{-13}$ в год. Объединив свои результаты с астрофизическими данными эти авторы показали, что относительные изменения во времени постоянной тонкой структуры меньше, чем $2{,}7\times10^{-13}$ в год. Во время этих же экспериментов часы подвергались воздействию годовых вариаций гравитационного потенциала Солнца из-за эксцентриситета орбиты Земли. Если бы EEP нарушал-

ся, то годовые вариации сказались бы на разности частот этих двух часов. Этого не произошло [9.20]. В более недавней работе сличения водородного мазера и стандарта частоты на ионах ртути Hg^+ установили верхний предел [9.21]

$$\frac{1}{\alpha}\left|\frac{d\alpha}{dt}\right| < 3{,}4 \times 10^{-14} \text{ в год.}$$

Более того, благодаря преимуществам отличной стабильности цезиевого и рубидиевого фонтанов, измерения на интервале 24 месяцев уменьшили верхний предел возможной вариации α. Теперь мы имеем [6.23]

$$\frac{1}{\alpha}\frac{d\alpha}{dt} < 3{,}1 \times 10^{-15} \text{ в год.}$$

Эти два последних результата основаны на том факте, что теоретическое выражение для сверхтонкого перехода в щелочных атомах (таких как H, Rb, Cs, Hg^+) содержит поправочный член, зависящий от произведения αZ, где Z – атомный номер элемента. Любые изменения α во времени привели бы к разному изменению частот рассматриваемых стандартов.

Один из постулатов теории относительности это постулат о постоянстве скорости света в локальных измерениях. Давайте мимоходом отметим, что это постоянство может быть проверено экспериментально только в том случае, если мы имеем независимые стандарты длины и времени. Для этих целей нам необходимы материальные стандарты длины, такие как линейки, плечи интерферометра или геодезические базисные линии. Постоянство c обычно проверяется двухсторонним распространением. Оно проверялось в известном эксперименте Майкельсона-Морли [9.22], который не обнаружил каких-либо эффектов из-за движения наблюдателя по отношению к гипотетическому эфиру. Однако мы также должны проверять, что скорость в прямом направлении равна скорости в обратном направлении, или обобщая – что скорость света действительно изотропна. Такая проверка была сделана в лаборатории реактивного движения (США) в 1990 году путем использования стабильности частоты двух водородных мазеров, разнесенных друг от друга на 21 км и сличавшихся односторонним методом через оптическое волокно [9.23]. В пределах экспериментальной неопределенности не было выявлено никакой анизотропии за время перемещения базисной линии двух мазеров благодаря вращению Земли, так что $|\Delta c|/c < 3{,}5 \times 10^{-7}$. Было предложено повторить этот эксперимент опираясь на спутниковый мазер [9.24]. Это уменьшило бы относительную неопределенность до $1{,}5 \times 10^{-9}$. В некоторых областях исследований долговременное сличение часов на борту спутников GPS с часами на наземных станциях привело к верхней границе $|\Delta c|/c < 4{,}9 \times 10^{-9}$

для всех направлений и к $|\Delta c|/c<1{,}6\times10^{-9}$ для экваториальных направлений.[9.25].

(b) Косвенное использование атомных стандартов времени

В других проверках теории относительности качество измерений времени используется для целей очень точного определения положения. Радиолокационная телеметрия может таким образом использоваться для измерения движения перигелия Меркурия более точно, чем это было возможно в оптической астрометрии с применением угловых измерений [9.26]. Относительная неточность в этой исторической проверке эйнштейновской теории была уменьшена до 10^{-3}. Лазерная локация лунных отражателей сделала возможным поиск возможного нарушения слабого принципа эквивалентности для массивных тел (эквивалентность инертной и гравитационной масс). Это мнимое нарушение носит название *эффект Нордтведта* [9.27]. Другие исследования были выполнены в поиске возможного изменения во времени гравитационной постоянной. Они основывались на точном измерении расстояний искусственных спутников, расположенных на орбите Меркурия.

Как показано в разделе 8.1.4a, с помощью РСДБ можно производить очень точные угловые измерения. Эти свойств были использованы для другой классической проверки общей теории относительности – отклонения света материей на более высоком уровне. Оказалось возможным измерение изгиба света от квазара при прохождении около Солнца, и даже около Юпитера [9.28].

Все эти эксперименты опираются на подробные модели орбитального движения. С этой целью используется атомная шкала TAI для того чтобы представить временные параметры теории так, как они определены в разделе 7.1. Так как наблюдения осуществляются на вращающейся Земле, то и параметры, описывающие это вращение необходимы с очень высокой точностью. Однако мы видели, что эти параметры определяются на основе техники времени и частоты. Совершенно ясно, что метрология времени касается такой работы исключительно сложным образом.

Одна из новых форм проверки заключается в измерении времени задержки сигналов от небесных объектов при их прохождении вблизи Солнца (*эффект Шапиро*). Время вовлекается в определение положения РСДБ, в орбитальные модели, в стабильность наземных и космических часов и в метод радиолокации. Этот метод задержки применяется к сигналам от миллисекундных пульсаров.

Мы также должны добавить сюда исследования, включающие в себя измерение времени миллисекундных пульсаров, обсуждавшееся в разделе 8.3.

(c) Предельные цели этих проверок

Со времени публикации эйнштейновской общей теории относительности было разработано много других теорий гравитации. В большинстве этих теорий гравитация рассматривается, как и в общей теории относительности, через проявление строения самого пространства-времени, описываемого математически метрикой. В дополнение к метрике некоторые другие теории предполагают существование полей, или даже заменяют метрику полями.

В слабых гравитационных полях и при малых скоростях по сравнению со скоростью света используется пост-ньютоновская метрика, например, (3.14) с низким уровнем приближения. Согласно работе Нордтведта [9.29] и Вилла [9.30] различные конкурирующие теории были закодированы путем введения десяти параметров, известных как PPN параметры, где PPN установлено для *параметризированного пост-ньютоновского* формализма. Эти параметры использовались для представления пост-ньютоновской метрики в единственном выражении, но они имели различные значения в различных теориях. Затем наблюдения приводили к оценкам их значений или значений одной из их комбинаций. Например, мы видели, что относительный гравитационный сдвиг частоты равен $\Delta U/c^2$ согласно общей теории относительности. Его выражение в PPN равно $(1+\alpha)\Delta U/c^2$. Параметр α (не путать с тонкой постоянной структуры) тогда является измеряемой величиной, которая согласно общей теории относительности равна нулю. Эксперимент Вессо и Левина показывает, что $|\alpha|<2\times10^{-4}$.

Вплоть до недавнего времени проверки, проводимые в солнечной системе для слабых полей и малых скоростей, в пределах экспериментальных неопределенностей давали PPN параметры, согласующиеся с общей теорией относительности. Это означает только то, что мы еще неспособны сделать различия между теориями, последствия которых могут отличаться при моделировании Вселенной и ее эволюции. Вот почему проверки должны включать в себя сильные гравитационные поля, такие как поля вблизи двойных пульсаров. Наблюдение и моделирование таких астрономических объектов как черные дыры также имеет ключевую важность для космологии.

Космология действительно является предельной целью этих исследований. Это величайшая мечта человека и не являются ли такие мечты славой человечества? В духе тех же исследований единая теория, объединяющая все микроскопические явления, а также все мировые элементарные частицы физики все еще не родилась, чтобы пожинать свои плоды. Лучшее понимание строения пространства-времени и гравитации было бы значительным успехом космологии.

Для тех, кого не убедили эти цели или для тех, кто вкладывает свое время и деньги в надежде на более быструю отдачу, мы можем добавить, что применение релятивистских теоретических рамок становится абсолютно необходимым. В качестве примера приведем сверхточное определение положения. В дополнение к его очевидному применению в транспорте оно является решающим для измерения крупномасштабных деформаций земной коры (относящееся к предсказанию землетрясений) и мониторинга уровня мирового океана, приводящего к климатическим изменениям. Поэтому озвучивание экспериментальных основ этих теорий имеет непреходящую важность.

9.2. Определение положения, геодезия и навигация

9.2.1. Принципы время частотных методов

Кроме многих применений в других областях, идея определения расстояний, основанная на принципах измерения двухстороннего времени распространения сигнала с помощью радио и лазерной техники широко используется в астрономии и геодезии. Далее возможен метод трехмерного определения положения при условии, что три цели доступны и наблюдатель не находится с ними в одной плоскости. Этот принцип управляет сверхточной геодезией путем лазерной локации искусственных спутников.

В спутниковой системе определения положения для наблюдателя более удобно использовать односторонние передачи со спутников. При этом также требуется много меньше излучаемой мощности. Однако неизбежное отсутствие синхронизации между передатчиком и приемником должно как-то компенсироваться. Во всемирных системах и в глобальной геодезии используются два решения. Одно из них основано на доплеровском сдвиге частот, передаваемых спутниками. В другом решении используется *псевдодальность*, которая определяется из разности времени приема, измеренного на приемных часах, и времени передачи, измеренного на спутниковых часах, что позволяет таким образом исключить ошибки синхронизации.

Для того чтобы представить простое описание доплеровского метода, давайте предположим, что частота ν_0 передатчика известна наблюдателю и наблюдатель фиксирован по отношению к земной поверхности. Здесь мы не будем рассматривать релятивистские эффекты, хотя на практике они должны быть приняты во внимание. Наблюдатель измеряет принимаемую частоту ν_{0R} и вычисляет разность $\Delta\nu = \nu_{0R} - \nu_0$. Затем эта разность интегрируется на временном интервале $\Delta t = t_2 - t_1$. После интегрирования мы получим число n периодов $P = \nu_0^{-1}$ потери (если $n<0$) или прироста (если $n>0$) на приеме в зависимости от того куда движется спут-

ник, от наблюдателя или к нему соответственно. Величина *cnP* есть просто разность между расстоянием до спутника на даты t_2 и t_1. Если на эти даты известно положение спутника в системе координат, установленной по отношению к Земле, то положение наблюдателя должно находиться на хорошо определенном гиперболоиде в этой системе координат. Так как спутник продолжает двигаться по орбите, то измерения повторяются и таким образом продолжают воспроизводиться гиперболоиды, имеющие только одну общую точку, которая должна являться положением наблюдателя. Однако на практике наблюдение одиночного спутника может привести к удовлетворительной точности только тогда, когда известна высота наблюдателя. Конечно, это описание сильно упрощено. Должны учитываться различные факторы, такие как возможная отстройка частоты, рефракция, собственное движение наблюдателя и т.д. Однако основные особенности системы достаточно ясны. Измерения с одним спутником дают «фиксированный» но не мгновенный способ, потому что спутник должен наблюдаться на части его орбиты. Полный набор направлений получается более быстро, если использовать низкоорбитальный спутник. Для глобального покрытия Земли низкоорбитальными спутниками требуются полярные орбиты.

Метод псевдодальности включает в себя несколько синхронизованных космических или наземных передатчиков, положение которых известно в земной системе координат. Если два передатчика одновременно излучают сигналы, то наблюдатель измеряет разность времени приема сигналов и затем находит разность расстояний до этих передатчиков. Тогда наблюдатель оказывается расположенным на известном гиперболоиде. Если передатчиков четыре, то положение наблюдателя определяется немедленно. Из-за геометрических ограничений необходимо, чтобы передатчики находились в космосе. Но для воздушной и морской навигации, когда известна высота, используется двумерная наземная система определения положения. В обоих случаях избыточная информация, получаемая из дополнительных передатчиков, обеспечивает точность и безопасность. Кроме определения положения в качестве подарка получаются сличения часов наблюдателя с часами системы. Как мы видели в главе 7, этот побочный продукт имеет очень важное значение для целей распространения времени.

Стоит отметить тесное соотношение между доплеровским методом и методом псевдодальности, хотя на практике это соотношение может укрываться за способом математического истолкования данных. Такие методы, применяемые с использованием искусственных спутников, опираются на точное описание орбит спутников в наземной системе координат. Это достигается путем создания сети следящих станций с известными положениями. Измерения немедленно передаются на центр управления, где они обрабатываются за самые краткие по возможности сроки для того,

чтобы передать орбитальные модели обратно на спутники и хранить их как можно более обновленными, готовыми для передачи потребителю.

Мы уже упоминали глобальную систему определения положения GPS, основанную на измерениях псевдодальности, потому что она может использоваться для сличения часов чтобы распространять время (главы 5 и 7) и измерять вращение Земли (глава 8). Можно отметить настоящий бум в росте числа пользователей GPS в ее стандартном назначении, т.е. для определения положения. В следующем разделе мы надеемся удовлетворить любопытство читателя, представив описание системы немного более детально, хотя этой теме уже посвящено много специализированных работ. В частности, ссылка [9.31] объясняет, как работает система и объясняет различные способы ее применения в геодезии. Далее мы обрисуем другие системы.

9.2.2. Глобальная система определения положения

(а) Спутники и их сообщения

Система Навстар (навигационное спутниковое время и дальность) сегодня более известная акронимом GPS, зародилась из исследований, начатых в 1965 году. Ее развертывание продолжалось до 1995 года. Это был также и исследовательский период, давший некоторые представления о размерах предприятия. GPS появилось в результате слияний проектов военно-морских и военно-воздушных сил США (проекты Тимэйшн и 621B соответственно)

В системе используются двадцать четыре спутника с почти круговыми орбитами на высоте около 20 000 км, распределенные в шести орбитальных плоскостях на интервалах 60° долготы и наклонением 55° к экватору. Орбитальный период этих спутников составляет половину звездного периода вращения Земли. (Последний равен 23 часа 56 мин. и соответствует возрастанию угла A_S, определенного в разделе 8.1.2, на 360 °). Следовательно, неподвижный наблюдатель на Земле каждые 23 часа 56 мин видит один и тот же самый спутник, двигающийся по небу по одной и той же траектории с тем же азимутом и восхождением. В конфигурации, подходящей для определения положения наблюдателя, с любой точки земной поверхности (кроме нескольких кратковременных исключений) всегда видны по крайней мере четыре спутника.

Первые одиннадцать спутников системы, известные как Block I, имели с самого начала рубидиевые стандарты, затем цезиевые стандарты, а также, в целях безопасности, резервные кварцевые часы. Спутники Block II запускались с 1985 года с двумя цезиевыми часами, двумя рубидиевыми часами и дополнительным кварцевыми часами. Они имели срок службы около восьми лет.

Основной принцип, разработанный до GPS, это синхронизация бортовых часов. На практике часы физически не синхронизированы, но спутники передают данные, которые можно использовать для расчета разности показаний между часами спутника и базовым временем системы, а именно временем системы GPS с неопределенностью в несколько наносекунд.

Сигналы GPS передаются на двух несущих частотах, одинаковых для всех спутников системы. Это частоты L1 на 1575,42 МГц и L2 на 1227,60 МГц. Цифровая информация, которую они передают, называемая навигационное сообщение, осуществляется переменой фазы несущей на 180°. Она состоит из следующих компонент:

- псевдо-случайный код C/A,
- псевдо-случайный код P,
- сообщение, несущее числовые данные.

Псевдослучайный код взаимно ортогональный и разный для разных спутников. Автокорреляционная функция каждого кода равна почти нулю, за исключением нулевой задержки. Это объясняет, каким образом приемники различают различные спутники, генерирующие свои коды, и как они измеряют время приема сигнала с каждого спутника (максимальная корреляция) с помощью своих внутренних часов.

C/A код (*Coarse Acquisition – грубый код* или *Clear Access – код общего пользования*) передается только на частоте L1, сдвигаемой по фазе на 90°. Он повторяется с периодом в 1 мс и включает в себя 1023-битовую последовательность.

Код P (*Precise – точный* или *Protected – защищенный*) передается и н а частоте L1 и на частоте L2. Его период повторения равен 267 дням, но однонедельный сегмент передается и изменяется каждую неделю. Он может быть зашифрован для исключения несанкционированного доступа потребителя.

Навигационное сообщение передается с низкой скоростью модуляции 50 бит $с^{-1}$. Оно состоит из 1500-бит *кадра* (продолжающегося 30 с), который подразделяется на пять *субкадров.* Содержание субкадров от 1 до 3 повторяется из кадра в кадр (если их содержание не было обновлено) и представляет информацию о передающем спутнике, т.е. его состоянии, элементах орбиты и коэффициентов для вычисления поправок на часы, используемые в данное время. Субкадры 4 и 5 обеспечивают информацию о системе в целом, т.е. модели для вычисления ионосферных задержек, разность между UTC и GPS временем, грубые эфемериды и состояние других спутников системы, информацию о кодах. Так как вся эта информация слишком объемная, чтобы поместить ее всю в двух субкадрах, то она передается по субкадрам 4 и 5 из двадцати пяти последова-

тельных кадров. Полный набор информации может быть получен таким образом только путем слежения за спутником в течение 12,5 минут. Это объясняет, почему слежение растягивается на 13 минутный период если на кону стоит применение очень точного измерения времени (см. раздел 5.6.3). На каждой частоте и L1 и L2 передаваемый сигнал является результатом модуляции соответствующей коду и сообщению. Отметим, что во время передачи одного бита сообщения, т.е. 20 мс, C/A код повторяется двадцать раз.

(b) Стандартная работа системы

Под «стандартной работой» мы понимаем применение системы для простого определения положения в реальном времени с использованием обычных GPS приемников. Мы не будем обсуждать множество других способов применение этих приемников.

Определение спутниковой орбиты и вычисление поправок часов, выполняемое станцией с ведущими часами, основано на слежении за спутниками с помощью пяти контрольных станций. Геометрические координаты этих станций известны с неопределенностью в несколько десятков сантиметров по отношению к опорной системе координат, которая называется WGS84.Было предпринято много попыток согласовать систему WGS84 с системой IERS, известной как ITRF и упоминавшейся в разделе 8.1.6. Сейчас отклонения систем друг от друга составляют порядка одного дециметра. Таким образом орбитальные параметры, вычисляемые и передаваемые в WGS84 системе, на практике являются и параметрами в ITRF и то же самое можно сказать о координатах наблюдателя.[1]

Время GPS согласовано с UTC (USNO) в пределах 100 нс по модулю одной секунды. Последняя следует за UTC в пределах 100 нс.

Орбиты описываются посредством шести параметров, похожих на параметры Кеплера, а именно: главная полуось, эксцентриситет, наклонение к экватору, прямое восхождение узла, аргумент перигея и средняя аномалия. Однако орбиты искусственных спутников возмущаются неоднородностями гравитационного поля Земли, давлением солнечного излучения, лунно-солнечным приливным потенциалом и остаточной атмосферой. Таким образом согласно методу, обычно принятому у астрономов, орбитальные параметры рассматриваются как переменные. Используя их значения в данный момент мы получим орбиты очень близкие к истиной орбите в диапазоне соседствующих времен. Передаются

---

[1] Национальные картографии и морские карты были установлены в региональных системах с помощью традиционных методов геодезии. Предполагается преобразование их в ITRF, хотя это еще не достигнуто. Расхождения между ними WGS84 и ITRF могут достигать несколько сотен метров. Это должно приниматься во внимание и в некоторые приемники необходимо вносить необходимые поправки.

их изменения скорости, и параметры должны меняться каждый час или около этого.

Неопределенность положения движущегося наблюдателя может быть снижена до 10 м при приеме P кода на обеих частотах. При приеме C/A кода неопределенность составляет около 30 м. До мая 2000 года, когда применялся режим SA (*Selective Availability – селективный доступ*), для не имеющего доступа потребителя неопределенность составляла около 100 м.

(c) Другие применения

GPS может использоваться в дифференциальном режиме для увеличения точности в применении к локальному позиционированию. Положение движущегося наблюдателя соотносится с точностью известных положений близлежащих неподвижных станций, одновременно следящих за тем же самым спутником. При этом значительно уменьшаются некоторые ошибки, включая ошибки из-за эфемерид, шума спутниковых часов, рефракции и SA, если он используется. Неопределенности могут быть уменьшены до нескольких метров в реальном времени. Дифференциальное определение положения используется также для определения относительного положения приемников времени и улучшения таким путем точности сличений часов. Так как эти приемники неподвижны, то может быть выполнено усреднение по времени и неопределенности их относительных координат могут быть уменьшены до дециметров [9.32].

Для высокоточных применений, таких как сантиметровая геодезия и изучение вращения Земли, используется слежение за фазой несущей частоты. Модуляция снимается путем умножения принимаемого сигнала самого на себя. Так как модуляция достигается путем обращения фазы, то получается чистая удвоенная частота. В этой области изучение координируется международной службой, известной как *международная GPS служба для геодинамики* (IGS). Эта служба устанавливает точные эфемериды GPS спутников, доступные в течение одной или двух недель после их установления. Они полезны в частности для уменьшения неопределенности в сличениях времени для ведения шкалы TAI.

(d) Гражданское дополнение к GPS

В вопросах применения для гражданских целей система GPS критиковалась за отсутствие точности, вызываемое включением SA, а также за отсутствие предупреждений в реальном времени о неисправностях с одним из спутников системы. Так как число видимых спутников только-только превышает минимально необходимое число четыре, то такая неисправность может привести к неправильному определению местоположения с

пагубными последствиями. Поэтому возникает вопрос, а не надо ли установить параллельную GPS гражданскую систему.

Сложность системы GPS обусловлена в частности ее военным назначением. Были сконструированы системы, работающие на тех же линиях, но без часов на борту спутников. Время обеспечивается единственными ведущими часами на земле и передается на спутниковые преобразователи. Однако этот тип систем не удовлетворяет нуждам вооруженных сил для полностью автономной работы, т.е. работы, для которой нет необходимости передачи данных с земных станций. Такая автономия, хотя и ухудшает качество, может оказаться жизненно необходимой во время конфликтов.

В настоящее время в разработке находятся две линии. Первая включает в себя совместное использование GPS и ее российского аналога ГЛОНАСС, который мы кратко обсудим позже. Число доступных спутников может возрасти до сорока восьми с десятком спутников или около того, видимых одновременно. Результирующая избыточность означает, что «больные» спутники могут быть обнаружены и исключены. Уже разработаны смешанные приемники GPS/ГЛОНАСС. Другое направление включает в себя установку гражданского дополнения к GPS, в котором передатчики будут установлены на геостационарных спутниках и будут передавать коды GPS [9.33]. В 2002 году европейское сообщество решило создать систему GALILEO, работающую на тех же принципах что и GPS и ГЛОНАСС. Система будет включать в себя 30 спутников и будет совместимой с GPS. Она обеспечит аналогичные возможности для сличения времени и сверхточных применений.

9.2.3. Другие системы определения положения

В СССР, а в последующем в России параллельно GPS была разработана система ГЛОНАСС. Она работает по тем же принципам, хотя спутники движутся по незначительно отличающимся орбитам на высоте около 20 000 км. Неподвижный наблюдатель также видит спутник, движущийся по одной и той же кажущейся траектории каждые 23 часа и 56 минут, но каждый день это новый спутник и спутники имеют однонедельный период. Спутники узнаются по частотам. Орбитальные модели передаются в координатной форме. В дополнение базовая шкала времени, известная как время ГЛОНАСС, не столь тесно совпадает с UTC. Геодезическая система координат также другая. Это приводит к сложной, но все же возможной задаче создания смешанных GPS/ГЛОНАСС приемников.

Прежде чем появились GPS и ГЛОНАСС, с 1964 года для морской навигации широко применялась доплеровская система Навсат. В этой системе шесть спутников с полярными орбитами на высоте 1100 км излучали частоты 150 МГц и 400 МГц, а также синхронизационные

маркеры. Сообщения об орбитальных моделях передавались фазовой модуляцией.

Другая доплеровская система, известная как DORIS, была разработана французским национальным центром изучения космоса (CNES) для определения положения с высокой точностью [9.34]. В ней используется противоположный Навсату метод в том смысле, что частоты излучаются наземными радиомаяками, положение которых должно быть определено, и принимаются спутниками. И радиомаяки, и спутники оборудованы сверхстабильными кварцевыми генераторами. Частоты близки к 400 МГц и 2 ГГц. Нестабильность частоты генераторов на радиомаяках и на спутниках должна быть менее 5×10^{-13} за время счета (интервал Δt определен в разделе 9.2.1) и 10^{-12} за время пролета (например за 1000 с). Бортовые частоты определяются путем сличений с наземными ведущими часами, настроенными на TAI. Неопределенность должна быть меньше 10 мкс. Основная сеть от сорока до пятидесяти радиомаяков на Земле служит для определения орбит спутников. Система была испытана при слежении с Земли за спутником Spot-2. Она затем была принята как одна из систем определения положения для океанографических спутников TOPEX/POSEIDON, которые будут обсуждаться ниже. В последнем случае было показано, что неопределенность радиальной компоненты положения спутника составляет около 3 см. Для двух других координат неопределенность составляет около 20 см. DORIS также определяет местоположение радиомаяков с неточностью до 2 см. Эта точность позволила системе DORIS выделиться из богатства средств, используемых IERS для конструирования ITRF и измерения вращения Земли.

Система PRARE (*Precise Range and Rate Range Equipment – оборудование для точного определения дальности и скорости)* была разработана в Германии для точного определения орбит спутников [9.35]. Она также включает в себя космический бортовой сегмент, сеть следящих станций и обрабатывающий центр. В ней объединяется метод псевдодальности и доплеровского измерения сигналов, передаваемых спутниками на частотах 2 ГГц и 8 ГГц. Эта сеть совместно с системой GPS используется европейским космическим агентством (ESA) в программе спутникового исследования Земли (ERS). Она имеет точности того же порядка, что и DORIS.

9.3. Интерферометрия со сверхдлинной базой (РСДБ)

В разделе 8.1.4 мы обсуждали роль, которую играет интерферометрия со сверхдлинной базой (РСДБ) в измерениях всемирного времени. Она применяется для изучения вращения Земли и конструирования небесных и земных координатных систем. Угловые разнесения между источниками должны быть большими и разнообразными для того чтобы выделить ве-

личины, которые необходимо определить. Мы уже видели, что этот способ измерений использует преимущество стабильности частоты водородных мазеров.

РСДБ сначала была создана для изучения строения небесных источников с малыми угловыми размерами, излучающих в радиодиапазоне спектра. Разрешающая способность улучшается при приеме более высоких частот и при большем разнесении антенн. При использовании антенн земного базирования в сантиметровом и дециметровом диапазонах достигается угловое разрешение 0,001″. Использование миллиметрового диапазона длин волн находится в состоянии разработки и угловое разрешение достигает 0,0001″. РСДБ проекты использования антенн наземного и космического базирования должны бы обеспечить угловое разрешение до нескольких десятых угловой микросекунды [9.36].

РСДБ используется также различными способами для определения положения небесных источников по отношению к компактным внегалактическим источникам, образующим первичную систему координат. В этом случае мишенями являются соседние источники и типичное угловое разрешение измерений должно составлять несколько градусов.

Эта техника может быть использована разными способами, для того чтобы расширить небесную систему координат ICRF до оптического диапазона (см. раздел 3.3.2b). В частности, космический астрометрический проект HIPPARCOS дает нам каталог угловых положений и собственных движений около 100 000 звезд с внутренней сходимостью на уровне 0,001" и 0,001″/год соответственно, но он неспособен установить начальное вращение и скорость вращения всего каталога в целом [9.37]. Каталог HIPPARCOS должен быть согласован с ICRF. Одна из технических возможностей, используемых для этих целей, это РСДБ наблюдение нескольких десятков радиоисточников, принадлежащих каталогу HIPPARCOS. Путем объединения различных методов была получена оценка, что согласование достигается в пределах 0,0005″ со средними данными наблюдений HIPPARCOS (1991 год). Однако они ухудшаются со скоростью около 0,00025" в год из-за неопределенностей в собственном движении звезд [9.38].

Дифференциальные РСДБ часто используются для слежения за космическими зондами для того, чтобы определить их положение относительно ICRF и таким образом определить положение других искусственных спутников, расположенных на орбитах вокруг других планет.

9.4. Программа TOPEX-POSEIDON

В качестве примера мы обсудим многие аспекты времени и частоты в программе TOPEX-POSEIDON.

Измерение вертикального расстояния между искусственным спутником Земли и поверхностью моря с помощью бортового альтиметра впер-

вые было испытано между 1975 и 1978 годами на американском спутнике Geos-3. Оно было продолжено со спутником Seasat в 1978 году. Хотя после нескольких месяцев спутник вышел из строя, все же были получены доказательства того, что эти измерения полезны для изучения океанов и их взаимодействия с атмосферой, для климатологии и внутренней геофизики. Geosat подтвердил эти выводы в 1985–1989 годах. Затем были разработаны более амбициозные программы в европейском космическом агентстве (ESA) в ERS (спутниковое изучение Земли), а в сотрудничестве между NASA и CNES (Франция) была разработана программа TOPEX-POSEIDON. Спутник ERS 1 был запущен в 1991 году и за ним последовал в 1995 году спутник ERS 2. Спутник по программе TOPEX-POSEIDON был запущен 10 августа 1992 года.

Радиолокационные пучки пересекают достаточно большие области поверхности моря для усреднения эффектов от морских волн. Высотные измерения включают в себя различные элементы, показанные на рис. 9.1.

- *Опорный эллипсоид* – это математическая поверхность, которая примерно соответствует поверхности моря и принимается по соглашению. Это устраняет необходимость опоры на геоцентр.
- *Геоид* – это поверхность постоянного гравитационного потенциала, после введения поправок на периодические члены, производимые приливным потенциалом. Он должен бы совпадать со средней поверхностью моря, если бы воды океанов были бы однородны и неподвижны.
- *Истинная поверхность моря* – определяется как поверхность, на которой усреднены эффекты от морских волн.
- *Орбита спутника* – устанавливается в геоцентрической системе координат. Ее высота вычисляется над опорным эллипсоидом.

Высотные измерения дают высоту спутника над истинной поверхностью моря. Если мы знаем орбиту спутника, то мы получим *высоту поверхности моря* над опорным эллипсоидом.

Два элемента имеют большую научную важность, это высота h_2 геоида над опорным эллипсоидом, которая может быть положительной или отрицательной и достигать значений в сотни метров или около того, и высота h_1 поверхности моря над геоидом, которая тоже может быть положительной или отрицательной и достигать значения около одного метра. Мы сразу же отметим, что измерения дают сумму этих двух элементов. Эта неоднозначность удаляется через их изменчивость. В течение нескольких лет геоид может приниматься как неизменный, и получаются данные об изменениях поверхности океана. Если мы предположим, что океанские эффекты имеют нулевое среднее время усреднения, то средняя поверхность представляет пространственную неоднородность геоида. В долгосрочных наблюдениях здесь остаются трудности. Например, таяние

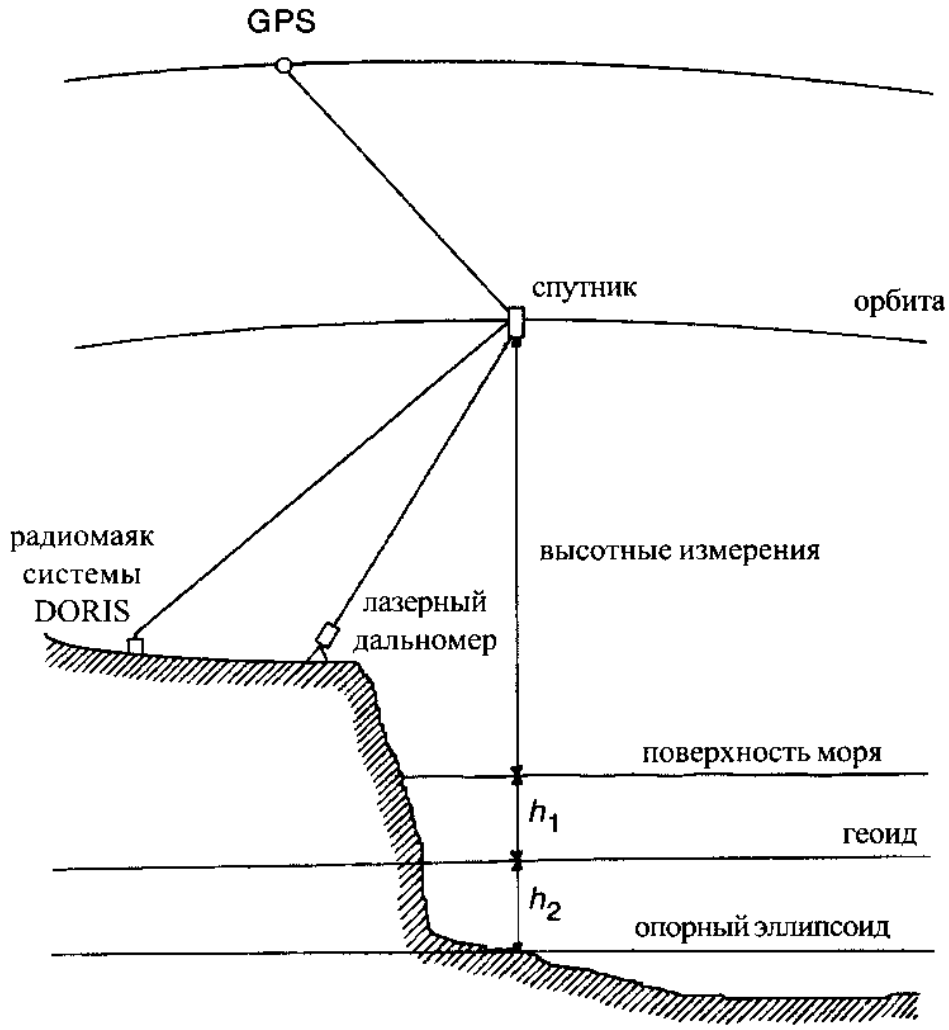


Рис. 9.1. Геометрические элементы измерения высоты над океаном

ледяных полярных шапок воздействует и на уровень моря и на форму геоида.

На форму отклика радиолокатора влияют условия на море. Кроме данных о высоте путем анализа отклика радиолокатора можно получить информацию о морских волнах, включая их высоту и направление движения.

На уровень моря влияют разности температур, соленость, атмосферные и динамические эффекты, такие как центробежные силы и силы Кориолиса. Поэтому легко видеть, что их изучение может обнаруживать течения, водовороты и перенос тепловой энергии, приводящий к климатическим изменениям. Однако величины таких эффектов не превышают нескольких дециметров. Для их изучения их необходимо измерять с неопределенностью в несколько сантиметров. Интересно перечислить научные темы, упомянутые в объявлении на предложения по исследованиям, объявленные CNES, пока TOPEX-POSEIDON находился в стадии подготовки [9.39]:

- провести изучение постоянных и изменчивых кругооборотов в океанах и их взаимодействие с атмосферой в глобальном и локальном масштабах и улучшить наше понимание климатической роли океана;
- разработать улучшенные объяснения океанических круговоротов в региональных масштабах за времена в несколько месяцев;
- получить улучшенное понимание переноса тепла в океане;
- провести изучение приливных моделей;
- изучить статистику океанских изменений;
- улучшить знания гравитационных полей и связанного с ними геоида;
- изучить физику волн.

Спутник TOPEX-POSEIDON имеет круговую орбиту с высотой около 1300 км и наклоном к экватору 66°. Поэтому он не охватывает полярные области. След его орбиты, т.е. вертикальная проекция орбиты на земную поверхность, повторяется каждые 10 дней и расстояние между следами на экваторе равно 316 км и уменьшается с повышением широты. Программа, планируемая раньше до 1997 года, все еще продолжалась в 2002 году. Геофизическая оценка программы и научные результаты ясно показывают, насколько она была успешной. Это засвидетельствовано одной тысячью двумястами страницами публикаций, посвященных этой теме в томах 99 и 100 *Журнала по геофизическим исследованиям* 1994 и 1995 годов и последующими публикациями во многих других научных журналах. В частности, неопределенность высоты поверхности моря над эллипсоидом составляет за один проход около 5 см. Она включает в себя неопределенность орбиты и неопределенность высотомерных измерений почти в равных пропорциях. Если применить усреднение за один месяц, то неопределенность может быть уменьшена до 1 см. Было также показано, что изменения среднего уровня моря могут быть выполнены с разрешением до 1 мм год$^{-1}$. Эти достижения вместе с успехом программ ERS 1 и 2 послужили толчком для создания проектов с применением следующих высотомерных спутников. (ERS 1 и 2 являются дополняющими друг друга, потому что они имеют различающиеся следы и повторяются разными способами.) В таких проектах изучались бы долговременные эффекты, включая долговременные изменения уровня моря, а также обеспечивался бы непрерывный мониторинг. 7-го декабря 2001 года NASA в сотрудничестве с CNES запустили первый, из планируемой серии, высотомерный спутник Jason-1.

Оставив в стороне научные результаты, которые не являются главной целью данного рассмотрения, мы обсудим теперь роль, которую играли время и частота в программе TOPEX-POSEIDON.

Альтиметр это ни что иное, как радиолокатор, измеряющий время пролета сигнала. Неточность часов альтиметра должна быть меньше чем

10^{-10}, если ее вклад должен быть пренебрежимо мал. Нестабильность его частоты за время пролета, равное примерно 10 мс, должна иметь примерно то же значение. Измеренные данные должны выражаться в TAI с точностью до нескольких десятков микросекунд.

Измерения времени очень важны при определении орбиты спутника. Точность, требуемая для TOPEX-POSEIDON (неопределенность несколько см в радиальном направлении) оставила далеко позади все, что ранее было достигнуто. Использовались три основные системы слежения, а именно: лазерные дальномеры с уголковыми отражателями на спутнике, бортовой приемник GPS и система DORIS. Эти время-частотные системы обсуждались в разделе 9.2. Для моделирования орбиты необходимы знания гравитационного потенциала Земли. Модель основана на слежении за очень большим числом спутников с совершенно разными орбитами. Такое слежение также включает в себя время-частотную технику определения положения. Более того, в настоящем случае модель была улучшена путем слежения за самим спутником TOPEX-POSEIDON. Такой анализ требует точного определения параметров вращения Земли, потому что орбиты описываются в невращающейся системе координат, в то время как положение спутника должно приводиться в опорной системе, сопровождающей Землю в ее вращении. Этот мониторинг вращения сейчас решается исключительно техникой и средствами времени и частоты (см. раздел 8.1.4). Параметры вращения могут быть определены из слежения системы DORIS за TOPEX-POSEIDON, но полезно объединить результат с результатами от других методов. Самые стабильные часы, которые могли бы быть построены, требуются для РСДБ методов.

Наконец разными способами в измерения вовлекается TAI. Она является основой для синхронизации (до 1 микросекунды на следящих станциях), опорой для точности по частоте и в применении временного аргумента в динамических теориях, используемых для орбитальных вычислений.

9.5. Телекоммуникация

Телекоммуникационные сети имеют структуру дерева, в которой окончания ветвей являются пользовательскими терминалами. Несколько иерархических уровней содержат узлы, в которых сообщения переключаются с одного канала на другой. В режиме аналоговых передач несколько сообщений передаются вместе через один и тот же коммуникационный канал путем применения SSB модуляции (модуляция с одной боковой полосой) без передачи несущей. Для того чтобы правильно воспроизвести сообщение, частоты генераторов, которыми оборудованы модулирующие/демодулирующие центры, должны быть синтонизированы почти до 10^{-8} [9.40].

Цифровая передача голоса является значительным совершенствованием. Звуковые частоты от 300 Гц до 3600 Гц усредняются со скоростью 8000 раз в секунду и кодируются в восьмибитовые слова. А потом оказывается возможным смешать сообщения вместе и передать их одновременно через один и тот же канал до 30 сообщений за одно и то же время посредством процесса уплотнения. Для того чтобы разделить каналы необходим прием опорной частоты. Однако между частотами уплотнения и разделения каналов допускается достаточно существенный сдвиг, потому что временной сдвиг не слышен.

Для передач цифровых данных, в которых потери информации из-за сдвига времени недопустимы, возникает необходимость в хорошей синхронизации, т.е. в хорошей настройке частоты по всей сети. Эта проблема привела к нынешнему развитию *синхронной цифровой иерархии* (SDH).

Здесь частоты в узлах и на терминалах должны быть настроены в пределах 10^{-11} на дневном усреднении. В дополнение международный союз электросвязи ставит условием, чтобы эти частоты были фиксированы по отношениям к частоте UTC и, следовательно, к частоте TAI в пределах 10^{-11} в нормализованном значении [9.41].

В контексте данной работы нужды телекоммуникаций могут показаться умеренными. Более изощренное решение требуется для поиска путей понижения стоимости. Тем не менее, цезиевые часы являются необходимыми на ведущем узле сети. Узлы других уровней и пользовательские терминалы оборудуются кварцевыми генераторами, которые автоматически подстраиваются по самой сети. GPS может дать вклад в контроль частоты опорного узла или узлов высокого уровня.

Несмотря на свое название SDH (синхронная цифровая иерархия) основана на синтонизации (т.е. к достижению совпадения частот). Использование синхронизации в собственном смысле иногда пропагандируется за те преимущества, которые оно могло бы принести. Иногда даже были некоторые разговоры о будущей необходимости доступа в реальном времени к UTC в пределах 100 нс для нужд телекоммуникации [9.42]. Это без сомнения проект, который довольно далеко удален в будущее, о котором мы не способны получить дальнейшие подробности.

Глава 10

Навигационная система GPS – генератор координатно-временного поля

За последние 10 – 15 лет в нашу жизнь буквально ворвались несколько технических новшеств, сразу получивших глобальное распространение и самое широкое применение, причем, не только для сложных научных и технологических целей, но и активно востребованных на бытовом уровне: персональные компьютеры, интернет, сотовые телефоны... В этот ряд вполне по праву можно включить Глобальные Навигационные Спутниковые Системы (ГНСС): американскую GPS и российскую ГЛОНАСС. Глобальность систем ГНСС обеспечивается функционированием на орбитах набора видимых из любой точки Земли спутников, непрерывно передающих высокоточные измерительные сигналы. Тем самым вокруг нашей планеты создано как бы *информационное координатно-временное поле*, находясь в котором пользователь с помощью специального приемника может черпать из него данные о своем положении в пространстве и времени.

Разрабатывавшаяся в первую очередь для военного применения (военно-морская и воздушная навигация, управление ракетами), система GPS вскоре стала активно использоваться и в различных областях гражданской сферы научно-технической и хозяйственной деятельности. Этому в значительной мере способствовало быстрое создание пользовательской приемной аппаратуры GPS самого широкого спектра – от прецизионных двухчастотных фазовых приемников, обеспечивающих миллиметровую точность координатных определений и наносекундную точность временной привязки, до массовых персональных портативных приемников размером с сотовый телефон, которые можно, например, вмонтировать в противоугонное устройство автомобиля. Вполне понятен, поэтому, интерес специалистов различного профиля, да и населения вообще, к устройству и принципам работы системы GPS, ее возможностям, техническим и потребительским свойствам.

С точки зрения тематики этой книги система GPS представляет двоякий интерес. С одной стороны, функционирование GPS основано на устройствах и методах измерений, применяемых в службе времени, которая в этом смысле является одним из ключевых элементов научно-технологической базы GPS. Но с другой стороны, и сама служба времени использует средства GPS для решения своих задач, получая возможность распространения сигналов точного времени в глобальных масштабах и выполнения непрерывных высокоточных сличений эталонов времени различных лабораторий мира.

Учитывая интерес к этой теме, мы посчитали полезным дать несколько более развернутое описание принципов действия и возможностей применения GPS в дополнение к тем сведениям, что содержатся в основных разделах книги.

10.1. Физические и геометрические принципы действия GPS

Хотя глобальные навигационные спутниковые системы являются чрезвычайно сложными и высокоорганизованными техническими системами, идеи и принципы, на которых основано их действие, весьма просты: это – определение местоположения объекта путем измерения расстояния до него от исходных точек, координаты которых известны. Сложность же реализации этих идей обусловлена стремлением сделать систему доступной в любое время на всей Земле и в окружающем пространстве *(глобальной)*. Для этого в качестве исходных точек выбраны искусственные спутники Земли, излучающие дальномерные радиосигналы, которые пользователь принимает на специальный приемник. Но спутники движутся по своим орбитам, значит система должна предоставлять пользователю информацию о координатах ИСЗ на любой момент выполнения измерений. В этом разделе мы последовательно рассмотрим, как решается эта задача, начав с пространственных геометрических сетей и локационных методов измерения расстояний.

10.1.1. Геодезические построения

В основе применяемого в ГНСС GPS метода определения местоположения лежит так называемая *линейная геодезическая засечка*. Ее суть сводится к известной со школы геометрической задаче: найти на плоскости положение точки P_0, если известны положения двух других точек P_1 и P_2 и расстояния от них ρ_1 и ρ_2 до точки P_0 (рис. 10.1). Искомая точка P_0, очевидно, принадлежит одновременно двум окружностям с радиусами ρ_1 и ρ_2, описанным из центров P_1 и P_2, т.е. является одной из двух точек пересечения этих окружностей. В аналитическом представлении эта задача выражается в виде системы двух уравнений:

$$\rho_1 = \sqrt{(X_1 - X_0)^2 + (Y_1 - Y_0)^2}$$
$$\rho_2 = \sqrt{(X_2 - X_0)^2 + (Y_2 - Y_0)^2}, \qquad (10.1)$$

где X_j, Y_j (j = 0, 1, 2) – прямоугольные координаты точек на плоскости.

Искомые координаты X_0, Y_0 точки P_0 получаются из решения системы (10.1) двух уравнений с двумя неизвестными.

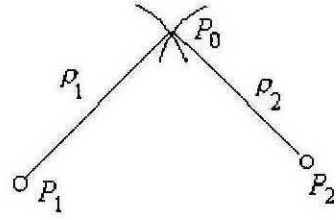


Рис. 10.1. Геометрическая засечка

При обобщении этой задачи от плоского построения к пространственному вводится третья координата Z. Стало быть, для определения теперь уже трех искомых координат X_0, Y_0, Z_0 точки P_0 понадобится решать систему из трех уравнений:

$$\rho_i = \sqrt{(X_i - X_0)^2 + (Y_i - Y_0)^2 + (Z_i - Z_0)^2}, \; i = 1, 2, 3. \qquad (10.2)$$

Следовательно, при решении пространственной линейной засечки должно быть 3 исходных пункта (которые, между прочим, не должны лежать на одной прямой, иначе система не будет иметь определенного решения). Разумеется, количество исходных точек, до которых измерены расстояния, может быть и больше трех, – тогда система (10.2) становится переопределенной и задача решается методом наименьших квадратов. Привлечение избыточных измерений, помимо повышения точности определения координат, дает еще возможность включения в систему уравнений дополнительных неизвестных параметров, определение которых необходимо для работы с GPS (в первую очередь – поправки часов на спутнике и в приемнике, о чем говорится в следующем параграфе).

С помощью описанного метода геодезической засечки в ГНСС решаются две главные задачи (рисунок 10.2):

- определение координат спутника по измеренным до него расстояниям от наземных пунктов с известными координатами (*прямая геодезическая засечка)*;
- определение координат наземного (или надземного) объекта по измеренным до него расстояниям от нескольких спутников, координаты которых известны *(обратная геодезическая засечка)*.

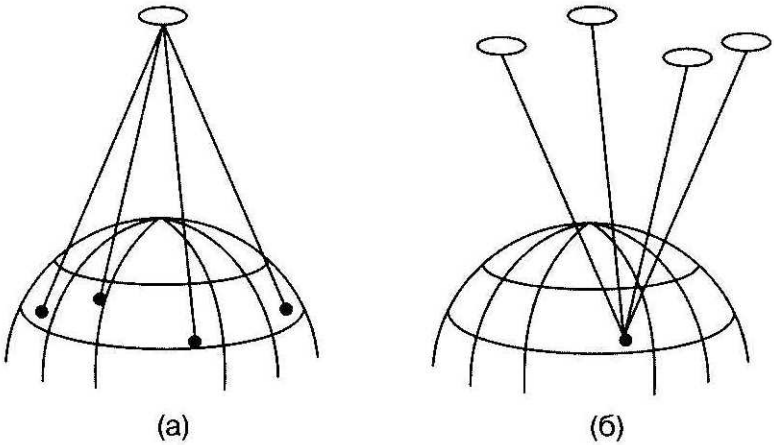


Рис. 10.2. Схемы определения координат спутника *(а)* и наземного пункта *(б)*.

Первую задачу решает служба управления системой, вторую – потребитель.

10.1.2. Методы измерения расстояний

Измерения расстояний до спутников производятся с помощью их локации электромагнитными волнами светового или радио диапазонов. Фактически измеряется время распространения волн между спутником и приемником, а искомое расстояние r вычисляется по формуле:

$$\rho = c\,\tau, \tag{10.3}$$

где c – скорость света.

Существуют однонаправленный и двунаправленный методы локации. При использовании метода *лазерной локации* ИСЗ световой импульс излучается наземным дальномером в момент t_E, достигает спутника и, отразившись от установленного на нем блока призм, возвращается к дальномеру в момент t_Q. Поскольку импульс пробегает трассу дважды (прямо и обратно), в формуле (10.3) следует полагать $\tau = ½\,(t_Q - t_E)$. Отметим, что моменты излучения и приема светового импульса регистрируются с помощью одного устройства – измерителя интервалов времени, благодаря чему здесь не возникает проблема взаимной синхронизации часов, о которой речь пойдет ниже. В то же время, метод лазерной локации ИСЗ требует весьма громоздкого и дорогого электронного и оптического оборудования, устанавливаемого обычно стационарно, и хотя на спутниках GPS имеются уголковые отражатели, лазерная локация их применяется только для решения специальных задач калибровки.

Радиодальномерный метод, применяемый в GPS, – однонаправленный: сигнал излучается передатчиком спутника и регистрируется приемником

пользователя. Значит, в формуле (10.3) следует полагать $\tau = t_Q - t_E$. Но это последнее соотношение справедливо только при условии, что оба момента t_Q, t_E выражены в одной и той же шкале времени. Реально же излучение сигнала фиксируется по часам, установленным на спутнике, а прием – по часам приемника. Этот, казалось бы, маленький нюанс на самом деле является ключевым, определяя облик системы GPS: ее устройство, методика измерений и их обработки подчинены необходимости взаимных сличений бортовых часов и часов пользователя. С другой стороны, пользователь получает дополнительную информацию о точном времени, что делает GPS не только системой позиционирования, но координатно-временной системой.

Рассмотрим этот вопрос несколько подробнее, опуская пока детали, связанные с искажающим влиянием внешней среды, и не конкретизируя способ формирования измерительного сигнала. Обозначим через t^S показания часов на спутнике в момент излучения им сигнала, а через t_R – показания часов приемника в момент приема этого сигнала. Обозначим, далее, через δ^S и δ_R поправки этих часов, которые нужно добавить к их показаниям, чтобы получить время по выбранной эталонной шкале, которую будем называть системной шкалой GPS (см. раздел 9.2.2). С учетом этих обозначений формула (10.3) примет вид:

$$\rho = c\,[(t_R + \delta_R) - (t^S + \delta^S)] = c(t_R - t^S) + c(\delta_R - \delta^S) = R + c\,\Delta\delta\,. \qquad (10.4)$$

В этой формуле R есть измеренное расстояние между приемником и спутником, которое называется *«псевдодальность»*. Оно отличается от действительной геометрической дальности ρ на величину $c\Delta\delta$, обусловленную неучтенным влиянием поправок часов спутника и приемника, которые к моменту измерений неизвестны. Впрочем, для выполнения навигационных и геодезических работ в реальном времени служба управления GPS ведет измерения и прогноз поправок часов каждого спутника и эти данные, а также прогнозированные параметры орбит спутников, передает для потребителей вместе с измерительными сигналами. Остается определить поправку часов приемника.

Из уравнений (10.2) и (10.4) имеем:

$$R_i = \rho_i - c\Delta\delta_i = \sqrt{(X_i^S - X_0)^2 + (Y_i^S - Y_0)^2 + (Z_i^S - Z_0)^2} - c(\delta_R - \delta^S)\,. \qquad (10.5)$$

Это и есть основное уравнение для решения навигационной задачи с помощью GPS. В нем – четыре неизвестных: координаты пункта X_0, Y_0, Z_0 и поправка часов приемника δ_R. Значит, для их определения необходимо составить и решить систему из четырех таких уравнений (i = 1,2,3,4), то есть использовать наблюдения минимум четырех спутников.

10.1.3. Орбиты спутников

В уравнениях (10.5) координаты спутников считаются известными. Их определение представляет собой самостоятельную задачу (см. 10.1.1), которую решает служба управления GPS. Поскольку спутники движутся по своим орбитам, пользователь должен знать параметры орбит, чтобы иметь возможность вычислить координаты спутников на момент измерения. Расчет орбитального движения ИСЗ представляет собой сложную задачу из-за влияния возмущающих сил гравитационной и иной природы, учет которых возможен только на основе более или менее приближенных моделей. Поэтому, служба GPS ведет непрерывное уточнение орбитальных параметров и передает их ближайший прогноз в составе спутниковых радиосигналов. Гораздо более точные элементы орбит, рассчитываемые позже («задним числом»), используют геодезисты при обработке накопленного материала своих измерений.

Чтобы понять смысл информации, передаваемой с борта спутника в составе навигационного сообщения, дадим начальные сведения из теории орбитального движения. В соответствии с законами Ньютона, на спутник действуют две уравновешивающие друг друга силы: центробежная $\mathbf{F}_c = m\mathbf{g} = m\ddot{\mathbf{r}}$ и сила тяготения $\mathbf{F}_g = -GMm\mathbf{r}/r^3$, где m – масса спутника; M – масса Земли; G – гравитационная постоянная; $\mathbf{r}$ – геоцентрический вектор спутника. Следовательно,

$$\ddot{\mathbf{r}} = -GM\frac{\mathbf{r}}{r^3}. \qquad (10.6)$$

Решением этого дифференциального уравнения второго порядка для траектории ИСЗ является эллипс. Согласно второму закону Кеплера, плоскость этого эллипса остается неподвижной относительно инерциальной системы координат и проходит через центр масс Земли, совпадающий с одним из фокусов эллипса. На рисунке 10.3 [10.5] светлым цветом показана плоскость земного экватора, а темным – плоскость эллиптической орбиты ИСЗ. Ось Z геоцентрической инерциальной системы координат направлена вдоль оси вращения Земли, а ось X – в *точку весеннего равноденствия* небесной сферы. Положение спутника задается с помощью шести Кеплеровых параметров. Два из них характеризуют размеры и форму орбитального эллипса (a – большая полуось; e – эксцентриситет), другие три – ориентировку эллипса (i – наклонение его плоскости к экватору; Ω – прямое восхождение узла орбиты, ω – аргумент перигея, разворот эллипса в его плоскости), наконец, шестой параметр T_0 – время прохождения спутником перигея (ближайшей к Земле точки эллипса) служит для отсчета текущего положения спутника на орбите.

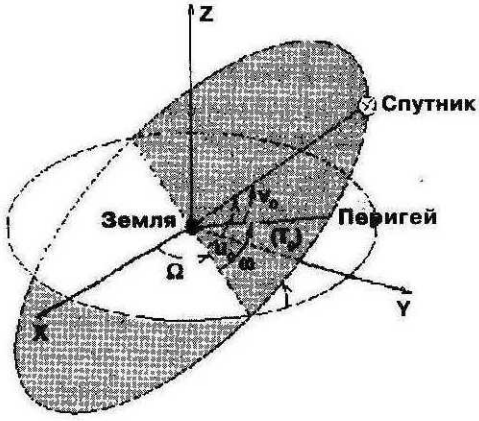


Рис. 10.3. Параметры орбиты ИСЗ.

Зная Кеплеровы параметры можно на любой момент рассчитать векторы координат и скорости ИСЗ (*вектор состояния*): $\mathbf{r}(t), \dot{\mathbf{r}}(t)$. Возможно и обратное преобразование – от вектора состояния к орбитальным параметрам. Но все это верно лишь для идеальных условий отсутствия возмущающих сил. Реально же на движение ИСЗ влияют несимметричность и аномалии гравитационного поля Земли, его изменения под действием приливообразующих сил, притяжения спутника со стороны Луны и Солнца. Среди влияний негравитационной природы важнейшим является световое давление, так как спутники GPS оснащены панелями солнечных батарей большой площади. С учетом этих возмущений уравнение (10.6) преобразуется к виду:

$$\ddot{\mathbf{r}} = -GM\frac{\mathbf{r}}{r^3} + \mathbf{a}(t, \mathbf{r}, \dot{\mathbf{r}}, p_0, p_1, p_2, \ldots) = \mathbf{f}(t, \mathbf{r}, \dot{\mathbf{r}}, p_0, p_1, p_2, \ldots). \qquad (10.7)$$

В этом уравнении дополнительное ускорение **a**, вызванное действием возмущающих факторов, зависит от параметров p_0, p_1, p_2, …, подлежащих определению в ходе решения. Учитывая то, что члены с возмущающими ускорениями в (10.7) существенно меньше главного члена, решение осуществляется по методу *оскулирующих элементов орбиты*, суть которого сводится к следующему. Сначала уравнения (10.7) решают методом численного интегрирования на некотором интервале времени, что дает значения векторов координат и скоростей $\mathbf{r}(t), \dot{\mathbf{r}}(t)$ на последовательные моменты времени t. Найденные значения, далее, преобразуют к наборам орбитальных элементов $a(t)$, $e(t)$, $i(t)$, $\Omega(t)$, $\omega(t)$, $T_0(t)$ по формулам Кеплера для невозмущенного движения. Эти элементы и называют оскулирующими. Таким образом, действительная орбита представляется в виде как бы огибающей, охватывающей непрерывную последователь-

ность невозмущенных орбит с мгновенными значениями оскулирующих элементов. Их использование облегчает вычисление координат спутника, и именно эти орбитальные элементы (преобразованные от инерциальной системы координат к земной) транслируются в навигационном сообщении.

10.2. Сигналы GPS

В этом разделе мы рассмотрим способы технической реализации изложенных выше принципов действия системы GPS: как формируется на спутнике и передается в эфир измерительный сигнал и как работает навигационный приемник пользователя.

10.2.1. Техника формирования кодовых последовательностей

Для того, чтобы измерить время распространения электромагнитных волн от источника до приемника, необходимо избрать некий информационный сигнал – процесс (или явление), позволяющий регистрировать моменты излучения и приема. В так называемых *импульсных* способах излучается короткий сигнал (импульс) и измеряется непосредственно время, затрачиваемое им на распространение вдоль измеряемой линии. При использовании *фазового* способа время измеряется не непосредственно, а через измерение сдвига фаз между принятыми и опорными колебаниями. В системе GPS применен комбинированный метод: излучаемый и опорный сигналы формируются в виде нерегулярной (но одинаковой для источника и приемника) последовательности импульсов, а измерительной информацией служит фазовый сдвиг между этими последовательностями. Такая нерегулярная последовательность импульсов носит название *псевдослучайной* (или псевдошумовой; по-английски – *Pseudorandom noise codes, PRN*), так как на первый взгляд похожа на чисто случайную (две разные последовательности такого рода некоррелированы), но все же формируется по определенному правилу и ее фрагменты через обусловленное время повторяются.

Кодирование излучаемого спутником радиосигнала псевдослучайной последовательностью импульсов, помимо задачи собственно дальномерных измерений, решает еще несколько задач:

- создание наилучших условий различения сигнала в аппаратуре приемника на фоне шумов (доказано, что псевдослучайные коды обладают такими свойствами);
- отождествление конкретного спутника, пославшего сигнал (каждому спутнику назначена своя кодовая последовательность);

- реализация режима ограниченного доступа к GPS, когда высокоточные измерения возможны лишь при санкционированном использовании системы (пользователь должен знать алгоритм построения кодовой последовательности).

Для генерирования псевдослучайной последовательности используется логическое устройство, называемое ленточный сдвиговый регистр. Он содержит некоторое количество однобитовых ячеек (на рисунке 10.4 они помечены номерами 1–5). При подаче на регистр управляющего (тактового) импульса содержимое каждой ячейки (1 или 0) перемещается в соседнюю (на рисунке – направо), а содержимое самой последней ячейки рассматривается как очередное состояние выходного сигнала. В освободившуюся первую ячейку посылается значение 1 или 0, получаемое путем логического суммирования содержимого двух других заданных ячеек (от их выбора и зависит получаемая на выходе псевдослучайная последовательность).

Номера ячеек	*1*	*2*	*3*	*4*	*5*
Исходное состояние	1	0	1	1	0
Следующее состояние	1	1	0	1	1

Рис. 10.4. Принцип работы ленточного сдвигового регистра

В примере, показанном на рисунке 10.4, заданную пару образуют ячейки с номерами 2 и 3. Под действием управляющего импульса содержимое ячеек 1–4 перемещается в ячейки 2–5, на выход устройства поступает 0 (из 5-й ячейки), а в ячейку №1 записывается значение 1 (логическая сумма из ячеек №№ 2 и 3). Можно убедиться, что при дальнейшем поступлении управляющих импульсов на выходе устройства образуется последовательность 1101110010…

Для формирования общедоступного кода *C/A* (*Clear Access / Coarse Acquisition*) используется комбинация из двух 10-битовых сдвиговых регистров, выходные сигналы которых складываются по модулю 2. Частота следования тактовых импульсов составляет 1,023 МГц, а период повторения псевдослучайной последовательности равен 1023 циклам, следовательно его продолжительность – 1 миллисекунда.

Формирование точного защищенного кода *P (Protected / Precise)* происходит по более сложному алгоритму: он образуется путем комбинирования двух последовательностей, каждая из которых генерируется парой сдвиговых регистров. Благодаря этому последовательность очень долго

не повторяется (более $2 \cdot 10^{14}$ циклов), что при частоте следования тактовых импульсов 10,23 МГц составляет 266,4 суток. Этим достигается высокая степень защищенности *P*-кода от несанкционированного доступа (сложность расшифровки его алгоритма). Каждому спутнику индивидуально назначается недельный отрезок из этой длинной последовательности. Старт недельного цикла всегда происходит по воскресеньям в 0 часов всемирного времени. Несмотря на высокую степень защищенности предусмотрена еще возможность дополнительного засекречивания *P*-кода – преобразования его в зашифрованный *Y*-код (режим *AS – Anti-Spoofing*).

10.2.2. Структура и передача в эфир навигационного сигнала

Формирование тактовых импульсов для воспроизведения кодовых последовательностей *C/A* и *P*-кодов, а также и других сигналов (несущих и бинарных) происходит на основе использования одного опорного высокостабильного генератора. Этот генератор, собственно говоря, и является основой бортовых «часов», обеспечивая синхронизацию и временную привязку всех процессов. Базовая частота, воспроизводимая опорным генератором, составляет f_0 = 10,23 МГц (с суточной нестабильностью 10^{-13}). Путем умножения или деления частоты f_0 (см. табл. 10.1) получают две несущие частоты, обозначаемые как *L1* и *L2*, тактовые импульсы для воспроизведения кодовых последовательностей и низкой частоты (50 Гц) для передачи навигационного сообщения.

Таблица 10.1. Компоненты спутникового сигнала.

Назначение	Коэффициент	Частота (МГц)
Базовая частота	f_0	10,23
Несущая *L1*	$f_1 = f_0 \times 154$	1575,42
Несущая *L2*	$f_2 = f_0 \times 120$	1227,60
P-код	f_0	10,23
C/A-код	$f_0 / 10$	1,023
Навигационное сообщение	$f_0 / 204600$	$50 \cdot 10^{-6}$

Несущие частоты, служащие для передачи сигнала в эфир, модулируются измерительными кодовыми последовательностями и бинарными кодами навигационного сообщения. При этом используется метод так на-

зываемой *бифазной модуляции*, что оказывается очень удобным, поскольку вся передаваемая в составе сигнала информация представлена в виде битовой последовательности из единиц и нулей. Идея метода заключается в управлении фазой синусоидального сигнала несущей при помощи кодов битовой последовательности $k(t)$:

$$L(t) = a \sin(2\pi f t + \pi k(t)).$$

Когда кодовый сигнал $k(t)$ изменяет свое состояние с 0 на 1 или наоборот, фаза скачком изменяется на 180^0 и, следовательно, знак синуса меняется на противоположный. Этот процесс схематически иллюстрирует рисунок 10.5.

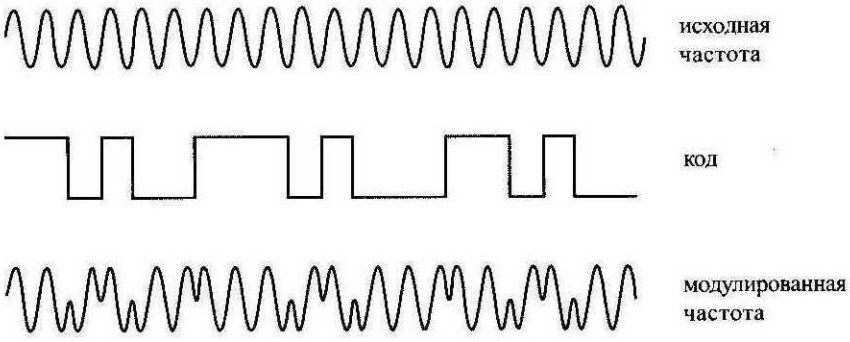


Рис. 10.5. Схема бифазной модуляции.

Обе несущие частоты *L1* и *L2* модулируются точным *P*-кодом, что позволяет в дальнейшем при обработке принятых сигналов учитывать их задержку в ионосфере (см. разд. 10.3.1). Общедоступный *C/A*-код передается только на одной частоте *L1*, причем с фазовым сдвигом на 90^0 по отношению к *P*-коду для удобства их последующего разделения. Информация навигационного сообщения *D* передается также на двух несущих частотах. Таким образом, комбинированные сигналы, передаваемые со спутника, могут быть представлены следующими соотношениями:

$$\begin{aligned} L1(t) &= a_1 P(t) D(t) \cos 2\pi(f_1 t) + a_1 C/A(t) D(t) \sin 2\pi(f_1 t), \\ L2(t) &= a_2 P(t) D(t) \cos 2\pi(f_2 t). \end{aligned} \qquad (10.8)$$

Обратимся теперь к структуре и содержанию навигационных сообщений. Объем одного сообщения (*кадра*) составляет 1500 бит. Кадр состоит из 5 *субкадров* (300 бит), содержащих по 10 *слов* (30 бит). Нетрудно подсчитать, что при тактовой частоте 50 Гц время передачи кадра составляет 30 секунд. Первые два слова каждого субкадра содержат телеметрическую и диагностическую информацию.

Основное содержание первого субкадра составляет информация о поправках часов спутника, представленных в виде коэффициентов квадратичного полинома, а также сведения о техническом состоянии систем спутника (*«здоровье»*). Второй и третий субкадры содержат текущие значения Кеплеровых элементов орбиты данного спутника (см. выше, п. 10.1.3), а также коэффициенты полиномов и периодических функций, описывающих изменения орбитальных элементов со временем.

Четвертый и пятый субкадры, в отличие от первых трех, содержат информацию не только о данном спутнике, а обо всех спутниках системы: приближенные значения элементов их орбит (*«альманах»*), статус работоспособности, поправки за влияние ионосферы и др. Объем этой информации достаточно большой – в четвертом и пятом субкадрах укладывается лишь $^{1}/_{25}$ ее часть, поэтому она передается порциями в последовательности из 25 кадров. На их передачу уходит 12,5 минут. Столько времени понадобиться приемнику для подготовительного этапа *захвата* и отождествления сигналов спутников, если приемник установлен на новом месте и в его памяти отсутствует свежие значения альманаха.

10.2.3. Прием и обработка сигналов

Посланные со спутников сигналы несут информацию, необходимую для измерения дальностей и определения навигационных параметров (координат, скоростей и точного времени), но сам процесс измерений происходит в приемнике. Сейчас в мире число типов и модификаций приемников исчисляется сотнями, в них применяются различные ухищрения, чтобы ослабить влияние искажающих факторов и в наибольшей степени удовлетворить требованиям различных категорий пользователей. Но, тем не менее, основной принцип измерений во всех навигационных приемниках общий, и все они состоят из радиотехнической части, обеспечивающей прием сигнала, генератора частоты, микропроцессора, запоминающего устройства и устройства управления.

Всенаправленная антенна приемника может быть совмещена с ним, либо выполнена в виде отдельного выносного блока, в котором размещается также предварительный усилитель сигналов высокой частоты. Фазовый центр антенны и является той точкой, от которой измеряется расстояние до спутника. Усиленные сигналы поступают в радиочастотный блок, где происходит их демодуляция, в результате чего выделяется цифровая составляющая (измерительные коды и навигационное сообщение) и аналоговая (синусоидальная). На основе кварцевого генератора частоты, задающего внутреннюю шкалу времени приемника, воспроизводится сетка опорных сигналов подобных тем, что излучают спутники (*реплика*). Сопоставление принятых и опорных сигналов и дает возможность

определить задержку сигнала на трассе, вычислить псевдодальности и искомые навигационные параметры.

Поскольку все спутники GPS передают сигналы на одной и той же паре несущих частот, в приемник поступает смесь сигналов всех спутников, находящихся над горизонтом. В процессе обработке принятых сигналов нужно решить следующие задачи:

- выделить в потоке сигналов составляющие, относящиеся к конкретным спутникам;
- разделить сигналы каждого спутника на кодовую последовательность, навигационное сообщение и чисто синусоидальную составляющую;
- определить временной сдвиг принятых сигналов по отношению к опорным.

Может показаться удивительным, но все эти три задачи, а также еще и некоторые вспомогательные, решаются одновременно и на основе одного общего метода корреляционной обработки принятого и опорного сигналов. Суть этого метода состоит в перемножении сравниваемых сигналов при помощи специального аналогового или цифрового устройства – коррелятора. На выходе этого устройства образуются нормированные средние значения результата перемножения, фактически представляющие собой последовательность коэффициентов корреляции. Интервал усреднения здесь составляет 1–2 мс (напомним, что в одной миллисекунде укладывается 1023 тактовых импульса *C/A*-кода). Коэффициент корреляции, как известно, характеризует тесноту связи двух совокупностей случайных величин, в нашем случае – степень «похожести» принятого и опорного сигналов. В приемнике поочередно генерируются и подаются на вход коррелятора опорные сигналы с кодовыми последовательностями разных спутников. В результате на его выходе выделяется только та составляющая, которая коррелирует с поданным в данный момент опорным сигналом. При этом степень корреляции и, следовательно, уровень выходного сигнала достигают максимума, если временное смещение между сравниваемыми сигналами близко к нулю. В процессе измерений с помощью автоматической плавной перестройки временной задержки опорного сигнала определяется то ее значение, при котором корреляция максимальна – именно это значение и представляет интересующий нас временной сдвиг между принятым и опорным сигналами, по которому и вычисляется псевдодальность.

Строго говоря, поскольку кодовые последовательности не чисто случайны, а через какой-то период повторяются, корреляционная обработка дает значения псевдодальностей с неопределенностью в целое число таких повторяющихся блоков (для *C/A*-кода – до 1 мс или, в пересчете на дальность, 300 км). Эта неопределенность снимается благодаря косвен-

ной оцифровке повторяющихся блоков с помощью *меток времени*, содержащихся в составе навигационного сообщения и несущих информацию о дате и точном времени. Метки времени привязаны к тактовым импульсам навигационного сообщения, когерентным с сигналами опорного генератора и со всеми другими спутниковыми сигналами [10.4].

Точность измерения описанным способом характеризуется погрешностью порядка 1^0 фазы сигнала или 0,3 % длительности периода, в данном случае – периода следования кодовых импульсов (см. п. 10.2.1). Следовательно, для *C/A*-кода эта погрешность составляет около 3 нс, а для *P*-кода – 0,3 нс, или, в пересчете на псевдодальность, 1 м и 0,1 м (здесь, понятно, не учтено влияние внешних источников погрешностей, о которых говорится в разд. 10.3).

В точных геодезических приемниках реализовано еще измерение фазового смещения самой несущей частоты по отношению к опорному сигналу. Поскольку периоды несущих в 154 и 120 раз короче тактовых периодов кодовых последовательностей (см. табл. 10.1), такие измерения позволяют повысить точность определения псевдодальностей до миллиметрового уровня. При этом, правда, удается измерить только дробную часть цикла в фазовом смещении, а целое число длин волн в измеряемой дальности остается неизвестным. Но эта проблема разрешается в ходе последующей совместной обработки измерений на двух или нескольких пунктах. (см. п. 10.4.2). Существуют, однако, и две чисто технические проблемы: как выделить синусоидальную составляющую из общего кодированного сигнала и как надежно разделить сигналы несущей частоты отдельных спутников – ведь ее значение для всех спутников одинаково.

Для решения первой задачи нужно, вообще говоря, точно знать структуру кодовых последовательностей. Но даже если она неизвестна (*P*-код не является общедоступным), есть простое и изящное решение – нужно лишь сквадратировать принятый сигнал, т.е. умножить его на самого себя, в результате чего модуляция исчезнет. Это легко понять из рассмотрения формул (10.8): коэффициенты $C/A(t)$, $P(t)$ и $D(t)$ в них, равные +1 или –1, при возведении в квадрат становятся все равными единице. Существуют и другие методы выделения чисто синусоидальных колебаний.

При решении задачи разделения несущих частот разных спутников используется тот факт, что вследствие движения спутников частоты принимаемых сигналов испытывают доплеровские смещения, разные для разных положений спутников и достигающих нескольких килогерц. Иначе говоря, реально значения несущих частот принимаемых сигналов разных спутников все-таки неодинаковы. В приемнике реализованы схемы автоподстройки частоты и фазы, которые корректируют опорные сигна-

лы индивидуально для каждого спутника в соответствии с уравнениями его движения, чем и достигается селекция.

Вычислительное устройство приемника управляет его работой по заданным программам и производит обработку кодовых измерений по алгоритмам, описанным в разд. 10.1. Текущие результаты – координаты и скорость приемника, а также точное время, отображаются на табло и записываются в запоминающее устройство. Фазовые измерения не обрабатываются, а лишь сохраняются в памяти.

10.3. Методы учета факторов, искажающих результаты измерений

Источники погрешностей, воздействующих на результаты измерений псевдодальностей, могут быть условно разделены на три группы, связанные, соответственно, со спутником, со средой распространения сигнала и с приемником. В процессе обработки измерений можно исключить или ослабить влияние погрешностей двумя путями: либо продуманным комбинированием наблюдений (например, образованием разностей), либо представлением погрешностей в виде моделей, параметры которых задаются эмпирически или включаются в число неизвестных при уравнивании измерений методом наименьших квадратов.

При прохождении через атмосферу скорость распространения радиоволн отличается от скорости света в вакууме, поэтому в вычисленные значения псевдодальностей необходимо ввести поправку. Принято значение этой поправки разбивать на две части, одна из которых относится к нижней (до 10 км) части атмосферы – *тропосфере*, а другая – к верхней, *ионосфере*, содержащей электрически заряженные частицы.

10.3.1. Влияние ионосферы

Ионосфера простирается до высот порядка 1000 км. Под воздействием космического излучения (прежде всего – солнечного) в ней образуются свободные электроны и ионы, оказывающие влияние на скорость прохождения радиоволн, причем это влияние зависит от частоты. Поправка к псевдодальности, учитывающая влияние ионосферы, достаточно точно выражается эмпирической формулой

$$\Delta\rho_{ion} = \pm\frac{k}{f^2}, \tag{10.9}$$

где f – частота сигнала; k – коэффициент, зависящий от концентрации электронов вдоль трассы распространения сигнала.

Знак в этой формуле зависит от того, какой сигнал используется для вычисления псевдодальности: фаза несущей частоты или кодовая после-

довательность, содержащая группу волн в некоторой полосе частот. Первому случаю соответствует *фазовая скорость* распространения, второму – *групповая скорость,* меньшая по величине. Интересно при этом, что поправки $\Delta\rho_{ion}$ оказываются, согласно теории, одинаковыми по абсолютной величине в обоих случаях, но при фазовых измерениях их следует вычесть, а при кодовых – прибавить к значению расстояния, рассчитанного с использованием значения скорости света в вакууме (10.3) .

Концентрация электронов на пути следования радиоволн в ионосфере сильно зависит от солнечной активности (изменения количества солнечных пятен с примерно 11-летним циклом), от времени года и времени суток, от географического положения станции. Значение поправки $\Delta\rho_{ion}$ в направлении зенита может колебаться от 1 до 20 м, а при низких углах наклона радиолуча – достигать 100 м. Это, конечно, очень большие величины, которые надо обязательно каким-то образом учитывать, чтобы не свести на нет высокую точность GPS измерений. Существуют для этого два метода: двухчастотный и модельный.

При использовании двухчастотных приемников, способных измерять псевдодальности по кодовым сигналам или по фазе несущей на двух частотах f_1, f_2 (см. табл. 10.1), можно так скомбинировать измерения, что влияние ионосферы исключится. Обратимся к выражению для псевдодальности (10.5), добавив в его правую часть необходимые поправки за влияние ионосферы $\Delta\rho_{ion}$ по формуле (10.9) и тропосферы $\Delta\rho_{tr}$ (о которой поговорим чуть позже). Пусть P_1 и P_2 – результаты кодовых измерений псевдодальности R на двух частотах:

$$P_1 = \text{с} - c(\text{д}_R - \text{д}^S) + \frac{k}{f_1^2} + \Delta\text{с}_{tr},$$
$$P_2 = \text{с} - c(\text{д}_R - \text{д}^S) + \frac{k}{f_2^2} + \Delta\text{с}_{tr}. \qquad (10.10)$$

Если умножить первое из этих выражений на $f_1^2/(f_1^2 - f_2^2)$, а второе на $-f_2^2/(f_1^2 - f_2^2)$ и затем сложить, то получим:

$$P_3 = \rho - c(\delta_R - \delta^S) + \Delta\rho_{tr}. \qquad (10.11)$$

Как видим, комбинированный результат измерения псевдодальности P_3 действительно не зависит от влияния ионосферы. Точно так же можно образовать комбинированный «ионосферо-независимый» результат фазовых измерений L_3. Собственно говоря, именно ради этого и ведется передача сигналов GPS на двух частотах.

Однако, такой метод применим только при работе с фазовыми двухчастотными приемниками, большинство же широко распространенных типов менее дорогих приемников работают по общедоступному коду *C/A*, кото-

рый передается только на одной несущей частоте. В таких приемниках учет влияния ионосферы производится с помощью математической модели, текущие параметры которой содержатся в навигационном сообщении. Они вычисляются службой управления GPS на основе собираемой метеорологической информации. Для вычисления ионосферной задержки Δt_{ion} используется формула, основанная на модели Дж. А. Клобушара:

$$\Delta t_{ion} = A_1 + A_2 \cos\left[\frac{2\pi(t - A_3)}{A_4}\right],$$

где A_1 – задержка кодовых сигналов в ночное время, принимаемая постоянной и равной 5 нс, а для дневной части суток (от 8 до 20 часов местного времени) добавляется второй член, в котором:

A_3 – постоянная, равная 14 часам;

$A_2 = \alpha_1 + \alpha_2\varphi_m + \alpha_3\varphi_m^2 + \alpha_4\varphi_m^3$;

$A_4 = \beta_1 + \beta_2\varphi_m + \beta_3\varphi_m^2 + \beta_4\varphi_m^3$;

t – местное время точки на линии «спутник–приемник», находящейся на высоте 350 км (примерно на этой высоте концентрация электронов имеет среднее значение);

φ_m – геомагнитная широта этой точки.

Коэффициенты α_i , β_i как раз и являются теми величинами, значения которых ежесуточно определяет ведущая станция управления GPS и закладывает на спутники для передачи в составе навигационного сообщения. С использованием этих данных и приведенных формул потребитель имеет возможность рассчитать ионосферную поправку $\Delta\rho_{ion} = c\,\Delta t_{ion}$.

10.3.2. Влияние тропосферы

Тропосфера является электрически нейтральной частью атмосферы, и потому запаздывание радиоволн в ней не зависит от частоты и одинаково для кодовых и фазовых измерений. Следовательно, для вычисления тропосферной поправки к измеренной псевдодальности $\Delta\rho_{tr}$ на первый план выходит метод моделирования.

Согласно теории, величина задержки в тропосфере зависит от показателя преломления вдоль трассы распространения волн, который в свою очередь зависит от температуры, давления и влажности. При установлении этих зависимостей принято рассматривать тропосферу как смесь сухого воздуха и водяных паров и рассчитывать задержки для каждого из этих компонентов по своим формулам. Отметим, что «сухая» часть задержки составляет около 90% от полной, и она достаточно точно определяется по метеоданным, измеренным вблизи приемника. «Влажная» часть

зависит от давления водяных паров, которое очень неустойчиво во времени и пространстве.

Существуют различные модели тропосферных задержек, отличающиеся гипотезами о характере изменения метеопараметров с высотой, наиболее употребительными из которых при GPS-измерениях являются две: Хопфилда и Саастамойнена. Приведем формулу для второй из них, чтобы дать представление о подобных моделях:

$$\Delta\rho_{tr} = \frac{0{,}002277}{\cos z}\left[P + \left(\frac{1255}{T} + 0{,}05\right)e - \operatorname{tg}^2 z\right].$$

Здесь z – зенитное расстояние наблюдаемого спутника, P – атмосферное давление, T – температура, e – давление водяных паров (температура выражена в Кельвинах, другие две величины – в миллибарах).

Отметим одно важное обстоятельство: величина тропосферной задержки зависит от угла наклона направления на спутник. Действительно: ведь чем сильнее наклон, тем длиннее путь через тропосферный слой, а вертикальный луч «прошивает» тропосферу по самому короткому пути. Это нашло отражение и в приведенной выше формуле: когда спутник находится в зените ($z = 0$), поправка самая маленькая. Указанное обстоятельство позволяет реализовать другой метод определения тропосферных поправок, без измерения метеопараметров. Для этого достаточно найти лишь поправку в зените $\Delta\rho_{tr}(0)$, полагая для других зенитных расстояний

$$\Delta\rho_{tr}(z) = \frac{\Delta\rho_{tr}(0)}{\cos z}.$$

Если измерения на пункте продолжаются достаточно долго и наблюдаются спутники при разных углах наклона, то поправку $\Delta\rho_{tr}(0)$ можно определить включив ее в число неизвестных при совместной обработке измерений методом наименьших квадратов. Такой способ даже точнее, чем по измерениям метеоданных, но его можно применять не в реальном времени в ходе измерений, а только в процессе их постобработки.

В среднем тропосферная поправка в зените составляет 2,3 м, а при угле наклона над горизонтом 15^0 – около 10 м.

10.3.3. Другие факторы

Еще одним вредным фактором, связанным со средой распространения радиосигнала GPS, является переотражение его от окружающих объектов (земной и водной поверхности, строений, деревьев и т.п.). В результате на антенну попадают не только прямые сигналы от спутника, но и смесь отраженных, прошедших больший путь, что приводит к дополнительной задержке суммарного сигнала. Это явление в русской литературе обозначается несколько неуклюжим, но уже ставшим привычным термином

«многопутность», являющимся калькой с английского термина *multipath* (иногда говорят также «многолучевость»). Величина погрешности при измерении дальностей по кодовым сигналам при неблагоприятных условиях может достигать 50 м и более. Фазовые измерения оказываются в более выгодном положении: ведь длина волны несущей составляет около 20 см, так что погрешность из-за многопутности может составить лишь ее долю (5–10 см), – знать целое число длин волн при фазовых измерениях вовсе не нужно, так как оно все равно восстанавливается в ходе последующих вычислений («меньше знаешь – меньше хлопот!»).

В отличие от тропосферных и ионосферных погрешностей, влияние многопутности не поддается удовлетворительному моделированию, поэтому борьба с ней ведется путем принятия защитных мер, таких как:

- расположение антенны, по возможности, вдали от отражающих объектов;
- устройство в антенне защитных экранов;
- удлинение сеансов измерений с тем, чтобы варьировать направления хода радиолучей от спутников.

Влияние других источников погрешностей (вариации хода часов на спутнике и в приемнике, неточность эфемерид, колебания фазовых центров передающей и приемных антенн, нестабильность аппаратурных временных задержек) сказывается, в основном, не на измеренных псевдодальностях, а уже на результатах вычислений, полученных по этим измерениям: координатах, скоростях и точном времени. Мы коснемся этих вопросов в следующих разделах, а здесь остановимся еще на одном факторе, который следует учитывать в самих измерениях – релятивистских эффектах. Их влияние сказывается на орбитах спутников, на распространении сигналов и на ходе часов спутника.

Первый из этих эффектов является следствием движения спутника в гравитационном поле Земли, он создает дополнительное возмущающее ускорение, которое не столь велико и учитывается при численном интегрировании орбит спутников.

Гравитационное поле вызывает также дополнительное пространственно-временное искривление пути распространения сигнала. Увеличение дальности при этом может быть выражено формулой [10.1]:

$$\Delta\rho_{rel} = \frac{2\mu}{c^2}\ln\frac{r^S + r_R + \rho_R^S}{r^S + r_R - \rho_R^S},$$

где r^S и r_R – расстояния от центра Земли до спутника и приемника, ρ_R^S – дальность между приемником и спутником; c – скорость света; $\mu = 3986005 \cdot 10^8$ м^3/с^2 – гравитационная постоянная Земли. Вычисленное по этой формуле значение $\Delta\rho_{rel}$ составляет около 19 мм.

Согласно теории относительности, ход времени в собственной координатной системе спутника отличается от хода в земной системе из-за разницы гравитационных потенциалов и из-за движения спутника. Для спутников GPS это приводит к смещению частоты его базового генератора, которое можно рассчитать по формуле:

$$\Delta f_{rel} = \frac{f_0^S - f_0}{f_0} = \frac{1}{2}\left(\frac{v}{c}\right)^2 + \frac{\mu}{c^2}\left[\frac{1}{R_E + h} - \frac{1}{R_E}\right],$$

где v – скорость спутника; R_E – радиус Земли; h – высота спутника над земной поверхностью; f_0^S – частота, сформированная на спутнике, а f_0 – принятая на земле.

Подставив в эту формулы численные значения входящих в нее величин, получим $\Delta f_{rel} = 4{,}464 \cdot 10^{-10}$. Для частоты базового генератора $f_0 = 10.23$ МГц абсолютное смещение составит $4{,}55 \cdot 10^{-3}$ Гц. Чтобы все излучаемые спутником сигналы воспринимались в масштабе земного времени, частота базовых генераторов спутников при настройке на земле была умышленно смещена до величины 10,22999999545 МГц.

10.4. Алгоритмы и точность координатно-временных определений

В предыдущих подразделах рассмотрены вопросы, связанные с измерением расстояния между приемником и спутником. Обсудим теперь, как по данным этих измерений определить интересующие пользователя навигационные параметры: координаты приемника и точное время. Идея решения этой задачи изложена в подразделе 10.1, она достаточно проста и основана на решении уравнений вида (10.5). В вычислительном устройстве приемника именно этот способ и применяется для определения в реальном времени текущих навигационных параметров по сигналам общедоступного *C/A*-кода. И если пользователя устраивает точность определения координат в несколько десятков метров, он вполне может ограничиться значениями, выдаваемыми приемником. Это есть, так сказать, штатный способ использования системы GPS. Между тем, потенциальная точность измерительной информации, содержащейся в сигналах GPS, гораздо выше, и для ее реализации разработаны специальные приемы, позволяющие довести точность координатных определений до миллиметрового уровня. Это – дифференциальные методы и использование измерений фазы несущей.

10.4.1. Абсолютный и относительные методы измерений

Метод определения навигационных параметров, впрямую основанный на решении системы уравнений вида (10.5), называют *абсолютным* в том

смысле, что он дает непосредственно значения координат в общеземной системе и поправку часов пользователя относительно эталонной шкалы времени. В противоположность этому, при использовании *относительных* методов искомые параметры определяют через разности измерений, полученных двумя приемниками, один из которых стоит на определяемой точке, а другой – на опорной, точные координаты которой известны. Смысл заключается в том, что при образовании разностей исключается влияние погрешностей, одинаково искажающих оба ряда измерений.

Пусть, например, с пунктов A и B одновременно измерены псевдодальности R_A^S и R_B^S до спутника S. Добавив к уравнению (10.5) поправки за влияние ионосферы $\Delta\rho_{ion}$, тропосферы $\Delta\rho_{tr}$ и совокупности других факторов $\Delta\rho_\Sigma$, запишем:

$$\begin{aligned} R_A^S &= \rho_A^S - c\delta_A + c\delta^S + \Delta\rho_{ion,A}^S + \Delta\rho_{tr,A}^S + \Delta\rho_{\Sigma,A}^S \\ R_B^S &= \rho_B^S - c\delta_B + c\delta^S + \Delta\rho_{ion,B}^S + \Delta\rho_{tr,B}^S + \Delta\rho_{\Sigma,B}^S \end{aligned}. \qquad (10.12)$$

Образуем разность псевдодальностей:

$$R_{AB}^S = R_A^S - R_B^S = \rho_{AB}^S - c\delta_{AB} + \Delta\rho_{ion,AB}^S + \Delta\rho_{tr,AB}^S + \Delta\rho_{\Sigma,AB}^S \qquad (10.13)$$

Здесь двойными индексами AB символически обозначены разности между величинами, относящимися к пунктам A и B. Мы видим, что при образовании разностей между уравнениями (10.12) исключились поправки часов спутника δ^S. Предположим теперь, что с пунктов A и B в это же время измерены псевдодальности еще и до другого спутника Q. Образовав разности между ними $R_{AB}^Q = R_A^Q - R_B^Q$, получим выражение, подобное (10.13), но с верхними индексами Q. А теперь возьмем разности между значениями R_{AB}^S и R_{AB}^Q:

$$R_{AB}^{SQ} = R_{AB}^S - R_{AB}^Q = \rho_{AB}^{SQ} + \Delta\rho_{ion,AB}^{SQ} + \Delta\rho_{tr,AB}^{SQ} + \Delta\rho_{\Sigma,AB}^{SQ} \qquad (10.14)$$

Разности псевдодальностей вида (10.13) принято называть *одинарными*, а вида (10.14) – *двойными* (по-английски *single difference* и *double difference*)[1]. Мы видим, что двойные разности не содержат уже и поправок часов приемников, т.е. влияние погрешностей хода часов спутников и приемников в них полностью исключено. Влияние других погрешностей в двойных разностях существенно ослабляется, особенно если пункты расположены не очень далеко друг от друга. При использовании фазовых измерений образование двойных разностей помогает еще разрешению неоднозначностей (об этом – в п. 10.4.2). Ценой же, которую за все это

[1] Иногда эти разности ошибочно называют «первые» и «вторые», но эти термины в математике используются для обозначения разностей между последовательными элементами одного ряда, здесь же речь идет о комбинировании разных рядов измерений.

приходится платить, является то, что из решения системы *разностных* уравнений вида (10.14) можно получить только *разности* координат между пунктами (ΔX_{AB}, ΔY_{AB}, ΔZ_{AB}). Но потом, прибавляя их к известным заранее координатам *исходного* пункта A (X_A, Y_A, Z_A), можно найти координаты *определяемого* пункта B (X_B, Y_B,Z_B). Это – один из видов относительного метода координатных определений с помощью GPS. Если имеется не один, а некоторое количество определяемых пунктов, образующих геодезическую сеть, то при использовании относительного метода необходимо наличие в сети минимум одного исходного пункта.

Метод образования двойных разностей обычно применяется при постобработке измерений, собранных с двух или нескольких пунктов. В некоторых современных типах навигационных приемников предусмотрена установка радиомодема, через который можно сразу передавать текущую измерительную информацию на вычислительную станцию. Но для выполнения навигационных определений в реальном времени сейчас наибольшее применение находят два относительных метода – *коррекции псевдодальностей* и *коррекции координат*.

Оба метода основаны на работе постоянно действующей опорной станции, обслуживающей конкретный регион. На станции непрерывно ведутся измерения сигналов всех видимых спутников GPS. Координаты станции хорошо известны и она оборудована высокостабильными часами. Благодаря этому можно предвычислять «точные» значения псевдодальностей на моменты наблюдения всех спутников. Считается, что отклонения реально измеренных псевдодальностей от расчетных обусловлены совокупным влиянием разных источников погрешностей, значительную долю которых составляют общие для наблюдений в данном регионе. Эти отклонения и представляют собой корректирующие поправки, которые распространяются на регион по радио или с помощью других средств связи. В приемнике пользователя полученные поправки учитываются в измеряемых псевдодальностях, и далее обычным образом из решения навигационной задачи определяются абсолютные координаты. Такова сущность метода коррекции псевдодальностей.

При использовании метода коррекции координат передаются поправки, представляющие собой разности между точными координатами опорной станции и их значениями, вычисляемыми по данным текущих измерений. В приемнике пользователя эти поправки просто прибавляются к значениям координат определяемого пункта, получаемых на основании собственных измерений. Этот метод проще, но эффект компенсации погрешностей может быть получен только при соблюдении условия, чтобы на опорной станции и определяемом пункте наблюдались одинаковые созвездия спутников.

10.4.2. Особенности обработки фазовых измерений

Мы уже говорили в п. 10.2.3 о том, что точные геодезические приемники позволяют измерять временное смещение сигнала, принятого со спутника, относительно опорного сигнала приемника не только с помощью кодовых псевдослучайных последовательностей, но и непосредственно по фазовому сдвигу самой несущей. Правда, при этом невозможно измерить количество длин волн, укладывающееся на линии «спутник–приемник», но зато «хвостик» – дробная часть фазового сдвига, – измеряется очень точно.

Как же выглядят результаты фазовых измерений L? В отличие от кодовых псевдодальностей они выражаются не в линейной мере (метрах), а в циклах с долями до 0,001 цикла. В момент, когда приемник осуществляет захват сигнала спутника, состояние счетчика целых циклов произвольно, и оно переносится в результат первого измерения L. В ходе последующих наблюдений этого же спутника расстояния до него меняются и эти изменения точно отражаются в изменениях величины L. Таким образом, неизвестным является только одно целое число циклов N, которое надлежит добавить к первому и всем остальным измерениям, чтобы получились правильные значения псевдодальностей. Это справедливо до тех пор, пока наблюдения спутника идут непрерывно и без сбоев, которые могут происходить, например, из-за затухания радиосигнала. В таких случаях счетчик не может определить количество пропущенных циклов несущей, и поэтому для последующего ряда наблюдений приходится вводить новый неизвестный параметр N'. Количество таких неизвестных равно количеству наблюдаемых спутников плюс общее количество перерывов. Определение их и представляет собой проблему *разрешения фазовых неоднозначностей.*

Для определения значений параметров N их можно включить в число неизвестных величин при обработке измерений методом наименьших квадратов. Но при частых сбоях количество дополнительных неизвестных может оказаться слишком большим, поэтому сначала выполняется анализ фазовых измерений с целью локализации пропусков и делаются попытки определения числа пропущенных циклов. Существуют разные приемы проведения такого анализа, основанные на сглаживании и прогнозе фазовых измерений, сопоставлении их с кодовыми измерениями и др.

Сама необходимость разрешения фазовых неоднозначностей обусловливает возможность применения к фазовым измерениям только относительного метода координатных определений. Уравнения для двойных разностей (10.14) при этом приобретают вид:

$$L_{AB}^{SQ} \cdot \lambda = \rho_{AB}^{SQ} + \Delta\rho_{ion,AB}^{SQ} + \Delta\rho_{tr,AB}^{SQ} + \Delta\rho_{\Sigma,AB}^{SQ} + N_{AB}^{SQ} \cdot \lambda, \quad (10.15)$$

где λ – длина волны несущей, а N_{AB}^{SQ} – двойная разность параметров неоднозначности.

В результате решения системы уравнений вида (10.15) методом наименьших квадратов получаются, конечно, вещественные значения неизвестных N_{AB}^{SQ}, в то время как по смыслу – это целые числа циклов несущей частоты. Наиболее простой прием – округление вычисленных значений до ближайшего целого, вовсе необязательно даст верное искомое значение параметра N_{AB}^{SQ}: согласно правилам теории вероятностей его следует искать внутри доверительного интервала, определяемого погрешностью вычисленного значения неизвестного. Это можно, например, сделать так: перебирая поочередно все целые числа внутри доверительного интервала, подставлять их в уравнения (10.15) на место N_{AB}^{SQ} и решать их. Вариант, который даст наименьшую среднюю квадратическую погрешность решения, и определит вероятнейшее значение целого числа циклов. Но, как уже говорилось, количество неизвестных вида N_{AB}^{SQ} на практике оказывается достаточно большим, так что число всевозможных комбинаций тестируемых целых чисел, с каждой из которых нужно решать систему уравнений, может достигать десятков тысяч. Это чересчур много даже для современной вычислительной техники, поэтому на практике используются различные упрощенные, но вполне эффективные стратегии поиска целочисленных значений N_{AB}^{SQ}.

Вполне очевидно, что разрешение фазовых неоднозначностей должно быть выполнено для измерений на каждой из двух несущих частот. Но после того, когда целочисленные значения N_{AB}^{SQ} установлены, можно образовать свободные от влияния ионосферы комбинированные значения фазовых измерений, подобные кодовым (10.11), и использовать уже их для решения уравнений (10.15).

10.4.3. Точность различных методов GPS-измерений

Многолетней практикой применения средств навигационной системы GPS выработаны достаточно четко очерченные методы для решения различных по назначению, требуемой точности и оперативности задач координатно-временных определений. В таблицах 10.2, 10.3 и дальнейших комментариях представлена классификация, предложенная в [10.6].

Непрерывные измерения ведутся на пунктах глобальных или региональных сетей с целью уточнения опорных систем координат, определения параметров вращения Земли, изучения дрейфа континентов и других геодинамических явлений, а также для контроля за деформациями инженерных сооружений.

Режимы *статики* представляют собой наиболее распространенный метод геодезической съемки с использованием фазовых измерений. Наблюдения на пунктах ведутся столько времени, сколько необходимо для разрешения фазовых неоднозначностей и достижения заданной точности. В режиме *быстрой статики* время пребывания на пункте сокращается за счет наблюдения большего количества спутников.

Таблица 10.2. Методы фазовых GPS-определений с пост-обработкой

Метод	Описание	Время измерений	Ср. кв. погрешность координат	Примечание
Непрерывный	Одновременные наблюдения на многих пунктах	Неограниченно, еженедельные решения	Несколько мм	Двухчастотные приемники, расстояния между пунктами – до глобальных
Статика	Одновременные либо попеременные наблюдения на нескольких пунктах	От 1 часа до многих суток	1 см	Одно- или двухчастотные приемники, расстояния между пунктами – до 100 км
Быстрая статика	Одновременные, попеременные и повторные наблюдения на нескольких пунктах	От нескольких минут до 1 часа	1–2 см	Одно- или двухчастотные приемники, расстояния между пунктами 10–30 км
Стоп-старт	Опорная станция, инициализация и дальнейшие измерения передвижным приемником	Инициализация от неск. минут до 1 часа; измерения – несколько минут	2–3 см	Двухчастотный приемник, расстояния между пунктами – до 10 км
Кинематика	Опорная станция, съемка маршрута передвижным приемником	Инициализация от неск. минут до 1 часа;	До 10 см	Двухчастотный приемник

Метод *стоп-старт* включает два этапа: сначала из обработки измерений на исходном пункте разрешаются фазовые неоднозначности (этап *инициализации*), а затем, не останавливая работу приемника, перемещают его для выполнения измерений на определяемых пунктах. Метод *RTK*

Таблица 10.3. Методы GPS-определений в реальном времени.

Метод	Описание	Время измерений	Ср. кв. погрешность координат	Примечание
Кинематика	Опорная станция, съемка маршрута передвижным приемником	Инициализация несколько минут	Менее 10 см	Двухчастотные фазовые приемники
RTK (Старт-стоп в реальном времени)	Опорная станция, инициализация и дальнейшие измерения передвижным приемником	Инициализация – несколько. минут; измерения – несколько минут	2–3 см	Фазовые приемники, расстояния между пунктами до 10 км
Дифференциальный	Опорная станция, передвижной приемник, топосъемка или навигация	Несколько минут, возможно накопление	0,5–5 м	Псевдодальности, расстояния между пунктами –любые
Автономный	Одиночный приемник, топосъемка или навигация	Несколько минут, возможно накопление	*C/A*-код: 100 м (при SA), 10 м (без SA)	Псевдодальности

позволяет выполнять относительные определения координат по фазовым измерениям в режиме реального времени: данные с опорной станции передаются через сеть мобильного телефона прямо в вычислительное устройство приемника на определяемом пункте.

Кинематический метод используется, в основном, для целей навигации: он позволяет вести непрерывные измерения во время движения приемника. Современные модели приемников позволяют выполнять текущие измерения в движении с высоким временным разрешением (до 0,1 с).

Дифференциальные и *автономные* измерения основаны на приеме кодовых сигналов и используются для определений координат относительным и абсолютным методами, о которых сказано в п. 10.4.1.

10.5. Организационная структура GPS

Глобальная навигационная спутниковая система GPS находится под управлением министерства обороны США. Она предоставляется в стандартном режиме для гражданского, коммерческого и научного использования бесплатно. Существует *«Интерфейсный контрольный документ GPS»* (ICD-200-C-002, 25.9.97), в котором зафиксированы важные для пользователей технические подробности об устройстве и функционировании системы и о взаимодействии с ней.

Принято рассматривать систему GPS состоящей из трех составных частей: космического сегмента, сегмента управления и сегмента пользователей. С недавних пор стали говорить еще о четвертом («наземном») сегменте, включающем международные службы и функциональные дополнения к системе GPS, организационно не связанные с ней, но обеспечивающие возможность получения более точных результатов измерений, чем только штатными средствами.

10.5.1. Космический, управляющий и пользовательский сегменты

Космический сегмент, иначе называемый *«орбитальная группировка»*, образуют действующие на орбитах спутники. Они распределены пространственно с таким расчетом, чтобы в любой точке Земли выше 15^0 над горизонтом находилось бы всегда от 4 до 8 спутников. Для этого запущено 24 основных и несколько резервных спутников (в 2001 г. их было всего 28), распределенных по 6 круговым орбитам на высоте около 20200 км над поверхностью Земли. Плоскости орбит наклонены к плоскости экватора под углом 55^0 и равномерно развернуты друг относительно друга так, что их узлы (точки пересечения орбит с экватором) отстоят по долготе на 60^0. Период обращения спутников составляет 11 час. 58 мин., т.е. ровно половину звездных суток, и значит видимая с земли конфигурация созвездия спутников повторяется каждый день на 4 минуты раньше, чем в предыдущий.

Первый космический аппарат (КА) GPS типа Block-I был запущен на орбиту 22 февраля 1978 г. С тех пор было последовательно разработано несколько новых модификаций КА: Block-II, Block-IIA, Block-IIR, имеющих все более сложное и надежное оборудование и рассчитанных на срок службы 7-10 лет. Сейчас в основном на орбитах действуют КА последних двух типов.

В состав конструкции КА входят:

- средства синхронизации и временного обеспечения (бортовые часы) на базе нескольких рубидиевых и цезиевых стандартов частоты со стабильностью $(2\div5)\cdot10^{-13}$, работающих параллельно для повышения надежности;
- формирователи навигационных сигналов;
- бортовое вычислительное устройство с резервными ЭВМ;
- радиопередатчик, излучающий измерительные и телеметрические сигналы, и приемник управляющих команд и данных от ведущей станции GPS;
- системы электропитания, терморегулирования, ориентации.

Общий вид КА Block-II показан на рисунке 10.6. Панели солнечных батарей могут поворачиваться вокруг оси *Y*, а радиоизлучение от антенной системы идет в направлении оси *Z*. В рабочем положении спутник ориентирован так, что ось *Z* направлена на центр Земли, а плоскости панелей обращены к Солнцу.

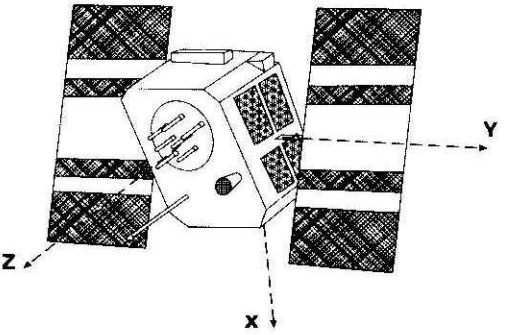


Рис. 10.6. Общий вид спутника GPS Block-II.

Сегмент управления GPS решает две группы задач:

- непрерывный *контроль за работоспособностью* аппаратуры спутников, выявление и дистанционное устранение сбоев, выдача команд по корректировке орбит;
- *эфемеридно-временное обеспечение* системы (уточнение текущих орбитальных параметров и показаний бортовых часов, загрузка этих данных на спутники для передачи в составе навигационного сообщения).

Сегмент управления состоит из сети наземных станций слежения, равномерно распределенных по Земле в близэкваториальных широтах. Сейчас сеть включает ведущую станцию и 5 станций слежения, 3 из которых

оснащены передающими антеннами для загрузки данных на спутники. Ведущая станция находится на авиабазе Фалкон (Шривер) ВВС США в районе г. Колорадо-Спрингс, штат Колорадо. Станции слежения расположены на атолле Диего-Гарсия (архипелаг Чагос в Индийском океане), на острове Вознесения (в Атлантическом океане), на Гавайях, на атолле Кваджалейн (в Тихом океане) и еще одна – совмещена с ведущей станцией. Антенны для загрузки информации на спутники расположены на станциях Диего-Гарсия, Кваджалейн и Вознесения. Они работают в автоматическом режиме и управляются из ведущей станции.

Станции слежения принимают сигналы спутников GPS при помощи прецизионных приемников и результаты измерения дальностей передают на ведущую станцию. Там по ним осуществляются точные расчеты и прогноз параметров орбит, ионосферной модели и корректирующих поправок для бортовых часов. Кроме того, на основе обработки метеорологической информации вычисляются текущие параметры модели тропосферных задержек. Все эти данные через радиостанции загрузки передаются на борт каждого спутника.

Сегмент управления устанавливает шкалу времени GPS, которая привязана к шкале времени UTC и поддерживается Военно-морской обсерваторией США. Начало отсчета этой шкалы времени установлено в полночь с 5 на 6 января.1980 г. Самой крупной единицей времени GPS является одна неделя (604800 с). Шкала времени GPS отличается от шкалы UTC во-первых – на целое число секунд (напомним, что в UTC иногда вводятся коррекции на 1 секунду), и во-вторых – на малую переменную величину, связанную с нестабильностью опорных генераторов, которую ведущая станция должна удерживать в пределах 1 мкс.

Для ограничения точности определения координат гражданскими пользователями может вводится режим селективного доступа SA (*Selective Availability*), при котором бортовая шкала времени спутников намеренно искажается по секретному алгоритму. В этом случае погрешность определения координат по *C/A*-кодовым измерениям составляет 50–100 м. Такой режим действовал на протяжении ряда лет, а с мая 2000 года он снят и указанные погрешности снизились в 10 раз.

Сегмент пользователей образует комплекс приемно-вычислительной аппаратуры – пользовательских навигационных приемников, способных выполнять измерения по сигналам спутников GPS. Может показаться не вполне логичным, что рассредоточенные по всему миру и никак не связанные между собой приемники разного типа и назначения считаются частью системы GPS, но это столь же верно, как, например, считать ваш личный телефонный аппарат частью телефонной сети, ее оконечным устройством.

Все разнообразие навигационных приемников GPS можно классифицировать по нескольким признакам:

- военного и гражданского назначения;
- навигационные и геодезические;
- кодовые и фазовые;
- одноканальные и многоканальные;
- универсальные и специализированные (например, для точных сличений разнесенных часов);
- технического назначения и бытовые.

Приемники каждой группы характеризуются разными функциональными возможностями в зависимости от предъявляемых к ним требований. Так, для военных приемников на первый план ставится условие надежности и оперативности их работы, а для геодезических – максимальной точности определений. При работе с кодовым приёмником нет необходимости вести непрерывный захват сигналов со спутников для вычисления местоположения. Это позволяет оборудованию работать на территориях с плотной застройкой и в местах с ограниченной видимостью, например, под деревьями. Для фазовых же приемников требуется непрерывный захват сигналов по меньшей мере четырех спутников, при этом должны одновременно работать несколько приемников.

В дешевых одноканальных приемниках измерение по сигналам каждого спутника происходит поочередно, а в многоканальных – сигналы после разделения распределяются по индивидуальным для каждого спутника каналам и их обработка ведется параллельно, что более эффективно. Особенно это преимущество проявляется когда приёмник передвигается или сигнал от спутников может прерываться.

Несмотря на богатое разнообразие конструктивных решений приемников GPS, принцип их работы одинаковый. Его можно пояснить с помощью обобщенной схемы устройства приемника [10.2], показанной на рисунке 10.7.

Антенное устройство приемника состоит из самой антенны и высокочастотного предусилителя. Антенна может быть выносной и соединяться с приемником коаксиальным кабелем, – тогда предусилитель располагается в корпусе антенны. Сейчас чувствительные элементы антенн спутниковых приемников изготовляют в виде печатных плат, рисунок которых обеспечивает благоприятную диаграмму направленности в форме полусферы. Для борьбы с многопутностью (см. п. 10.3.3) в антенном устройстве устанавливают защитные экраны.

Работа всех систем приемника синхронизируется по опорному кварцевому генератору с относительной стабильностью порядка 10^{-7}. В некоторых приемниках предусмотрена также возможность подключения внеш-

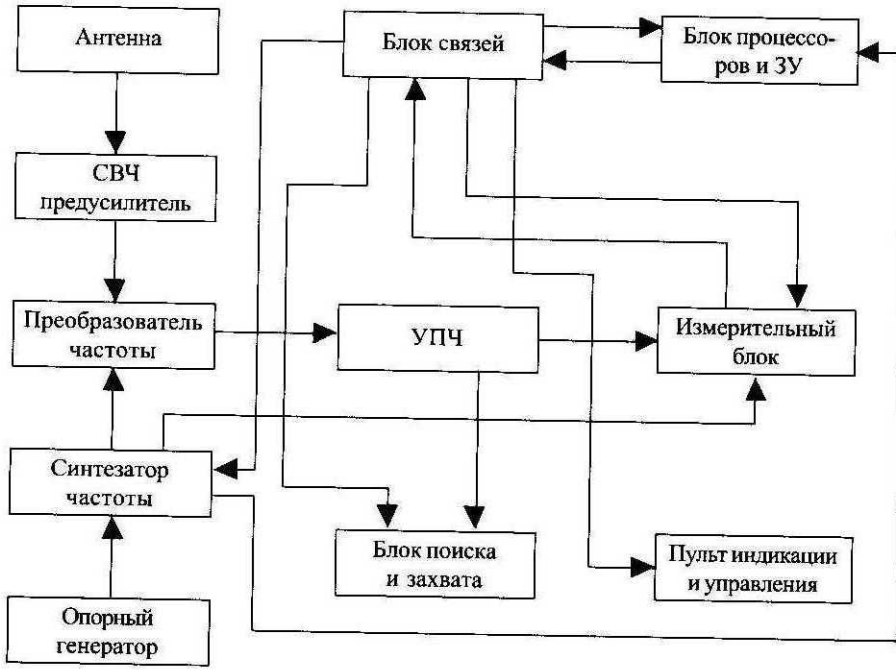


Рис. 10.7. Обобщенная структурная схема приемника GPS.

него стандарта частоты, что важно, например, при использовании приемника для сличений высокоточных часов. Необходимая для работы приемника сетка частот формируется с помощью синтезатора частот.

Радиочастотная часть приемника действует на основе супергетеродинного принципа, при котором частота принятого сигнала преобразуется в более низкую промежуточную, а затем сигнал усиливается в усилителе промежуточной частоты (УПЧ). С выхода УПЧ сигнал поступает в блок поиска и захвата сигналов спутников и в измерительный блок, действующие на основе описанного в п.10.2.3 корреляционного метода обработки. В многоканальных приемниках предварительно осуществляется селекция сигналов отдельных спутников, распределение их по разным каналам и дальнейшая параллельная обработка.

Запись накапливаемой в процессе измерений информации осуществляется в запоминающее устройство, из которого ее позже (или в темпе измерений) можно выбрать в компьютер через последовательный порт. Темп следования записей (шаг между последовательными данными измерений) задается пользователем. Обычно он составляет 30 секунд, но его можно выбрать и более частым, например, при измерениях во время движения (напомним, что время, необходимое для получения одиночных измерений при использовании *C/A*-кода, составляет 1–2 мс, см. п. 10.2.3).

10.5.2. Функциональные дополнения к GPS

Обратим внимание на последнюю строку таблицы 10.3: она характеризует как раз тот базовый («штатный») метод навигационных определений, который был изначально положен в основу при создании системы GPS. Как видим, при использовании общедоступного кода *C/A* даже при выключенном искусственном зашумлении *SA* средняя квадратическая погрешность (СКП) абсолютных координатных определений составляет 10 м. Практика показывает, что использование закрытого *P*-кода дает не очень большой выигрыш в точности (СКП = 4 м). Между тем, есть немало областей применения GPS, в которых допустимая погрешность навигационных определений в режиме реального времени не должна превышать 1 м или даже меньше. Вот два наиболее очевидных примера: управление заходом самолета на посадку и плавание в узкопроливных зонах. Для решения подобных задач сейчас широко применяются *дифференциальные системы*, основанные на относительных методах определений, которые рассмотрены в п. 10.4.1. Используется даже специальная аббревиатура для общего обозначения такого рода систем – DGPS (*Differential GPS*). Организационно они не связаны с ГНСС GPS, и потому рассматриваются как ее функциональные дополнения.

Мы уже говорили, что относительные методы основаны на одновременной работе двух приемников: одного – на исходной (опорной) станции с известными координатами, и другого – на определяемом пункте. В отличие от измерений внутри локальных сетей пунктов, при работе с дифференциальными системами только один из этих приемников принадлежит пользователю, а измерения на опорных станциях и забота о передаче корректирующих поправок всем потенциальным пользователям являются задачами DGPS. Но в аппаратуре пользователя, конечно, должна быть предусмотрена возможность приема и обработки данных конкретной системы DGPS.

Еще одной (и не мене важной!) задачей дифференциальных систем является текущий *контроль целостности* навигационного поля GPS, заключающийся в проверке качества и метрологической достоверности передаваемых измерительных сигналов и информации в навигационном сообщении. Фактически, это дополнительный контроль к тому, что ведется средствами сегмента управления системы GPS и, в некоторой степени, самой аппаратурой пользователя, но его задача более ответственна: свести до минимума риск использования навигационной аппаратурой ошибочных сигналов и данных во избежание катастрофических последствий. Эффективность контроля средствами DGPS достигается, в первую очередь, за счет весьма высокой оперативности передачи информации потребителям, исчисляемой секундами. Важно и то, что работа служб DGPS не только обеспечивает строгий контроль сигналов системы GPS, но и способствует улучшению характеристик ее целостности, так как по-

зволяет использовать даже аномальные спутниковые сигналы благодаря оперативной выдаче корректирующих данных.

Оперативность передачи данных DGPS потребителям достигается за счет использования выделенных частотных линий радиоканала, либо на частотах любительских радиостанций, либо по системам спутниковой связи (например, INMARSAT). Освоена также передача цифровых данных на частотах FM-радиостанций.

Дифференциальные системы условно подразделяются на широкозонные (радиус рабочей зоны до 6000 км), региональные (500–2000 км) и локальные (50–200 км).

Примером широкозонной системы DGPS является создаваемая WAAS (США), предназначенная для обслуживания североамериканского континента и Северной Атлантики. Она состоит из наземного и космического сегментов. Наземный сегмент образуют станции трех типов: *контрольные*, предназначенные для мониторинга состояния навигационного поля GPS, *главные*, ведущие централизованную обработку данных мониторинга, и *станции передачи* данных космическому сегменту. Последний образуют геостационарные ИСЗ GEOS, в задачу которых входит ретрансляция сформированных наземным сегментом дифференциальных поправок и информации о целостности системы, а также передача GPS-подобного сигнала на частоте *L1* для повышения доступности и надежности навигационных определений. Похожие широкозонные системы создаются в Европе (EGNOS) и в Японии (MSAS).

Региональные системы типа StarFix и SkyFix хотя и имеют меньший радиус действия, чем широкозонные, но их сеть охватывает многие регионы всех континентов. В действующих и проектируемых системах такого рода для передачи дифференциальных поправок чаще всего также предусмотрено использование геостационарных спутников связи. Что же касается локальных сетей DGPS, то они классифицируются по функциональному назначению на морские, авиационные и геодезические. Обычно они состоят из одной контрольно-корректирующей станции, а передача дифференциальных поправок и информации о целостности навигационного поля осуществляется через радиостанции в диапазоне средних волн и УКВ.

10.5.3. Международная служба IGS

Если главной задачей дифференциальных систем DGPS является обеспечение в режиме реального времени надежных навигационных измерений с приемлемой для практики точностью, то Международная геодинамическая служба IGS (*International GPS Service for Geodynamics*) предназначена для решения с помощью GPS технологий большого круга научно-технических и прикладных задач с максимально достижимой точностью. Требование работы в режиме реального времени здесь не ставится, но все же IGS выда-

ет прогнозированные координаты спутников с погрешностью 25 см (сопоставим с погрешностью эфемерид навигационного сообщения GPS в 3 м!), а современное организационно-техническое устройство IGS обеспечивает получение «окончательных» данных (СКП координат спутников 5 см, поправок часов 0,1 нс) с задержкой всего 13 суток [10.7, 10.8].

Идея добровольного объединения усилий мировой научно-технической общественности для создания IGS возникла в конце 1980-х годов, когда был осознан потенциально высочайший уровень точности, который может дать система GPS. После подготовительных мероприятий и экспериментов служба начала регулярную работу с 1994 г. Заложенная тогда организационная структура и схема функционирования IGS к началу 2002 года выглядит так (см. рис. 10.8).

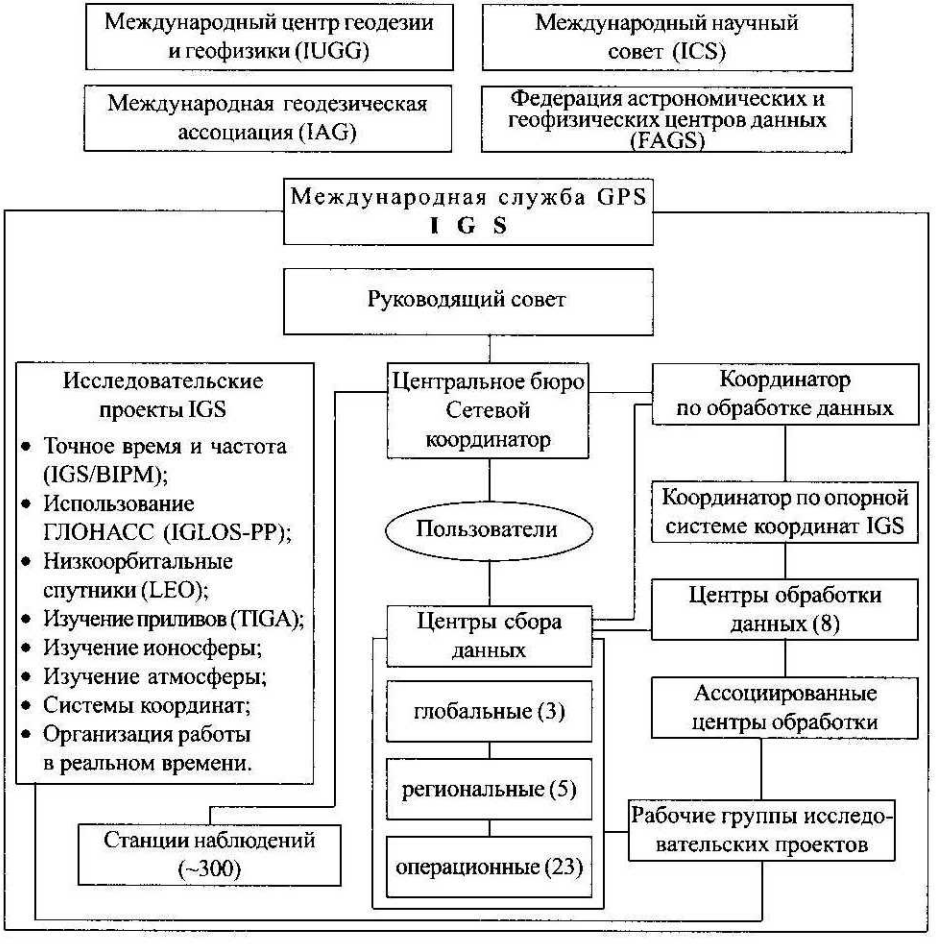


Рис. 10.8. Структура Международной службы IGS.

Работа службы основана на наблюдениях около 300 станций, расположенных по всему миру и принадлежащих примерно 70 организациям разных стран (рис. 10.9). Станции оснащены фазовыми двухчастотными GPS-приемниками. Наблюдения ведутся непрерывно, результаты ежесуточно (а по возможности – ежечасно) передаются через Интернет в центры сбора данных либо непосредственно, либо через операционные центры, осуществляющие методическое сопровождение нескольких станций.

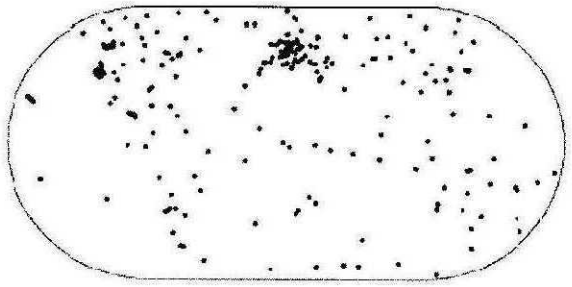


Рис. 10.9. Сеть постоянно действующих станций IGS.

Самые полные наборы данных содержат глобальные центры, два из которых расположены в США и один во Франции. Они оснащены мощной вычислительной техникой и обеспечивают открытый доступ к текущим и архивным данным одновременно большому числу пользователей. Глобальные центры автоматически сличают между собой содержимое баз данных.

Накапливаемые в центрах сбора данных результаты наблюдений поступают в центры обработки, иначе называемые центрами анализа (*Analysis Centers*). Глобальные и ассоциированные центры обработки действуют при научно-исследовательских институтах и лабораториях разных стран (из 8 глобальных центров 4 находятся в США, 2 – в Германии и по одному в Швейцарии и Канаде). В ходе обработки наблюдений вычисляются высокоточные орбиты спутников, поправки бортовых и наземных (подключенных к приемникам) часов, параметры вращения Земли (координаты полюса и длительность суток), изменения координат станций и атмосферные параметры. Результаты вычислений, полученные в разных центрах анализа, подвергаются совместной статистической обработке, в результате которой выводятся официальные данные IGS. Они направляются в те же центры сбора данных, в которых хранятся наблюдения, и таким образом, становятся доступными всем желающим. Существуют 3 типа выходных данных IGS, различающиеся по точности и оперативности вычислений. Их характеристики представлены в таблице 10.4.

Таблица 10.4. Характеристики официальных данных IGS.

Данные	Прогноз	Срочные	Окончательные	Единицы измерений
Задержка	Реальное время	17 часов	13 суток	
Орбиты	25,0	5,0	< 5,0	см
Часы	5,0	0,2	0,1	нс
Движение полюса	–	0,2	0,1	0,001".
Длительность суток	–	30,0	20,0	мкс/сут
Координаты станций (в плане / по высоте)	–	–	3,0 / 6,0	мм
Тропосферные поправки (в зените)	–	–	4,0	мм

Текущей деятельностью IGS руководит ее Центральное бюро, подотчетное избираемому Руководящему совету, в который входят специалисты IGS по научным направлениям и представители международных союзов и организаций.

Помимо текущей оперативной работы, связанной с выполнением наблюдений, их обработкой и выдачей данных потребителям, IGS ведет исследования по расширению и совершенствованию своей деятельности, для чего в ее составе образуются рабочие группы по выполнению исследовательских проектов (перечень ныне осуществляемых проектов дан на рис. 10.8).

10.6. Области применения GPS

По расчетам американских аналитиков к 2003 году объем продаж GPS приемников достигнет 4,7 млрд. долларов в автомобильном бизнесе, 3,8 млрд. долларов в туристическом бизнесе и более 6 млрд. долларов в сфере картографии и контроля грузоперевозок. Похоже, однако, на то, что уже сейчас процесс проникновения GPS-технологий в нашу жизнь становится стихийным и трудно контролируемым. Миниатюризация электроники захватывает и эту область: GPS-приемники в виде модулей-чипов уже устанавливаются в фотоаппаратах и камерах, часах, мобильных телефонах и т.п.

Вот далеко не исчерпывающийся список основных направлений использования средств глобального позиционирования в гражданской сфере:

- высшая геодезия и геодинамика: определение формы и размеров Земли, деформации ее поверхности, тектонические движения континентальных плит, предсказания землетрясений и т.п.;

- картографирование, геодезическая и аэрофотосъемка, кадастровая съемка, проектирование инженерных сооружений; геоинформационные системы (ГИС), содержащие электронные версии разнообразной информации о местности;
- мониторинг деформаций инженерных сооружений и естественных объектов в целях предупреждения техногенных катастроф;
- высокоточные сличения эталонов времени лабораторий всего мира; синхронизация систем связи и энергетических систем, обеспечивающая надежность их функционирования и минимизацию энергетических потерь;
- воздушная навигация: полет по маршруту (прокладка и отображение курса, определение расстояний до пункта назначения и оценка времени прибытия в аэропорт назначения); заход на посадку (здесь в первую очередь предъявляются требования к достоверности навигационной информации и обнаружению отказов с помощью контроля целостности системы);
- навигация морских и речных судов, обеспечение безопасности плавания в узкопроливных зонах; мониторинг рыболовецких флотов, квотирование и контроль добычи морских продуктов; определение рельефа дна и расчистка фарватеров, отслеживание загруженности портов, береговые службы контроля прохождения грузов;
- навигация в наземных условиях и диспетчерские службы грузовых автомобильных перевозок; выбор оптимального маршрута, отслеживание местоположения грузовика, оснащенного GPS-приемником; контроль целостности груза и своевременности прибытия и т.д.
- на железнодорожном транспорте – повышение безопасности движения, повышение эффективности работы диспетчерских служб, оптимизация режимов ведения поездов.
- гидрографические применения: постройка морских сооружений, прокладка трубопроводов, создание морских нефтяных платформ, сейсмологические исследования дна моря;
- в сельском хозяйстве – организация землепользования, оптимизация обработки посевных площадей и др.
- массовые применения средств GPS бытового назначения.

Мы постарались упорядочить список сфер применения GPS технологий в направлении от самых точных задач фундаментального назначения к задачам, требующим меньшей точности, но более высокой оперативности вплоть до работы в режиме реального времени. Сейчас даже трудно предположить, какие области применений GPS могут открыться в ближайшее время в сфере технологии и услуг населению. Как один из эффектных примеров такого рода применений приведем, в заключение, описание наруч-

ных часов со встроенным GPS-приемником (рис. 10.10). В дополнение к обычным функциям часов (время, дата, хронометр, секундомер, будильник и пр.) такие часы указывают потребителю его положение и скорость на маленьком дисплее На нем отображается направление к цели и время ее достижения. Часы хранят и поддерживают собственную шкалу времени, которую можно синхронизировать по шкале GPS и тогда часы будут правильно показывать местное время в любой точке Земли.

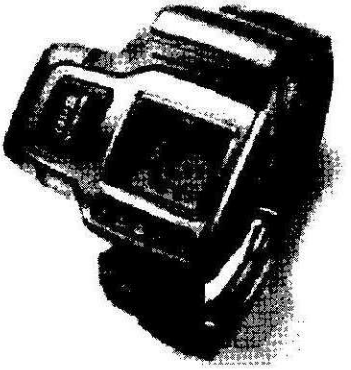


Рис. 10.10. Наручные часы со встроенным GPS-приемником

Часы работают как компас. Для пользования такой функцией достаточно пройти несколько метров и тогда стрелки покажут направление на северный полюс и направление движения. Часы могут использоваться как альтиметр, причем в отличие от обычного альтиметра их показания не зависят от атмосферного давления. Часы хорошо приспособлены для отдыха на природе, включая туризм, путешествия, хождение под парусами, пилотирование и др. Таким образом реализуются возможности GPS приемника, размеры и портативность которого соответствуют наручным часам.

Становится очевидным, что общество стоит на пороге массового внедрения в повседневную жизнь приемников GPS и других космических навигационных систем.

Приложение 1

Акронимы для лабораторий времени

Следующий список представляет акронимы лабораторий *k*, поддерживающих приближение UTC(*k*) к UTC или независимую локальную атомную шкалу TA(*k*), вместе с их названием и местонахождением.

AMC	Дублирующая станция ведущих часов, Колорадо Спрингс, CO, США
AOS	Астрономическая обсерватория, Жерокощиве, Боровиц, Польша
APL	Лаборатория прикладной физики, Лаурель, MA, США
AUS	Консорциум австралийских лабораторий
BEV	Федеральное управление по эталонам и измерениям, Вена, Австрия
BIRM	Пекинский институт радиометрологии и измерений, Пекин, КНР
CAO	Астрономическая обсерватория Каглиари, Каглиари, Италия
CH	Консорциум швейцарских лабораторий
CNM	Центр национальной метрологии, Мексико-сити, Мексика
CRL	Исследовательская лаборатория связи, Токио, Япония
CSAO	Шанхайская астрономическая обсерватория, Линтонг, КНР
CSIR	Совет по научным и техническим исследованиям, Претория, Южная Африка
DLR	Немецкий научно-исследовательский институт по воздухоплаванию и космонавтике, Оберпфафенхофен, Германия
DTAG	Немецкая телеком АГ, Дармштадт, Германия
F	Национальная комиссия по часам, Париж, Франция
GUM	Главная служба мер, Варшава, Польша
IEN	Национальный электротехнический институт Галилео Феррари, Турин, Италия
IEAG	Главная станция федерального агентства по картографии и геодезии, Ветцель, Германия
IGMA	Военный географический институт, Буэнос-Айрес, Аргентина
INPL	Национальная физическая лаборатория, Иерусалим, Израиль

IPQ	Португальский институт качества, Монте де Капарика, Португалия
JATC	Объединенная комиссия по атомному времени, Линтонг, КНР
KRIS	Корейский институт стандартов и науки, Таеджон, Южная Корея
LDS	Университет Лидса, Лидс, Великобритания
MSL	Лаборатория измерительных стандартов, Лоуер Хатт, Новозеландия
NAO	Национальная астрономическая обсерватория, Мисудзава, Япония
NIM	Национальный институт метрологии, Пекин, КНР
NIST	Национальный институт стандартов и технологии, Боулдер, CO, США
NPL	Национальная физическая лаборатория, Теддингтон, Великобритания
NPLI	Национальная физическая лаборатория, Нью-Дели, Индия
NRC	Национальный исследовательский совет, Оттава, Канада
NRLM	Национальная исследовательская лаборатория по метрологии, Цукуба, Япония
OMH	Будапешт, Венгрия
ONBA	Морская обсерватория, Буэнос-Айрес, Аргентина
ONRJ	Национальная обсерватория, Рио де Жанейро, Бразилия
OP	Обсерватория Парижа, Париж, Франция
ORB	Королевская обсерватория Бельгии, Брюссель, Бельгия
PTB	Физико-технический федеральный институт, Брауншвайг, Германия
ROA	Военно-морская обсерватория , Сан Фернандо, Испания
SCL	Лаборатория стандартов и поверок, Гонконг, КНР
SO	Шанхайская обсерватория, Шанхай, КНР
SP	Шведский национально-исследовательский институт испытаний, Борас, Швеция
SU	Институт метрологии времени и пространства (ИМВП), ГП “ВНИИФТРИ”, Менделеево, Московская обл. Россия
TL	Телекоммуникационная лаборатория, Чунг-Ли, Тайвань
TP	Институт радиотехники и электроники, академия наук , Чешская республика
UME	Национальный метрологический институт, Гебзе-Кокаели, Турция
USNO	Военно-морская обсерватория, Вашингтон, округ Колумбия, США
VSL	Лаборатория ван Свидена, Дельфт, Нидерланды

Приложение 2

Сокращения

Шкалы времени

Сокращения, написанные жирными буквами, являются официальными на всех языках.

EAL	Свободная (неуправляемая) атомная шкала
ET	Эфемеридное время
GMST	Гринвичское среднее звездное время
GMT	Гринвичское среднее время
TA(*k*)	Атомное время, устанавливаемое лабораторией *k*
TAI	Международное атомное время
TCB	Барицентрическое координатное время
TCG	Геоцентрическое координатное время
TDB	Барицентрическое динамическое время
TDT	Земное динамическое время (стало TT в 1991 году)
TT	Земное время
UT	Всемирное время
UTC	Всемирное координированное время
UT1	Всемирное время 1 (существуют также виды 0 и 2)

Разное

BIH	Международное бюро времени
BIPM	Международное бюро мер и весов
CCDS	Консультативный комитет по определению секунды (стал CCTF в 1997 году)
CCIR	Международный консультативный комитет по радио
CCTF	Консультативный комитет по времени и частоте
CGPM	Генеральная конференция по мерам и весам
CIPM	Международный комитет по мерам и весам
CNES	Национальный институт изучения космоса, Франция
GLONASS	Глобальная навигационная спутниковая система ГЛОНАСС
GPS	Глобальная система навигации и определения положения
IAU	Международный астрономический союз

ICRF	Международная небесная координатная система
IERS	Международная служба вращения Земли
ITRF	Международная земная координатная система
ITU	Международный союз электросвязи
IUGG	Международный союз геодезии и геофизики
JD	Юлианская дата
LHA	Лаборатория атомных часов, Орсэ, Франция
LPTF	Лаборатория первичного времени и частоты, Париж
MJD	Модифицированная юлианская дата
NASA	Национальное управление по аэронавтике и космосу
SI	Международная система единиц
TWSTFT	Двусторонняя спутниковая передача времени и частоты
URSI	Международный союз радионауки
VLBI	Интерферометрия со сверхдлинной базой (РСДБ)
VLF	Сверхдлинные волны (СДВ)
WGS	Всемирная геодезическая система

Приложение 3

Определение основных еденицСИ

Метр (м)

Метр есть расстояние, проходимое светом в вакууме за интервал времени 1/299 792 458 секунды. (17-я генеральная конференция по мерам и весам, 1983 год)

Килограмм (кг)

Генеральная конференция одобрила прототип килограмма, принятый международным комитетом. Этот прототип с тех пор рассматривается как единица массы. (1-я CGPM, 1889 год).

Килограмм является единицей массы; он равен массе международного прототипа килограмма. (3-я CGPM, 1901 год).

Секунда (с)

Секунда есть длительность 9 192 631 770 периодов излучения, соответствующего переходу между двумя сверхтонкими уровнями основного состояния атома цезия 133. (13-я CGPM, 1967 год).

Ампер (А)

Ампер есть постоянный ток, который при протекании по двум прямым параллельным проводникам бесконечной длины с пренебрежимо малым поперечным сечением и расположенным в вакууме друг от друга на расстоянии 1 метр, вызывает между этими проводниками притяжение с силой, равной 2×10^{-7} ньютона на один метр длины. (9-я CGPM, 1948 год).

Кельвин (К)

Кельвин, единица термодинамической температуры, есть 1/273,16 доля термодинамической температуры тройной точки воды. (13-я CGPM, 1967 год).

Моль (моль)

(1) Моль есть количество вещества в системе, содержаще столько элементарных частиц, сколько атомов в 0,012 килограмма углерода 12; обозначение «моль».

(2) Если используется моль, то элементарные частицы должны быть указаны и они могут быть атомами, молекулами, ионами, электронами, другими частицами или оговоренными группами таких частиц. (14-я CGPM, 1971 год)

Кандела (кд)

Кандела есть световая интенсивность в данном направлении от источника, излучающего 1/683 ватта на стерадиан монохроматического излучения на частоте 540×10^{12} Гц..(16-я CGPM, 1979 год).

Приложение 4

Международные службы

Время

Международное бюро мер и весов, секция времени, Pavillon de Breteuil, F-92312, Sevres Cedex (France)

Телефон	Коммутатор МБМВ	+33 1 45 07 70 70
Факс	Общий МБМВ	+33 1 45 34 20 21
	Секция времени МБМВ	+33 1 45 07 70 59
Интернет	`http://www.bipm.org`	
	или анонимный ftp 62.161.69.5	
	(поддиректория TAI)	
E-mail	`tai@bipm.org`	

Вращение Земли (Всемирное Время)

Международная служба вращения Земли (IERS – русск. МСВЗ)

Центральное бюро

Федеральное управления по картографии и геодезии (BKG), Richard Strauss Allee 11, D-60598, Frankfurt-am-Main (Germany)

Телефон	+49 69 6333 273
Факс	+49 69 6333 425
E-mail	`richter@ifag.de`

Центр обработки ориентации Земли

Обсерватория Парижа, отдел фундаментальной астрономии, 61, avenue de l'Observatoire, 75014/ Paris (France)

Интернет	`http://hpiers.obspm.fr`
	или анонимный ftp hpiers.obspm.fr 145.238.100.28
E-mail	`gambis@obspm.fr`

Быстрая служба предсказания вращения параметров Земли

Военно-морская обсерватория, отделение ориентации Земли,
3450 Massachusetts Avenue NW, Washington DC 20392-5420, USA.

Интернет	`http://maia.usno.navy.mil`	
	или анонимный ftp `maia.usno.navy.mil`	192.5.41.22
E-mail	`jimr@maia.usno.navy.mil`	